HAL
open science

Metabarcoding in the abyss : uncovering deep-sea biodiversity through environmental DNA

Miriam Isabelle Brandt

To cite this version:

Miriam Isabelle Brandt. Metabarcoding in the abyss : uncovering deep-sea biodiversity through environmental DNA. Agricultural sciences. Université Montpellier, 2020. English. NNT : 2020MONTG033 . tel-03197842

HAL Id: tel-03197842

https://theses.hal.science/tel-03197842

Submitted on 14 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE POUR OBTENIR LE GRADE DE DOCTEUR
DE L'UNIVERSITÉ DE MONTPELLIER

En Sciences de l'Évolution et de la Biodiversité

École doctorale GAIA

Unité mixte de recherche MARBEC

Pourquoi Pas les Abysses ? L'ADN environnemental pour l'étude de la biodiversité des grands fonds marins

Metabarcoding in the abyss: uncovering deep-sea biodiversity through environmental DNA

Présentée par Miriam Isabelle BRANDT
Le 10 juillet 2020

Sous la direction de Sophie ARNAUD-HAOND
et Daniela ZEPPILLI

Devant le jury composé de

Sofie DERYCKE, Senior researcher/Professeur rang A, ILVO, Belgique	Rapporteur
Naiara RODRIGUEZ-EZPELETA, Senior researcher/ Chercheur rang B, AZTI, Espagne	Rapporteur
Jan PAWLOWSKI, Associate Professor/ Professeur rang A, Université de Genève	Examinateur
Anne CHENUIL, Directrice de recherche, CNRS IMBE	Examinateur, Président du jury
Frédérique VIARD, Directrice de recherche, CMRS UPMC	Examinateur
Sophie ARNAUD-HAOND, Cadre de recherche des EPIC, IFREMER MARBEC	Directrice de thèse
Daniela ZEPPILLI, Cadre de recherche des EPIC, IFREMER LEP	Invitée et co-encadrante de thèse

UNIVERSITÉ DE MONTPELLIER

Any intelligent fool can make things bigger, more complex, and more violent. It takes a touch of genius — and a lot of courage — to move in the opposite direction.

Ernst Friedrich Schumacher

General abstract

The abyssal seafloor covers more than half of planet Earth. It can host a large number of, mostly small and still undescribed, organisms (~50,000-5 million individuals/m^2), contributing to key ecosystem functions such as nutrient cycling, sediment stabilisation and transport, or secondary production. Technological developments in the past 30 years have allowed remarkable advances, yet due to the vastness and remoteness of deep-sea habitats, ecological studies have been limited to local or regional scales. Indeed, we have so far explored less than 1% of the deep seafloor, although the latter is under increased threat from a variety of anthropogenic pressures.

This PhD aimed at bringing new perspectives for the study of biodiversity and biogeography in the deep-sea, to bridge this large knowledge gap, and advance toward the development of efficient biomonitoring protocols to preserve this vast and elusive backyard.

We investigated the potential of multi-marker environmental DNA (eDNA) metabarcoding to assess the extent and distribution patterns of biodiversity in this remote ecosystem. Using mitochondrial and nuclear marker genes, this PhD aimed at producing and testing an optimized eDNA metabarcoding workflow for deep-sea sediments, on a bioinformatic, molecular, and sample processing level, applicable to multiple life compartments including bacteria, protists, and metazoans.

Biodiversity assessment with eDNA is confronted with the difficulty in defining accurate "species-level" molecular Operational Taxonomic Units (OTUs), as numerous sources of error induce frequent overestimations. The first part of this thesis describes how newly developed bioinformatic tools can be combined in order to get access different levels of biotic diversity, and underline the advantages of clustering and LULU-curation for producing metazoan biodiversity inventories at the level of the morphospecies.

Moreover, the accuracy of protocols based on eDNA in deep sea sediments still needs to be assessed, as results may be biased by ancient DNA, resulting in biodiversity assessments not targeting *live* organisms.This thesis assessed the potential bias of ancient DNA by 1) evaluating the effect of removing short DNA fragments, and 2) comparing communities revealed by co-extracted DNA and RNA in five deep-sea sites. Results indicated that short DNA fragments do not affect alpha and beta diversity, but that DNA obtained from 10g of sediment should be favoured over RNA for logistically realistic, repeatable, and reliable surveys. Results also

confirm that increasing the number of biological rather than technical replicates is important to infer robust ecological patterns.

Sieving sediment to separate benthic size classes increased the number of detected meiofauna OTUs, but was not essential for achieving robust biodiversity estimates, and should be avoided if unicellular organisms are also of interest. More importantly, aboveground water and superficial sediment detected significantly different communities in all taxonomic compartments, even when large volumes of water were sampled, emphasising that eDNA metabarcoding of aboveground water samples is not suitable for benthic biodiversity surveys.

Finally, this thesis applied the optimized eDNA metabarcoding protocols to investigate the influence of biotic and abiotic factors on the extent and distribution of deep-sea metazoan biodiversity on an East-West transect ranging from the Central Mediterranean to the Mid-Atlantic Ridge. Results, consistent to morphology-based studies, confirm that small-scale biotic and abiotic factors lead to significant vertical changes in metazoan richness and community structure within the sediment, and highlight that regional beta-diversity patterns result from a combined influence of past biogeography and present day processes.

This thesis opens the way to large-scale eDNA-based studies in the deep-sea realm, thus contributing to a better understanding of biodiversity, biogeography, and ecosystem function in this vast and still poorly known biome.

Résumé général

Les fonds abyssaux couvrent plus de la moitié de la planète Terre. Ils peuvent héberger un grand nombre d'organismes (~ 50000 à 5 millions d'individus/m^2), pour la plupart petits et encore non décrits, contribuant à des fonctions écosystémiques clés telles que le recyclage du carbone ou la productivité secondaire.

Les développements technologiques ont permis des avancées remarquables, mais l'immensité et l'éloignement des habitats profonds ont restreints les études aux niveaux local et régional. Nous avons exploré moins de 1% des fonds marins, alors que ces-derniers forment l'un des plus grands biomes sur Terre et sont de plus en plus sous pression anthropique.

Cette thèse vise à apporter de nouvelles perspectives pour l'étude de la biodiversité et de la biogéographie en environnements profonds, pour combler ce déficit de connaissances et permettre le développement de protocoles de biosurveillance efficaces.

Nous avons étudié le potentiel du métabarcoding d'ADN environnemental (ADNe) pour évaluer l'étendue et la distribution de la biodiversité en environnements profonds. À l'aide de gènes marqueurs mitochondriaux et nucléaires, cette thèse vise à produire un protocole de métabarcoding d'ADNe pour les sédiments des grands fonds, optimisé à un niveau de traitement bioinformatique, moléculaire et d'échantillonnage, et applicable à plusieurs compartiments du vivant.

L'évaluation de la biodiversité avec l'ADNe est confrontée à la difficulté de définir des Unités Taxonomiques Opérationnelles (OTU) au niveau de «l'espèce», car de nombreuses sources d'erreur induisent de fréquentes surestimations. Le premier chapitre de cette thèse décrit comment des outils bioinformatiques nouvellement développés peuvent être combinés afin d'accéder à divers niveaux de diversité biotique, et soulignent les avantages du clustering et de l'outil LULU pour produire des inventaires de biodiversité métazoaire plus proches du niveau de l'espèce morphologique.

De plus, la précision des protocoles basés sur l'ADNe dans les sédiments profonds devait être évaluée, car les résultats peuvent être biaisés par de l'ADN ancien archivé dans le sédiment, ce qui conduirait à des estimations de biodiversité passée plutôt que présente. Dans un second temps, nous avons donc estimé le biais de l'ADN ancien en 1) évaluant l'effet de l'élimination de courts fragments d'ADN, et 2) en comparant les communautés révélées par l'ADN et l'ARN co-extraits. Les résultats indiquent que les fragments d'ADN courts n'affectent pas la diversité

alpha et bêta, que l'ADN obtenu à partir de 10 g de sédiments est plus approprié que l'ARN pour des études exhaustives logistiquement réalistes, et que les réplicas biologiques plutôt que techniques sont importants pour inférer des patrons écologiques fiables.

Le tamisage des sédiments séparant les organismes benthiques par classe de taille a augmenté le nombre d'OTU métazoaires détectées, mais n'était pas essentiel pour obtenir des patrons écologiques robustes, et devrait être évité si des taxons unicellulaires sont également ciblés. De même la comparaison de communautés détectées par des échantillons d'eau affleurante et de sédiments a montré que l' eaux affleurante ne représente pas une alternative aux sédiments pour effectuer des inventaires de diversité benthique.

Enfin, les protocoles optimisés de métabarcoding d'ADNe ont été appliqué pour étudier l'influence de facteurs biotiques et abiotiques sur la biodiversité métazoaire des grands fonds, allant de la Méditerranée centrale à la dorsale médio-atlantique. Les résultats confirment que des facteurs agissant à très petite échelle (cm) conduisent à des changements verticaux significatifs de la richesse et de la structure des communautés dans les sédiments, et soulignent que les tendances régionales de diversité bêta résultent d'une influence combinée de la biogéographie passée et de phénomènes actuels.

Cette thèse ouvre la voie à des études de biodiversité globale dans les environnements profonds, contribuant ainsi à une meilleure compréhension de la biogéographie et des fonctionnements écosystémiques dans ce vaste biome encore mal connu.

Declaration

I declare that this doctoral thesis, entitled "Metabarcoding in the abyss: uncovering deep-sea biodiversity through environmental DNA", has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where states otherwise by reference or acknowledgment, the work presented is entirely my own.

My doctoral studies led me to participate in the following scientific publications:

Brandt MI, Trouche B, Quintric L, Wincker P, Poulain J, and Arnaud-Haond S. A flexible pipeline combining clustering and correction tools for prokaryotic and eukaryotic metabarcoding. BioRxiv, 717355, ver. 3 peer-reviewed and recommended by PCI Ecology. doi: 10.1101/717355. 2020. *In revision at Mol. Ecol. Res.*

Brandt MI, Trouche B, Henry N, Liautard-Haag C, Maignien L, de Vargas C, Wincker P, Poulain J, Zeppilli D, and Arnaud-Haond S. An assessment of environmental metabarcoding protocols aiming at favouring contemporary biodiversity in inventories of deep-sea communities. *Front. Mar. Sci*., 7, 2020. doi: 10.1101/836080

Brandt MI, Pradillon F, Trouche B, Henry N, Liautard-Haag C, Cambon-Bonavita MA, Wincker P, Poulain J, Arnaud-Haond S, and Zeppilli D. Evaluating sediment and water sampling methods for the estimation of deep-sea biodiversity using environmental DNA. In preparation. 2021. *Accepted with minor revisions at Sci. Rep.*

Brandt MI, Zeppilli D, Günther B, Trouche B, Henry N, Simier M, Orejas C, Pradillon F, Fabri MC, Cambon-Bonavita MA, Guieu C, Quintric L, Belser C, Poulain J, Wincker P, Maignien L, de Vargas C, and Arnaud-Haond S. Multi-marker eDNA metabarcoding reveals large and small-scale metazoan metazoan biodiversity patterns in the deep-sea Atlantic-Mediterranean transition zone. *In preparation.*

Saeedi H, Reimer JD, **Brandt MI**, Dumais PO, Jażdżewska AM, Jeffery NW, Thielen PM, Costello MJ. "Global marine biodiversity in the context of achieving the Aichi Targets: Ways forward and addressing data gaps," *PeerJ*, **7**, e7221, 2019. doi: 10.7717/peerj.7221

Cowart DA, Matabos M, **Brandt MI**, Marticorena J, and Sarrazin J, "Exploring Environmental DNA (eDNA) to Assess Biodiversity of Hard Substratum Faunal Communities on the Lucky Strike Vent Field (Mid-Atlantic Ridge) and Investigate Recolonization Dynamics After an Induced Disturbance," *Front. Mar. Sci*., 6, 2020. doi: 10.3389/fmars.2019.00783

Arnaud-Haond S, **Brandt MI**, Trouche B, Atteia A, Poulain J, Wincker P, Zabel M, Wenzhofer F, Glud R N., Zeppilli D. Distribution and Drivers of Benthic Biodiversity in Hadal and Abyssal Realms through eDNA-based Inventories. *In preparation*.

Lins-Pereira L, Zeppilli D *et al.* Methods to ecological survey of abyssal infauna in the context of deep-sea polymetallic nodule mining impact. *In preparation.*

Trouche B, **Brandt MI**, Orejas C, Pradillon F, Fabri MC, Cambon-Bonavita MA, Guieu C, Quintric L, Belser C, Poulain J, Wincker P, Arnaud-Haond S & Maignien L. Biogeographical patterns of benthic prokaryote communities in the Atlantic-Mediterranean transition. *In preparation.*

Date: 10 February 2020, revised 12 January 2021.

Print name: Miriam Isabelle Brandt

Signature:

Table of contents

General abstract ii
Résumé général iv
Declaration vi
Table of contents viii
Glossary xiv
Chapter I. Introduction and literature review 1
I.I. Why study biodiversity? 1
Estimating biodiversity 2
The species concept 5
Global biodiversity estimates, and their limitations 7
I.II. Estimating biodiversity in the 21st century: the revolution of DNA-based taxon identification approaches 12
Species identification with DNA barcodes 12
From barcoding of single species to metabarcoding of whole communities 15
Challenges and uncertainties of metabarcoding approaches 16
Advantages and applications of eDNA metabarcoding for biodiversity assessments 24
I.III. The deep sea, the last frontier on earth 25
I.IV. Oceanic regions and their associated deep-sea ecosystems 29
Abyssal plains 31
Mid-ocean ridges 32
Active and passive continental margins 34
Seamounts 36
Hadal trenches 37
Other known benthic ecosystems: OMZs and organic falls 37
I.V. Deep-sea benthic biodiversity patterns 39
Benthic size classes and their main differences 40
Ecological patterns at regional (~100-10,000 km) spatial scales 43
Ecological patterns at habitat to small spatial scales 46
I.VI. Aims and objectives 49

Objectives 50

Chapter II. Bioinformatic pipelines combining correction and clustering tools allow for flexible and comprehensive prokaryotic and eukaryotic metabarcoding 52

Résumé en français 54

Abstract 55

1 Introduction 56

2 Materials and methods 59

2.1 Preparation of samples 59

Mock communities 59

Environmental DNA 59

2.2 Amplicon library construction and high-throughput sequencing 60

Eukaryotic 18S V1-V2 rRNA gene amplicon generation 60

Eukaryotic COI gene amplicon generation 60

Prokaryotic 16S rRNA gene amplicon generation 61

Amplicon library preparation 61

Sequencing library quality control 61

Sequencing procedures 62

2.3 Bioinformatic analyses 62

Reads preprocessing 62

Read correction, amplicon cluster generation and taxonomic assignment 63

2.4 Statistical analyses 65

3 Results 66

3.1 Alpha diversity in mock communities 66

3.2 Alpha-diversity patterns in environmental samples 69

3.3 Patterns of beta-diversity between pipelines 72

3.4 Taxonomic assignment quality 74

4 Discussion 76

4.1 ASVs *vs* OTUs: a choice depending on taxon of interest and research question 76

4.2 Importance of parameter adjustment for LULU curation 78

4.3 Taxonomic resolution and assignment quality 80

5 Conclusions and perspectives 81

Data accessibility 82

Acknowledgements 82

Conflict of interest disclosure 82

Author contributions 83

Chapter III. An assessment of environmental metabarcoding protocols aiming at favouring contemporary biodiversity in inventories of deep-sea communities 84

Résumé en français 86

Abstract 88

1 Introduction 89

2 Materials and methods 92

2.1 Collection of samples 92

2.2 Nucleic acid extractions and molecular treatments 92

eDNA with the 10 g-PowerMax kit 92

Size-selection of eDNA extracts 92

Ethanol reconcentration of eDNA extracts 92

Joint environmental DNA/RNA with the 2 g- RNeasy PowerSoil kit 93

2.3 PCR amplification and sequencing 93

2.4 Bioinformatic analyses 93

2.5 Statistical analyses 95

3 Results 96

3.1 High-throughput sequencing results 96

3.2 Alpha diversity between processing methods 97

3.3 Effect of molecular processing methods on beta-diversity patterns 100

3.4 Extraction kit *vs* nature of nucleic acid 101

4 Discussion 103

Acknowledgements 106

Author contributions 107

Data accessibility 107

Conflicts of Interest 107

Chapter IV. Evaluating sediment and water sampling methods for the estimation of deep-sea biodiversity using environmental DNA 108
Résumé en français 110
Abstract 111
1 Introduction 112
2 Materials and methods 114
2.1 Sample collection 114
2.2 Nucleic acid extractions 115
2.3 PCR amplification and sequencing 115
2.4 Bioinformatic analyses 116
2.5 Statistical analyses 117
3 Results 118
3.1 High-throughput sequencing results 118
3.2 Alpha diversity between sampling methods 119
3.3 Effect of sampling method on community structures 124
4 Discussion 127
4.1 Importance of substrate nature 127
4.2 Sieving sediment is not essential for comprehensive benthic biodiversity surveys 128
4.3 Adjusting water sample volume and filter mesh size to target organisms 130
Acknowledgements 132
Author contributions 132
Additional Information 133
Chapter V. Multi-marker eDNA metabarcoding reveals large and small-scale metazoan biodiversity patterns in the deep Atlantic-Mediterranean transition zone 134
Résumé en français 136
Abstract 138
1 Introduction 139
2 Materials and methods 141
2.1 Preparation of samples 141
Environmental DNA 141

Mock samples 142
Organic matter content and sediment grain size 143
2.2 PCR amplification and sequencing 143
Eukaryotic 18S V1-V2 rRNA gene amplicon generation 143
Eukaryotic COI gene amplicon generation 144
Eukaryotic 18S V9 rRNA gene amplicon generation 144
Prokaryotic 16S V4-V5 rRNA gene amplicon generation 144
Amplicon library preparation 144
Sequencing library quality control 145
Sequencing procedure 145
2.3 Bioinformatic analyses 145
2.4 Statistical analyses 146
3 Results 148
3.1 High-throughput sequencing results 148
3.2 OTU richness decreases with depth in the sediment 149
3.3 Numbers and nature of OTUs change across the Gibraltar Strait 151
3.4 Influence of small and large-scale factors on community structures 153
4 Discussion 157
Acknowledgements 161
Author contributions 162
Data accessibility 162
Chapter VI. General discussion 163
Conclusions and future directions 175
Bibliography 177
Acknowledgements 209
Appendix 210
Résumé substantiel en français 211
Supplementary material Chapter II. 217
Supplemental tables 217
Supplemental figures 224
Supplementary material Chapter III. 227

Supplemental materials and methods.....227
Supplemental tables 229
Supplemental figures 236
Supplementary material Chapter IV. 242
Supplemental materials and methods.....242
Supplemental tables 244
Supplemental figures 249
Supplementary material Chapter V. 252
Supplemental tables 252
Supplemental figures 256

Glossary

18S rDNA or 18S SSU rRNA gene	Gene coding for the 18S ribosomal RNA, a part of the eukaryotic small ribosomal subunit (40S).
BOLD	Barcode Of Life Database
COI, CO1, COX1	Cytochrome c Oxidase subunit I gene, a mitochondrial gene found in all life forms, encoding for the main subunit of the cytochrome c oxidase complex, the last enzyme in the respiratory electron transport chain.
DHAB	Deep Hypersaline Anoxic Basin
DNA barcode	Standard DNA fragment allowing taxonomic identification. Selection of informative DNA regions is crucial. An ideal DNA barcode should have low intra-specific and high inter-specific variability, and possess conserved flanking sites for developing universal PCR primers allowing wide taxonomic application. For animals, the most commonly used barcodes are the cytochrome c oxidase subunit I gene and the 18S SSU rRNA gene.
Barcoding	Method of species identification using a genetic barcode.
Bulk DNA	DNA extracted from a pool of organisms.
Chimera	Unique sequence resulting from the recombination in vitro (e.g., during PCR) of multiple parent sequences.
Degenerated primer	Primer with wobbles (N), i.e. equimolar mixtures of two or more primer sequences with different bases at a given position within the primer; or including Inosine, a reduced nucleotide that is able to bind on any nucleotide.
Environmental DNA (eDNA)	DNA extracted directly from an environmental sample, i.e. soil, water, or air.
GC content	Guanine-cytosine content, the percentage of nitrogenous bases in a DNA or RNA molecule
HTS	High-Throughput (DNA) Sequencing
Macrofauna	Organisms in the 1 cm- 2 cm size range
Meiofauna	Organisms in the 32 µm – 1 cm size range
Megafauna	Organisms larger than 2 cm in size

Metabarcoding	Barcoding of samples that contain more than one organism (bulk DNA or eDNA, allowing for the simultaneous identification of several taxa within the same sample, using High-Throughput Sequencing technologies.
Multiplexing	Pooling of sample DNA libraries at equal concentration for simultaneous sequencing. In order to reassign each sequence to a unique sample after DNA sequencing (demultiplexing), sample specific indices or tags are added to the DNA fragments of each sample.
NGS	Next-Generation (DNA) Sequencing, a synonym for HTS
OM content	Organic Matter percent content
OMZ	Oxygen Minimum Zone
OTU	Operational Taxonomic Unit
PCR	Polymerase Chain Reaction, a technique relying on DNA polymerase enzymes, used to make numerous copies of a specific DNA fragment. This exponentially amplifies the amount of DNA, allowing its detailed analysis.
Primer	Oligonucleotides (short single-stranded DNA or RNA fragments), that are the starting point for DNA synthesis (DNA replication).
Quality-check	Procedure of sequence controlling and selection, based on their length and quality scores.
RDP	Ribosomal Database Project
Sequencing run	Sequencing instrument parameters and run type (single read or paired end) used to perform high-throughput sequencing
Sequence tag	Sample-specific index or tag used to associate a sequence to a unique sample.
Tag-switching (Cross-talk)	The assignment of a sequence to a wrong sample, due to recombination of sequence tag arising from either cross-contamination with tagged primers or from mixed clusters on the flow-cell.

Chapter I.
Introduction and literature review

I.I. Why study biodiversity?

How many species are there on Earth? What influences their distribution? These two simple questions about Life remain enigmatic. These questions are important, as biodiversity − the variety of Life on Earth − is a major component of Earth's life-support system, the product of over 3 billion years of evolution, but is increasingly under threat from human activities.

Natural ecosystems provide great benefits to human societies, and there is global consensus that human well-being depends on healthy ecosystems (Stokstad 2005). Global efforts to understand and preserve our natural world have increased since the 1970s, with a milestone set during the Rio Earth Summit, the first United Nations Conference of the Parties (COP). This conference was the starting point of the United Nations Framework Convention on Climate Change (UNFCCC) and the Convention on Biological Diversity (CBD), which are the world's first commitments to sustainable development and the conservation of biological diversity.

Since then, a lot of research has highlighted the relationship between biodiversity and ecosystem health. Biodiversity encompasses variation among genes, species, and functional traits (Cardinale et al. 2012) .The value of biodiversity lies in the myriad of roles all these life forms perform, and which make complex biotic systems possible. The functioning of these ecosystems, i.e. the way they store resources, produce biomass, decompose and recycle nutrients, is tightly linked to the biodiversity they harbour. Research has also shown that higher levels of diversity are associated with higher ecosystem stability through time (Cardinale et al. 2012). Finally, natural ecosystems, and thus the biodiversity they harbour, also provide a series of so-called "ecosystem services". These range from valuable goods in industry or agriculture, to clean drinking water, or the regulation and stability of fundamental Life equilibria such as disease outbreak mitigation, pollination, or climate stability (Palmer et al. 2004; Cardinale et al. 2012; Rohr et al. 2020).

Despite the Rio agreements, a 2010 review of the state of biodiversity showed ongoing declines and increasing levels of anthropogenic pressure (Butchart et al. 2010). A primary

consequence of these stressors is an accelerated loss of populations and species in the past century (Cardinale et al. 2012). This is associated with increased rates of resource collapse, decreased ecosystem productivity, decreased resistance and resilience capacities, and decreased water quality (Worm et al. 2006). Overall, the impacts of diversity loss on ecosystem functioning might be among the major drivers of global environmental change (Cardinale et al. 2012). In order to maintain ecosystem services, and preserve human well-being, it is therefore essential that we acquire a better understanding of the natural patterns and processes that sustain ecosystem functioning, among which biological diversity.

Estimating biodiversity

Walter Rosen first introduced the term *biodiversity*, contraction of biological diversity, in 1986, during the National Forum on Biological Diversity in Washington DC. The United Nations Environment Program (UNEP) now defines biological diversity as "*the variability among living organisms from all sources […] and the ecological complexes of which they are part; this includes diversity within species, between species, and of ecosystems*". Hubbell (2001) proposed a simpler and more precise definition, describing biodiversity as "*species richness and their relative abundance in space and time*".

The biodiversity concept, although intuitively easy to grasp, is hard to define mathematically due to the various definitions it has been given in the past, and the confusion in terminology that still exist nowadays (e.g species richness vs. species diversity). Diversity is usually defined at different spatial scales. Whittaker (1960) first introduced this concept as he recognized that total species diversity in a landscape could be considered to consist of conceptually different components. He used alpha, beta, and gamma to refer to these components. In this way, *Gamma diversity* is the total species diversity in a landscape or ecosystem. It is composed of the local species diversity (measured in spatially limited and homogenous samples at the habitat-scale), called *Alpha diversity*, and the compositional differences among these local systems, called *Beta diversity*.

Consequently, when measuring biological diversity in natural samples, one has to be aware of three central elements (Purvis and Hector 2000). First, *species richness*, or the number of species in a sample, is the simplest and most intuitive component of biodiversity. It however assumes that species definition and classification is well known (which is a matter of debate, see below!) and that all species are equivalent (each species weighs equally in the richness value, regardless its abundance). Simple counts of species in samples underestimate true species

richness and strongly depend on sampling effort, as it is hardly possible to sample all species in an ecosystem. Thus, there are potentially numerous undiscovered species in any species inventory. Generally, two approaches can be used to estimate species richness in incomplete samples: an asymptotic approach via species richness estimators (Fisher's alpha, Chao estimators, Jackknife estimators, Coverage-based estimators) or a non-asymptotic approach via rarefaction and extrapolation (Chao and Chiu 2016).

Second, sample *evenness* or *equitability* estimates the relative abundance of species in a sample, i.e. the extent to which individuals are distributed evenly among species. Indeed, individuals from a very abundant species contribute less to biodiversity than individuals from a rare species. Sample diversity is thus higher when individuals are distributed evenly among species. Equitability estimators therefore evaluate the deviation of the observed species distribution from a uniform distribution, and the Pielou index, J, is the most widely used equitability index (Purvis and Hector 2000).

Ideally, *alpha* (local) diversity should be a measure of both species richness and species relative abundances. Commonly used alpha diversity indices such as the Shannon or Simpson indices, evaluate both richness and equitability. It is important to note that alpha diversity indices mainly differ in the weight they give to abundant vs. rare species, for e.g. the Simpson index is more sensitive to dominant species. As they quantify diversity in different ways, they are thus only a *proxy* of the variable they try to quantify. Modern diversity estimators have unified classical indices, using measures of entropy that can be expressed as an "effective number of species", i.e. Hill numbers of order q (Grabchak et al. 2017; Chao et al. 2014). These measures only depend on q, the exponent of the species frequencies in the index. This q, is what determines the sensitivity of the index to the species frequencies (typically $q=0$-2). In this framework, when $q=0$, diversity is species richness and the same importance to all species, thus greatest possible weight is given to rare species. When $q=1$, diversity is the Shannon index, all individuals weigh the same, thus species weigh differently depending on their relative abundance. When $q=2$, diversity becomes the Simpson index, which gives less weight to rare species, explaining why it is usually called "the number of very abundant species" (Tuomisto 2010).

These classical diversity measures are so-called species-neutral diversity measures, as they do not consider interspecific distances. However, two species of a same genus are obviously more related than two species of different families, and some species assemblages gather species with similar or very distinct functions in the community. Phylogenetic and functional

diversity measures take into account the phylogenetic and/or phenotypic differences by evaluating *disparity,* the mean divergence between species. The last component of biodiversity thus measures the extent to which species are different, giving insights into evolutionary history of the community or allowing evaluating ecosystem productivity, functioning, and resilience.

Disparity can also be measured among samples to describe the degree of compositional differentiation of communities according to changes in the environment, i.e. *beta diversity*. Ecologists usually look at community differentiation in two possible ways: turnover by reference to a specific gradient, or variation in community structure (Anderson et al. 2011). When looking at turnover, the idea is to measure the change in community structure from one sampling unit to another, along a spatial, temporal, or environmental *predefined* gradient. The change can be measured via species identities, relative abundances, biomass, or percentage cover of individuals. Overall, turnover can always be expressed as a rate, as in a distance-decay plot. When looking at variation, the idea is to evaluate community structure among a set of sample units within a given category (space, time, habitat type, experimental treatment…). Variation is measured among all possible pairs of units, without a reference to a gradient, and it quantifies the proportion of unshared species among all sampling units.

Pairwise dissimilarities between sampling units form the basis of the analysis of *beta diversity*. The different approaches mentioned above (turnover-based vs. variation-based studies) differ essentially by the type of distance metric calculated. For turnover-based studies, pairwise, non-Euclidean distances are usually computed, and analysed with respect to the predefined gradient. Linear or non-linear models (regressions or distance-decay models) consequently allow evaluating the turnover or rate of turnover along the chosen gradient. In variation-based studies, the gradient is unknown and the variation, i.e. pairwise dissimilarities, visualized in ordinations (i.e. 2-D representations of dissimilarity). Ordinations can be unconstrained or constrained. In unconstrained ordinations, one does not impose the nature of the ordination axes and only associates environmental variables by *a posteriori* superimposing environmental labels or vectors. This is called indirect gradient analysis. In constrained ordinations, also called direct gradient analysis, one tries to partition variation according to some factors or continuous environmental variable, resulting in the fact that the ordination axes are defined (constrained). This approach however requires good knowledge of the studied samples, or a predefined working hypothesis. There are numerous ordination techniques that differ essentially in the type of distance (Euclidean or non-Euclidean distances) they use to calculate pairwise differences (Legendre and Cáceres, De 2013). Anderson et al. (2011)

emphasize the importance of comparing ordinations based on different dissimilarity measures, as they correspond to different underlying ecological hypotheses. More importantly, it is crucial to distinguish between incidence-based and abundance-based measures, as dramatically different results can be obtained when relative abundance data are included. Second, inclusion vs. exclusion of joint-absence information (number of species absent in both compared units) is also to be considered, as this information can be relevant when studying the disappearance of species following an environmental impact, predation, or biological invasion.

Overall, making an informed choice of the indices to use for a particular study necessitates an understanding of which aspects of diversity the indices quantify, and what is needed to answer a specific ecological question. There is no universal measure, and it should always be kept in mind that most indices are only a *proxy* of diversity itself. Moreover, in specific applications such as environmental impact or conservation studies, other measures than the "generalist" summarized here are likely more appropriate. For example, biotic indices based on indicator organisms are commonly used in impact assessments (Washington 1984; Aylagas et al. 2014; 2018), while a measure of endemicity may be more useful for conservation management efforts (Costello and Chaudhary 2017).

The species concept

The Linnaean project, initiated some 265 years ago, is arguably the longest-running, most successful, and most impactful biology megaproject of all (Blaxter 2016). Carl Linnaeus proposed a binomial system to "name" groups of organisms that are recognizable as distinct natural types (i.e. species), therefore allowing to communicate complex concepts across the globe. The *species* constitutes the core of biodiversity inventories for biological and ecological studies, and helps organizing agriculture, trade, and industry (e.g. species used for the production of biomaterial) as well as measuring the impact of human activity on the Earth's ecosystems (e.g. biomarker taxa and pathogenic or invasive species).

While biotic diversity can be valued and assessed at various levels, including that of the individual organism and the genetic locus, the key level remains the species, and some authors conclude that species richness, while not perfect, is the best metric (Freudenstein et al. 2017). A main issue in answering the question of *how many species* is that it requires a consensus on the definition of "species". Before Darwin, the delimitation of species focused largely on phenotypic uniqueness, i.e. common morphology. Darwin (1859) added *history* into the species definition, by depicting connections through time between species and their "offspring". Within

an evolutionary framework, the possibility arose of delimiting species, at least conceptually, by their unique history, and the word *lineage* became common to speak of populations and species through time.

In the mid-20th century, evolutionary synthesists such as Mayr and Simpson recognised populations (groups of similar individuals) as fundamental units in nature, and outlined that it is the relationship among populations, i.e. interbreeding, that is critical in the concept of species (Freudenstein et al. 2017). Mayr (1942) developed his Biological Species Concept (BSC), and defined species as "*groups of actually or potentially interbreeding natural populations, which are reproductively isolated from other such groups*". Simpson (1951) formalized his Evolutionary Species Concept (ESC) as "*a phyletic lineage (ancestral- descendent sequence of interbreeding populations) evolving independently of others, with its own separate and unitary evolutionary role and tendencies*".

Recently, specialists from the eukaryotic (Freudenstein et al. 2017; Costello and Chaudhary 2017), micro-eukaryotic (Fenchel 2005), and prokaryotic (Rosselló-Mora and Amann 2001) worlds have stressed the importance of *role* in the second part of Simpson's sentence. Indeed, while there has been an increasing trend toward viewing species only as historical lineages (Freudenstein et al. 2017), these authors argue that this contradicts the original species concept and misaligns with the key position of species as units of biodiversity. They re-define the ESC as "*A species is a lineage or group of connected lineages with a distinct role.*", or call it the "pheno-phyletic" or "phylo-phenetic" species concept.

Other species definitions gave similar emphasis on role, like Van Valen's (1976) Ecological Species Concept ("*a lineage, or a closely related set of lineages, which occupies an adaptive zone minimally different from that of any other lineage in its range and which evolves separately from all lineages outside its range*"), or Levin's (2000) ecogenetic concept that considered ecological function as part of species definition. Cohan's bacterial "ecotype" is also similar to this view, as ecotypes are necessarily ecologically distinct (Cohan 2001; 2002). Later in his career, even Mayr (1982) came to view role as critical with his revised definition of species as "*a reproductive community of populations (reproductively isolated from others) that occupies a specific niche in nature.*"

Thus, it seems that there is increasing recognition that ecological function should be part of species definition. Simpson described role as "definable by their equivalence to niches" (Freudenstein et al. 2017), although this has encountered criticism as ecological niches are difficult to precisely delimitate (Hengeveld 1988). Freudenstein *et al.* (2017) view role broadly

as "*the ways in which individuals interact with their environment and the total complement of expressed properties (beyond genotype) that they exhibit*", and call it an "*extended phenotype*". What does this extended phenotype encompass? Most intuitively, morphological and physiological features, as those are related to ecological role (Simpson, 1961). Phenotypic change is related to genotypic change, but the latter are not strictly linked, as epigenetic manipulation of the genome and extra-genomic determinants (ecological, cultural, parental inheritance) of phenotypical characteristics have been described (Danchin et al. 2011). It is thus apparent that phenotypic as well as genotypic features have to be considered to determine role during species description and detection. Thus, although based on the ESC, methods describing biodiversity solely through genetic *proxies* are inherently limited as they do not encompass the complete "*extended phenotype*".

Global biodiversity estimates, and their limitations

How many species?

As there is yet no consensus on the definition of "species" for prokaryotes, the single-celled but extremely diverse Archaea and Bacteria (Konstantinidis et al. 2006), the question of *how many* has mostly been asked for eukaryotes, i.e. protists, plants, fungi, and animals (Pimm 2012).

So far, expert opinions and extrapolations from macro-ecological patterns or from species description rates have been the main approaches used to estimate total species richness both in the terrestrial and marine realms (Table 1). At the beginning of the 1990s, global species richness estimates were hardly more than "*educated guesses"*, ranging from around three to over fifty million with no associated estimates of uncertainty (Mora et al. 2011). As of today, species richness estimates have not converged (Table 1), and range from ~2.0 million to 76.5 million eukaryotic species globally (Caley et al. 2014; Costello et al. 2012; Vargas et al. 2015). Given that ~1.2 million eukaryotes have been catalogued so far (Mora et al. 2011) and that some authors predicted up to 1 trillion (10^{12}) species of prokaryotes on Earth (Locey and

Lennon 2016), it is clear that biodiversity research has still a long way to go, and many uncertainties to clarify.

Table 1. An overview of biodiversity estimates and utilized methods in the past three decades. Adapted from Mora *et al.* (2011) and Appeltans *et al.* (2012).

Number of species estimated	**Method**	**Reference (year)**
5 - >50 million globally	Extrapolation from the frequency of large to small species	May RM (1988)
3-5 million	Ratio of numbers of tropical to temperate and boreal species	May (1990)
5-15 million globally	Analysis of available global estimates	Stork N (1993)
12.5 million	Compilation and extrapolation from regional estimates	Hammond (1992)
> 10 million marine species	Extrapolation of deep-sea benthos samples	Grassle & Maciolek (1992)
1.4-1.6 million marine species	Extrapolation from proportion of brachyurans in Europe	Bouchet (2006)
1-1.4 million marine species	Census of Marine Life, extrapolation based on regional estimates	Costello et al. (2010)
8.7 ± 1.3 million globally of which 2.2 ± 0.8 are marine	Extrapolation of number of species based on patterns in higher taxonomic levels	Mora et al. (2011)
1.8-2.0 million species globally of which 0.3 million are marine	Prediction based on description rates	Costello et al. (2012)
0.5 ± 0.2 million marine species	Prediction based on description rate	Appeltans et al. (2012)
0.7-1.0 million marine species	Prediction based on expert opinions	Appeltans et al. (2012)
16.5 million terrestrial and 60 million marine eukaryotic species	Extrapolation of Mora et al.'s estimates based on metabarcoding-revealed protistan knowledge gap	De Vargas et al. (2015)
Up to 1 trillion (10^12) microbial species	Universal dominance scaling law	Locey & Lennon (2016)

Cataloguing biodiversity requires extraordinary knowledge of individual taxa, and thus extraordinary amounts of taxonomic experts (Pimm 2012). Although there is growing concern of declining taxonomic expertise, evidence shows that the numbers of scientists describing new

species, taxonomic publications, and species discovery rates have been increasing in the past decades (Costello and Chaudhary 2017; Costello et al. 2012; Appeltans et al. 2012). Current catalogues of biodiversity are available for terrestrial and marine realms (e.g. Catalogue of Life, and WoRMS, the World Register of Marine Species). Major sources of uncertainty in these catalogues are fourfold: 1) frequent occurrence of synonyms may inflate diversity estimates; 2) significant amounts of cryptic diversity may have been undetected by morphology-based approaches; 3) the potential hyper-diversity of small organisms may have been overlooked; and 4) under-sampled habitats such as the deep-sea may harbour large amounts of unknown biodiversity.

The synonyms

A potential large amount of extant named species may be synonyms, i.e. duplicate names for the same biological entity (Alroy 2002). Disconcertingly high proportions of synonyms have been reported for marine species (40%), but also for terrestrial animals (e.g. 31% of insect species), plants (78%), or freshwater fish (81%). Overall, synonyms may inflate current catalogues by about 20%, thus future species discoveries will be balanced by recognition of synonyms (Costello and Chaudhary 2017).

The cryptic species

Cryptic species refer to species that can only be differentiated by genetic but not by morphological features (Costello & Chaudhary 2017). Advances in DNA analysis have revealed high levels of cryptic diversity across the tree of life, likely driven by habitat heterogeneity and fragmentation (Poulin and Pérez-Ponce de León 2017). However, conflating cryptic diversity and cryptic species is misleading, as much of the detected cryptic genetic diversity does not result in formally described species, as this requires the characterisation of biological and ecological traits (Costello and Chaudhary 2017; Pante et al. 2015).

Additionally, genetic markers also have downfalls for species discrimination: levels of mutation rates in mitochondrial DNA (mtDNA) sequences differ across animal species (Galtier et al. 2009; Nabholz et al. 2008), and some basal metazoan lineages exhibit such low rates of evolution that species cannot be distinguished on the basis of mtDNA sequences (Shearer et al. 2002; Shearer and Coffroth 2008; Huang et al. 2008). MtDNA diversity is thus highly variable among taxonomic groups, not consistently correlated with population size (Bazin et al. 2006; Mulligan et al. 2006), depends on life-span (Nabholz et al. 2009), and is affected by bacterial

symbionts (Hurst and Jiggins 2005). Thus, the detection of cryptic species cannot solely rely on molecular proxies, as these need to be combined with morphological and ecological analyses to establish accurate species boundaries. Nevertheless, true cryptic species, i.e. only discernible through genetic data, do exist, and may balance the decrease in the number of catalogued species resulting from the discovery of synonyms (Costello and Chaudhary 2017).

The small

A major source of uncertainty in the current catalogue of diversity is related to the fact that it was built on three centuries of morphological information and thus exhibits a strong bias towards large organisms. The record for vertebrates may be close to complete, but this is likely not the case for taxa with smaller body size (Blaxter 2003; Cristescu 2014). Indeed, it has now become clear that the overwhelming majority of organisms are microscopic, i.e. smaller than 1 mm (Bacteria, Archaea, protists, but also most Metazoa). Lack of easily recognized morphological characters, incompleteness of early descriptions, phenotypic plasticity, and the high numbers of organisms compared to the relatively few numbers of taxonomists are all factors suggesting that current inventories may underestimate microscopic biodiversity (Blaxter 2016). Recent estimates suggest the presence of 10 million insect species globally (Hebert et al. 2016b), and possibly over 1 million species of nematodes (Blaxter 2016). Similarly, DNA-based studies in the marine realm showed an unprecedented eukaryotic genetic diversity in planktonic and benthic environments, emphasizing the protistan knowledge gap, and suggesting that eukaryotic diversity may increase with decreasing body size. In the Tara Oceans expedition (https://oceans.taraexpeditions.org/en/m/about-tara/), protists accounted for over 85% of the diversity, raising previous biodiversity estimates to 16.5 million terrestrial and 60 million marine eukaryotic species (Vargas et al. 2015). Similarly, the BioMarks project on benthic diversity in European coastal waters concluded that 30%-70% of protists remain to be discovered (Forster et al. 2016). Finally, the latest estimate of bacterial diversity based on high-throughput molecular data predicted up to 1 trillion (10^{12}) species of prokaryotes on Earth (Locey and Lennon 2016).

These DNA-based approaches thus suggest an enormous amount of undescribed microbial species. However, high levels of alpha diversity do not imply high global (gamma) diversity. Studies have shown that small organisms, while exhibiting high local species richness, display decreasing diversity at larger spatial scales (Azovsky 2002). Biodiversity patterns at the microscopic scale differ markedly from those at the macroscopic scale: small species are often

found to be cosmopolitan, i.e. thriving wherever local habitat conditions are suitable (Finlay et al. 2004), to exhibit high and random dispersal, asexual reproduction and increased horizontal gene transfer, as well as short generation times combined with large population sizes (Finlay 2002). Together, these characteristics support high genetic diversity, enabling rapid adaptation to changing environmental conditions, but also lower speciation rates due to higher gene flow (Costello and Chaudhary 2017). Indeed, speciation rates have actually been found to be higher in multicellular eukaryotes compared to prokaryotes (Lynch and Conery 2003). Thus, while high levels of genetic diversity may be detected in small organisms, these do not necessarily point toward high species diversity (Rossberg et al. 2013). Indeed, studies in vertebrates and plants have found no correlation between genetic and species richness (Costello and Chaudhary 2017). Moreover, it has been found that body size does not predict species richness in the Metazoa (Orme et al. 2002), highlighting that small does not necessarily mean species rich. Nevertheless, morphology-based investigations showed that 37% of meiofauna species (42 µm - 1 mm size ranges) sampled in a well-known ecosystem (Western Mediterranean shallow water) were new to science, indicating that much of the small diversity remains to be described (Curini-Galletti et al. 2012).

The under-sampled

Another bias in the catalogue of diversity is related to the fact that humans are terrestrial animals and have therefore more extensively explored terrestrial habitats compared to aquatic ones. Consequently, biodiversity research has mostly been focused on terrestrial, usually temperate fauna, and targeted mostly mammals, birds, and arthropods (Table 1; Stork 1993; May 1988; Hendriks and Duarte 2008; Hendriks et al. 2006). However, of the 36 animal phyla described today, all but one are found in the marine environment, and 40% are exclusively marine (Pimm 2012). This highlights the extreme breadth of oceanic biodiversity, but also the fact that terrestrial species have evolved from marine ones (Costello and Chaudhary 2017). The last decades were marked by the will to elucidate marine diversity and initiatives like the Census of Marine Life and the World Register of Marine Species (WoRMS) greatly improved our knowledge of the marine realm (Table 1). More marine biodiversity research was performed in the last 60 years than never before, yet, only 16% of described species are marine. This low proportion of marine species may be due to under-sampling and still disproportionally small research efforts (particularly in the deep-sea; Hendriks et al. 2006), or to the biological reality

that diversity is higher on land than in the sea, due to higher productivity, higher habitat complexity, and thus more ecological niches (Costello and Chaudhary 2017).

I.II. Estimating biodiversity in the 21st century: the revolution of DNA-based taxon identification approaches

Species identification with DNA barcodes

Since the description of Deoxyribonucleic Acid (DNA) in 1953, significant advances in molecular biology have allowed researchers to develop techniques to exponentially amplify DNA molecules, so-called Polymerase Chain Reaction (PCR), and determine the order of their nucleotides (their building blocks), a process known as DNA sequencing. During the late 1990s, microbiologists used these advances to survey the diversity of bacteria and archaea using the 16S small subunit (SSU) ribosomal RNA (16S rRNA) gene, showing that prokaryote diversity was at least 100 times higher than previously expected (Blaxter 2003). Microbiologists soon used sequence data for "*species*" descriptions, the bacterial taxa being defined as phylotypes or "Molecular Operational Taxonomic Units" (MOTUs or OTUs). These DNA-based diversity estimation methods were built on the observation that there generally is a gap between the distributions of intraspecific and interspecific divergence in gene sequences, termed the *barcode gap* by Meyer and Paulay (2005, Fig. 1). The term "barcode" is a figurative analogy to commercial barcodes found on price tags, where the width and spacing among parallel lines identify products. Similarly, the sequences of nucleotides (Adenine (A), Thymine (T), Guanine (G), and Cytosine (C)) in barcode genes is taxon specific. Consequently, DNA sequences of barcode genes enable species identification and recognition, while complementing formal species description by providing molecular diagnostic characters (Hebert et al. 2003a; Bucklin et al. 2011).

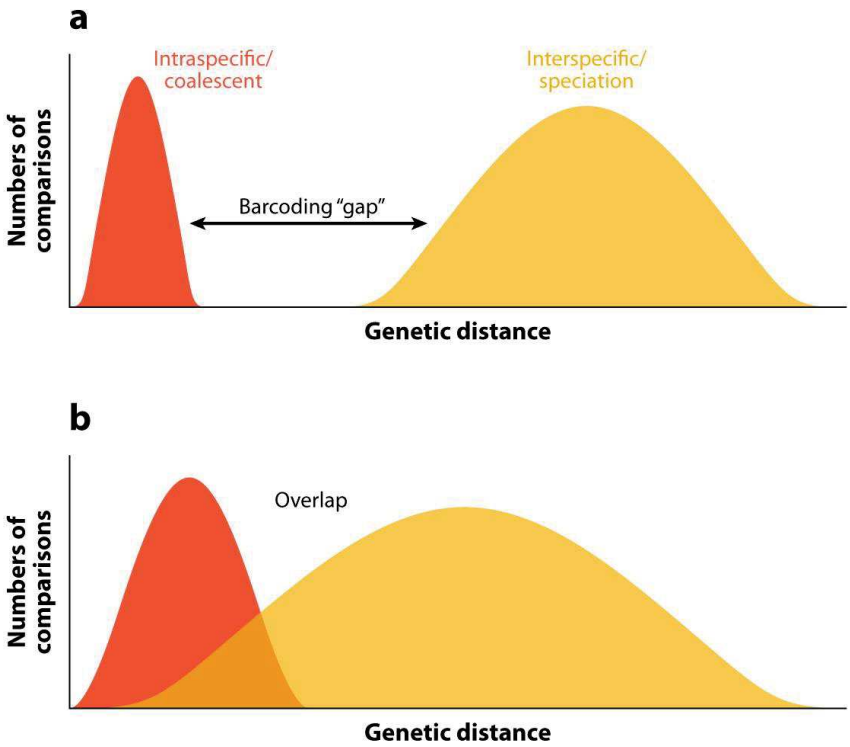


Figure 1.Schematic representation of the barcoding gap.
Frequency distributions of genetic distances within (red) and between (yellow) species. (a) Ideal world for barcoding, with discrete distributions of intraspecific and interspecific variation and no overlap. (b) A common situation with significant overlap between intra- and interspecific variation and no barcode gap. From Bucklin et al. (2011).

In 2003, Hebert et al. proposed to extend this approach to eukaryotes and suggested that the mitochondrially encoded Cytochrome Oxidase I gene (COI) could serve as a DNA barcode for all animal taxa (Hebert et al. 2003b). Since then, much work has been undertaken for determining standardized DNA barcodes for all domains of life. An ideal barcode gene should have three main characteristics. First, as PCR amplification depends on primers, short DNA fragments that bind to the DNA to be amplified, a barcode gene should possess conserved flanking sites to allow successful primer binding across broad taxonomic levels, thus avoiding the non-detection of taxa due to unsuccessful primer binding (primer bias). Second, it should possess a strong enough phylogenetic signal, i.e. have mutation rates (and thus intraspecific variation) that allow discrimination of closely related taxonomic groups (ideally species). Finally it should display a *barcode gap* (Fig. 1), i.e. marked divergence and no overlap between intra- and interspecific genetic distances (Bucklin et al. 2011).

For animals (metazoans), the mitochondrial genome has several advantages over the nuclear genome, such as lack of introns, mostly uniparental (maternal) inheritance and thus

little recombination, and predominantly neutral evolution, allowing it to serve as a "molecular clock". These features combine with the presence of high copy numbers in every cell, making amplification more successful , and with the presence of conserved regions allowing the design of "universal primers" that amplify a broad range of taxa (Folmer et al. 1994; Geller et al. 2013). Moreover, due to elevated mutation rates, the mitochondrial COI gene offers the best species-level resolution in most taxa except for ctenophores, sponges, nematodes, and some benthic cnidarians (corals and anemones), for which COI is either difficult to amplify or not resolutive enough (Bucklin et al. 2011; Blaxter 2016). However, these advantages are not universally valid (Galtier et al. 2009), and studies have reported a lack of conserved regions leading to considerable taxonomic bias during PCR (Deagle et al. 2014), or the presence of nuclear mitochondrial pseudogenes (NUMTS) leading to considerable overestimation of biodiversity (Song et al. 2008). Consequently, various variable regions of the 18S SSU ribosomal RNA (rRNA) gene have been increasingly used as barcodes (18S V1-3, V4-5, V7, or V9), particularly in taxa for which COI is difficult to amplify (including unicellular eukaryotes). However, as rRNA genes evolve more slowly than protein-coding genes, they tend to provide less taxonomic resolution, leading to the potential underestimation of diversity (Tang et al. 2012). Among the variable regions used in 18S barcoding, the V1-3 region was found to show greatest sequence variability and thus highest taxonomic identification power, although currently under-represented in taxonomic databases (Tanabe et al. 2016).

For plants, two plastid genes (*matK* and *rbcL*) were selected as core barcodes, supplemented by more variable barcodes from non-coding regions (plastid inter-genic spacer, or nuclear ribosomal internal transcribed spacer) to allow more precise differentiation at lower taxonomic levels. Similarly, the nuclear ribosomal internal transcribed spacer (ITS) is the standard DNA barcode for fungi, but secondary markers are being looked for, as this region is sometimes too variable for robust species-level identification (Hebert et al. 2016a; Bellemain et al. 2010).

Overall, as there is no unique, ideal, and universal barcode gene, it is thus widely recommended to use sequences from multiple loci (Bucklin et al. 2011; Cowart et al. 2015). The great taxonomic coverage but low species-level resolution of slowly evolving genes, such as rRNA genes, are well complemented by mtDNA or plastid loci that allow deeper taxonomic identification (Hebert et al. 2016a).

As research in barcode genes intensified, so did the effort in developing well curated sequence databases, which are essential for the taxonomic identification of sequence data.

Consequently, efforts to establish large public databases have been considerable in the past ten years, and have led to the development of BOLD (Barcode Of Life Database), compiling standard barcode sequences for animal COI, fungal, and plant sequence data. Moreover, curated and taxon-specific mitochondrial and ribosomal references databases for prokaryotes and eukaryotes have emerged, with most notable examples being MIDORI for COI, SILVA, GreenGenes, and PR2 (Protist Ribosomal Reference Database) for ribosomal DNA (rDNA), and ITS2 (Internal Transcribed Spacer 2 Ribosomal DNA database) for internal transcribed spacer sequences.

From barcoding of single species to metabarcoding of whole communities

Since 2005, the development of high-throughput sequencing (HTS) technologies, has allowed producing millions of DNA sequences from individual samples. This high throughput allows reliable, rapid, and inexpensive analysis of community samples, representing a new generation of sequencing technologies that are becoming increasingly available for the investigation of biodiversity at inter- and intraspecific scales. HTS can be applied to marker gene analysis (i.e. metabarcoding), allowing the description of biodiversity at the species-level, while total DNA approaches (e.g. shotgun sequencing, RNA sequencing, restriction site-associated DNA sequencing, coined RAD seq), are effective tools for resolving individual and population-level genetic diversity. Molecular biodiversity assessment can be performed from bulk DNA extracted from a collection of organisms, approach termed DNA metabarcoding, or from environmental samples where DNA is extracted directly from air, water, or soil samples, termed environmental DNA (eDNA) metabarcoding. Metabarcoding studies mainly differ in (1) the type of barcode gene (genetic marker) used, (2) the precision of the taxonomic identification they allow considering the reference databases available and genetic marker used, and (3) the level of degradation of the DNA extract, determining the length of the barcode region to be used (Taberlet et al. 2012a).

Figure 2 illustrates a typical workflow of high-throughput metabarcoding studies based on eDNA, allowing the estimation of alpha and beta diversity, taxonomic community profiling, but also connectivity studies (through OTU networks), or coalescence analyses (via phylogenetic reconstructions of the marker genes).

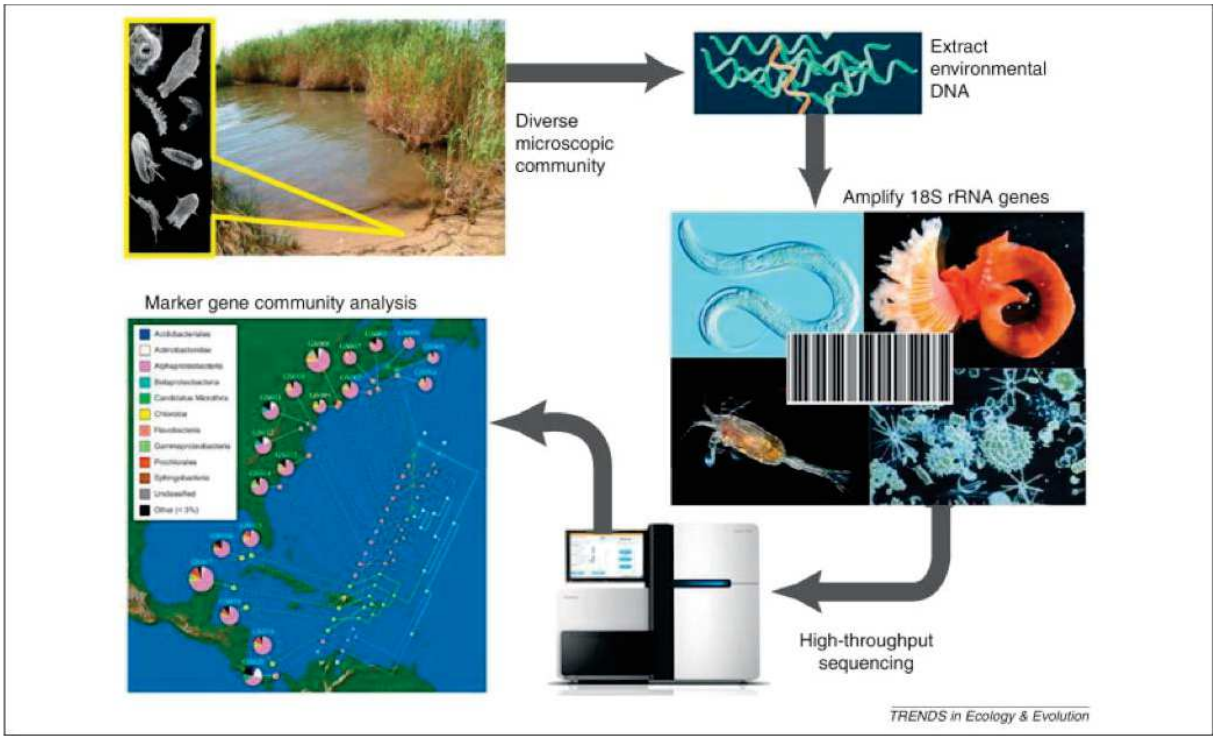


Figure 2. Typical workflow for high-throughput eDNA metabarcoding studies. Environmental samples such as sediments are usually frozen upon collection (–80°C to preserve RNA) and brought back to the lab for extraction of eDNA. Marker genes (e.g. 16S, 18S rRNA, COI) are amplified from genomic extracts using primer pairs targeting a wide variety of taxa. Following high-throughput sequencing (typically conducted on Illumina platforms), sequences are bioinformatically processed, and molecular entities defined. The latter are then taxonomically assigned, and used to conduct α- and β-diversity analyses, summarize community taxonomy, and interpret assemblages in a phylogenetic context. From Bik et al. 2012a.

Challenges and uncertainties of metabarcoding approaches

After PCR amplification of barcode fragments, DNA amplicon libraries can be prepared in numerous ways for HTS, all generally involving the ligation of sequencing platform-specific adapters, sample-specific indexes, DNA purification, and pooling of libraries at equal concentration for multiplexed sequencing. Following HTS, typically conducted on Illumina platforms, the user is confronted with tens to hundreds of millions of raw sequences that need to be bioinformatically processed to produce a list of putative taxa. The bioinformatic analysis of metabarcoding data has evolved a great deal in recent years, with a plethora of algorithms developed for each processing step. Bik et al. (2012a) provide an overview of bioinformatic processing steps and the tools and suites available, but many other algorithms and pipelines were made available in the last years such as USEARCH (Edgar 2010) , VSEARCH (Rognes

et al. 2016), OBITOOLS (Boyer et al. 2016), DADA2 (Callahan et al. 2016), FROGS (Escudié et al. 2018), or the web-application SLIM (Dufresne et al. 2019).

First, bioinformatic processing usually includes various quality-filtering steps, where primers, sample tags, and sequencing adaptors are removed from raw sequences (Fig 3). These are then trimmed to remove low-quality ends and/or quality-filtered (based on nucleotide Quality-scores or error rates). Next, a key step is to decide on the molecular entity that will serve as a *proxy* for taxa in the dataset. This can result from either grouping (clustering) processed sequences within a user-defined similarity threshold, resulting in Operational Taxonomic Units (OTUs), or denoising sequences, resulting in Amplicon Sequence Variants (ASVs in Callahan et al. 2017), also called ZOTUs in (Edgar 2018d). Illumina sequence correction algorithms such as Deblur (Amir et al. 2017), UNOISE2 (Edgar 2016c), or DADA2 (Callahan et al. 2016) are relatively new, but are increasingly popular as they effectively remove sequencing errors by applying a data-based and quality-aware correction algorithm.

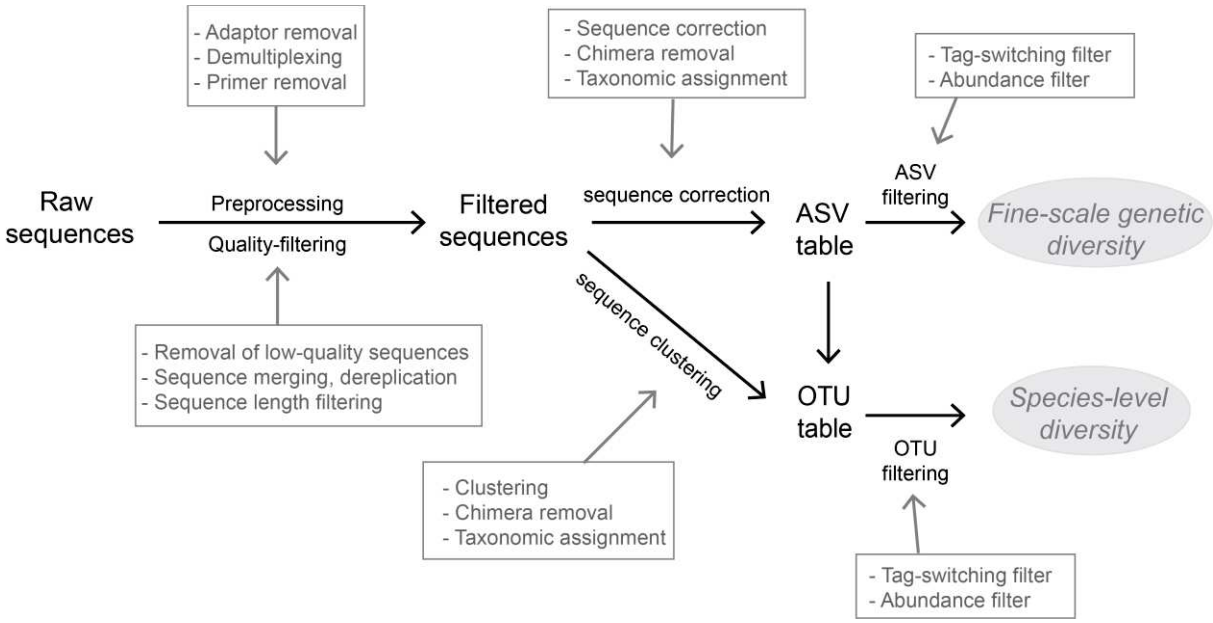


Figure 3. Typical bioinformatic processing steps for analysis of metabarcoding data. ASV: Amplicons Sequence Variants, OTU: Operational Taxonomic Unit.

Biological biases in eDNA metabarcoding: dead or live biodiversity?

Numerous biological biases affect the number and abundance of molecular clusters retrieved by a metabarcoding analysis (Fig. 4). First and most intuitively, the size, biomass, and spatial distributions of organisms will affect their detection rate. Genetically, the characteristics

of the gene region chosen as barcode will influence amplification success and resulting cluster abundances. In addition, eDNA is a complex mixture of genomic DNA present in living or inactive cells, extra-organismal (e.g., organelle) DNA, and extracellular DNA originating from the degradation of organic material and biological secretions (Torti et al. 2015). Extracellular DNA is very abundant in the environment, for e.g. it has been shown to represent 50-90% of the total DNA pool in marine sediments (Corinaldesi, Tangherlini, Manea, & Dell'Anno, 2018; Dell'Anno & Danovaro, 2005). However, extra-organismal and extracellular DNA may not only comprise DNA from contemporary communities, as DNA can persist in the environment due to adsorption onto clay particles, low temperatures, high salt concentrations, or the absence of UV light (Torti et al. 2015; Nagler et al. 2018). Up to 125,000-year-old ancient DNA (aDNA) has been reported in oxic and anoxic marine sediments at various depths (Boere et al. 2011; Lejzerowicz et al. 2013a; Coolen et al. 2013). Ancient DNA may thus bias eDNA metabarcoding biodiversity inventories towards describing past, rather than present communities, particularly in environments known to favour DNA persistence such as marine sediments. In contrast, aDNA bias will likely not be an issue in studies targeting aquatic environments, as it has been shown that DNA molecules released in the water column degrade rapidly (Dejean et al. 2011; Collins et al. 2018).

Technical biases in the number of molecular entities

Numerous technical biases affect the number of molecular clusters retrieved by a metabarcoding analysis (Fig. 4). First, the taxonomic composition retrieved from metabarcoding data can be biased by the specificity of PCR primers, as primer mismatch can hinder PCR amplification and thus species detection. Taxon detection can also suffer from strong sampling effects due to insufficient sequencing depth, or because DNA extractions, typically performed on small amounts of material, make large organisms not necessarily well represented in eDNA extracts (Creer et al. 2016; Cordier et al. 2019b).

Several studies have shown that spurious clusters are a serious issue in molecular biodiversity inventories, and highlighted the need for stringent quality filtering steps and/or clustering programmes in order to avoid overestimation of the number of OTUs/ASVs, and approach a 1:1 correspondence with species sampled *in situ* (Clare et al. 2016; Edgar 2013; Bokulich et al. 2013). Although sequence-denoising algorithms effectively remove sequencing errors, they do not remove errors originating from PCR amplification or tag-switching. Chimeras, DNA artefacts generated during PCR and derived from the mixture of two or more

template molecules due to incomplete extension in previous PCR cycles, are a well-known problem leading to the over-inflation of cluster richness, and many tools have been developed to detect and remove them during bioinformatic processing (Edgar et al. 2011; Bik et al. 2012a). In addition, HTS sequencing is performed on pooled equimolar sample libraries. Tag-switching (also called cross-talk), i.e. the assignment of sequences to the wrong sample, is a common phenomenon in these multiplexed sequencing libraries, and can cause a substantial amount of false positives (Schnell et al. 2015). The problem is particularly severe if samples from different origins but similar ecosystems are multiplexed in the same sequencing run. It is thus essential to implement a "tag-switching filter" during bioinformatic processing (Fig. 3). Although not often used in practice, such filters have been developed and are usually based on OTU filtering based on cumulative frequency (Edgar 2016b; 2018a; Wangensteen and Turon 2016).

Even after all these filtering steps, many OTU/ASV table entries are singletons (i.e., have total abundance of 1), or comprise clusters with low sequence ("read") counts. Small counts are more likely to be spurious, especially singletons, either because the OTU/ASV itself is spurious (e.g., an undetected chimera), or because of tag switching. It is thus current practice to remove singletons and filter molecular clusters based on their relative abundance per sample or in the total dataset (Wangensteen and Turon 2016). These minimal abundance filters (Fig. 3) have to be chosen with caution as they significantly affect qualitative detection measures. To avoid the arbitrary filtering based on relative abundance, Frøslev et al. (2017) have developed an alternative curation algorithm that filters OTUs/ASVs based on their identity and co-occurrence rates to more abundant OTUs/ASVs. As this tool was developed on plant ITS2 data, it still needs to be adjusted to other taxonomic compartments, as minimum identity thresholds vary among marker genes and taxa. The applicability of LULU to metazoans is one of the goals of chapter 2.

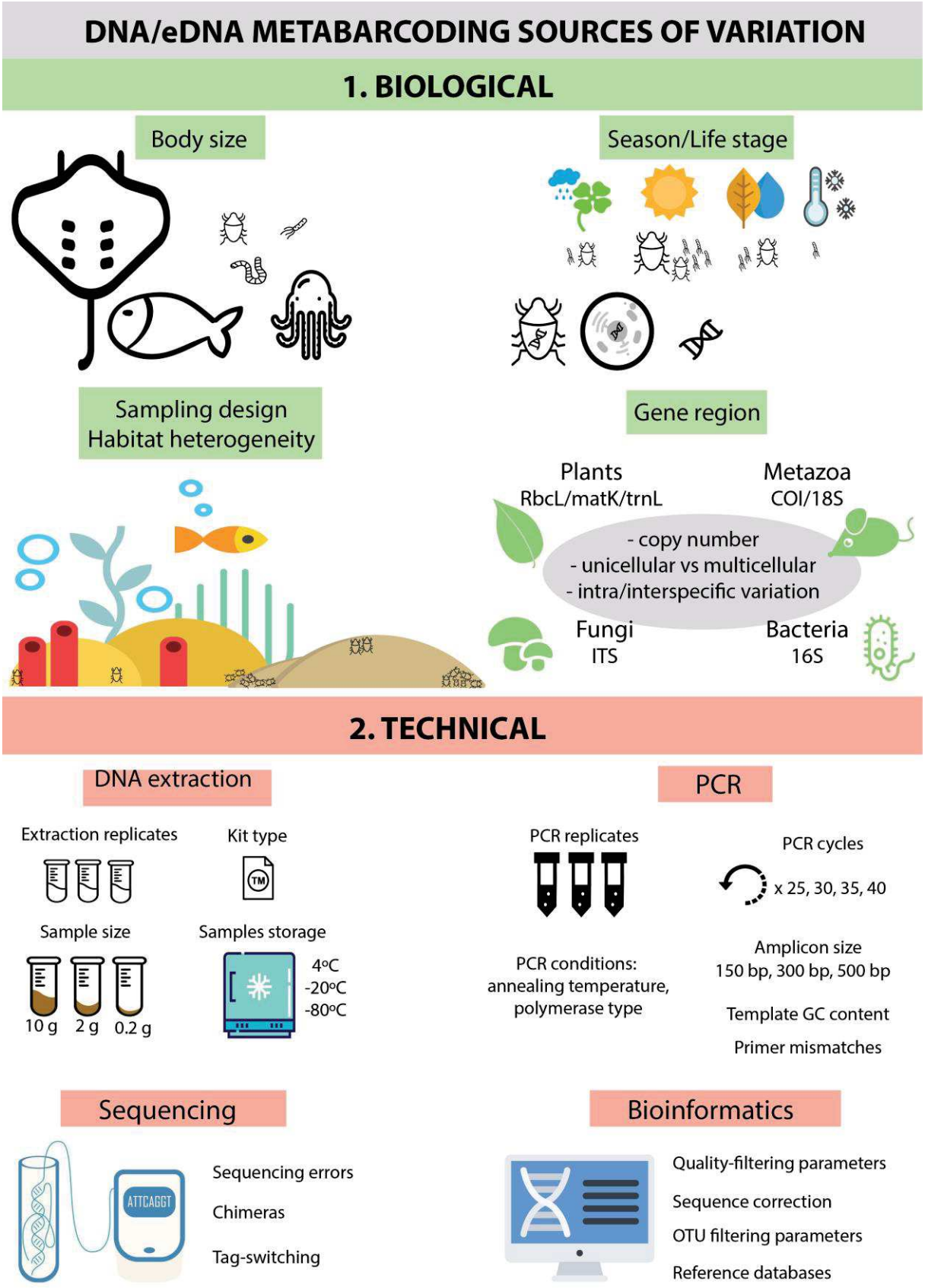


Figure 4. Sources of biological and technical variation in a metabarcoding workflow that can affect the number and the abundance of molecular entities.

Technical biases in cluster abundances

The abundance of sequences in ASVs or OTUs is not only influenced by species abundance, but also the number of copies of the marker gene in the genomes, and by the number of cells for multicellular organisms (Fig. 4). The latter is known to vary widely among eukaryotes (Bik et al. 2013; Weber and Pawlowski 2013) and to a lesser extent for prokaryotes (Klappenbach et al. 2001). Other PCR-related biases also affect the number of sequences produced from each template DNA molecule (Fig. 4). Primer mismatches (decreasing PCR efficiency), unevenness in the oligonucleotide mixture of degenerate primers, template sequence lengths (shorter sequences amplify more efficiently), GC content of the template DNA, type of polymerase used (Fonseca, V. G. 2018; Nichols et al. 2018; Lamb et al. 2019; Pinol et al. 2015) are all factors leading to uneven amplification of template DNA, and as PCR amplification is exponential, this can lead to large biases in read counts.

Studies targeting particular taxonomic groups, such as insects (Piñol et al. 2019; Krehenwinkel et al. 2017) or fishes and amphibians (Pont et al. 2018; Jo et al. 2017; Evans, N T et al. 2016) found correlations between biomass and read abundance using taxon-specific primer pairs. This seems however unlikely to achieve for studies using "universal" (let alone degenerate) primers to encompass the broadest possible range of diversity. Authors have therefore generally concluded that metabarcoding assessments should rely on presence-absence metrics, particularly for metazoans (Lamb et al. 2019; Elbrecht and Leese 2015; Edgar 2017b).

Because of the issues described above, many diversity metrics are invalid, meaningless, or hard to interpret, as neither cluster abundance nor incidence can truly accurately be determined from HTS data. For example, some alpha diversity metrics, like Chao1/Chao2 estimators, explicitly use singleton counts or frequencies in their formulas. When singletons or low abundance clusters are discarded, these calculations are invalid. As singletons are suspect for reasons detailed above, metrics including them are misleading or meaningless.

The above considerations show that it is impossible to measure meaningful and accurate values for *any* diversity metric using HTS data. Diversity metrics can nevertheless be compared among samples analysed through standard sampling, molecular, and bioinformatic pipelines because the errors and biases are mostly *systematic*, i.e. occur in the same way and at the same magnitude in all samples. To ensure that this is truly the case, it is thus crucial to standardize sampling and molecular protocols and to normalise sequencing depth among samples before calculation of biodiversity metrics. This can be done via rarefaction to the lowest sequencing depth or other normalisation methods based on relative abundance. Some authors have

proposed different approaches in visualizing alpha diversity patterns between samples, for example, by extrapolating rarefaction curves (Hsieh et al. 2016), or by visualizing octave plots (Edgar and Flyvbjerg 2018). For beta diversity, shared ASV/OTU presence can be effectively compared with the Jaccard index, a dissimilarity that measures the commonness between samples once double-absences have been removed. In the cases where cluster abundance is considered meaningful, then Bray-Curtis or weighed Jaccard dissimilarities can be computed on relative abundance (or other normalized) data.

Biological interpretation of molecular entities

The relevance of clustering sequences into OTUs is now being discussed as the reproducibility and comparability of ASVs across studies challenge the need for clustering sequences (Callahan et al. 2017; Edgar 2018d). This may be true for prokaryotes, for which optimal clustering thresholds for species definition were found to be >99% (Edgar 2018d). However, it has to be kept in mind that the construction of OTUs, apart from reducing noise due to sequencing or PCR errors, also allows to reduce noise due to intraspecific variation. For metazoans, this is critical, as intraspecific polymorphism is known to be higher than in prokaryotes and varies strongly across taxa and gene regions due to both evolutionary and biological specificity (Bucklin et al. 2011; Phillips et al. 2019). This likely results in very different numbers of ASVs produced among individuals and/or species. Metabarcoding inventories based on ASVs, while accurately resolving fine-scale genetic variation, may thus be biased in favour of taxa with high levels of intraspecific diversity, even though the latter are not necessarily the most abundant ones (Bazin et al. 2006). The biological applicability of ASVs *vs*. OTUs for metazoans is further investigated in chapter 2.

While ASVs may “achieve the best possible phenotype resolution”, this will occur “at the expense of an increased tendency to split species and strains into multiple [ASVs]” (Edgar 2018d) due to cryptic diversity and/or intraspecific diversity. Lumping and/or splitting of species will also occur in OTU datasets, at any clustering threshold. Indeed, OTU clustering thresholds are usually determined based on the barcode gap observed in the marker gene used, i.e. its level of intra vs interspecific divergence. However, there is no consensus on OTU delimitation thresholds, as there is no uniform interspecific divergence threshold across taxonomic groups in barcode genes (Meyer and Paulay 2005; Brown et al. 2015; Candek and Kuntner 2015). Even within a single animal order, there can be large differences in this threshold value between families (Tempestini et al. 2018), highlighting that OTU delimitation

thresholds are data-dependent. Imposing a "universal" clustering threshold on metabarcoding datasets is thus also introducing bias, penalizing groups with lower interspecific divergence, and overestimating species diversity in groups with higher interspecific divergence. However, this can be alleviated with tools such as swarm v2, a single-linkage clustering algorithm (Mahé et al. 2015). Based on network theory, swarm v2 aggregates sequences iteratively and locally around seed sequences and determines coherent groups of sequences, independent of amplicon input order, allowing highly scalable and fine-scale clustering.

Inaccuracy in taxonomic assignments

Although ecological patterns can be investigated without taxonomic identities, species names are useful for inferring biological traits or ecosystem function, as behind each name, there is a phenotype (with all its variability and life forms), an ecological role, and a geographic distribution. Numerous approaches therefore exist to link the detected genetic entities to a Linnaean taxonomy by comparing query sequences to sequences present in reference databases. They include sequence alignment-based (identity-based) methods such as BLAST (Altschul et al. 1990), probabilistic classifiers such as the RDP Bayesian classifier or SINTAX, or phylogenetic (tree-based) assignment methods (Bik et al. 2012a; Edgar 2016a). A study comparing taxonomy prediction algorithms on 16S rRNA and ITS sequences found that alignment-based methods provided similar accuracy than probabilistic methods, although the latter have the advantage of providing a confidence level for each taxonomic rank (Edgar 2018b).

The limitations in taxonomic assignment quality are therefore mostly due to the limited amount of data available (both query and reference) rather than algorithms. Indeed, the assignment accuracy of all these methods is dependent on the quality of the reference database, the database coverage of target groups, the length of the query sequences, and the nature of the marker gene used as barcode (Macheriotou et al. 2019). Although considerable efforts have been undertaken to produce large, public, and curated databases, annotation errors may still be widespread, for e.g., one in five taxonomy annotations in SILVA and Greengenes were found to be wrong (Edgar 2018c). Moreover, many taxa are drastically under-represented in public databases, leading to poor accuracy in taxonomic assignments, especially in studies targeting poorly-known ecosystems (Bik et al. 2012a). Edgar (2018b) highlighted that the length of query sequences and the nature of genetic markers also strongly affect taxonomic accuracy. Longer query sequences provided higher taxonomic accuracy (genus accuracy was ≤50% on 16S V4

sequences, it increased to ~60-70% when using full-length 16S sequences), and more variable loci always provided higher accuracy at lower taxonomic ranks (genus accuracy was close to 90% with ITS sequences). The taxonomic resolution of a study should therefore be adjusted according to the marker gene and study objectives (i.e. which taxonomic ranks are actually needed).

Advantages and applications of eDNA metabarcoding for biodiversity assessments

Despite these limitations, metabarcoding techniques provide several key benefits for achieving comprehensive biodiversity assessments. First, as metabarcoding does not require specimen isolation, it represents a practical and efficient tool in large and hard-to-access ecosystems. For example, it has been successfully applied to study pro- and eukaryote biodiversity in the marine realm, both in the water column (Pernice et al. 2015b; Salazar et al. 2016; Sunagawa et al. 2015; Vargas et al. 2015; Bakker et al. 2017), and on the seafloor, from coastal (Fonseca, V. G. et al. 2010; Cowart et al. 2015; Chariton et al. 2015; Forster et al. 2016) to deep-sea environments (Bik et al. 2012b; Sinniger et al. 2016; Pawlowski et al. 2011; Cordier et al. 2019a).

Another main advantage is the possibility to study the diversity of various biological compartments simultaneously from a single sample by targeting the appropriate barcode genes. Multigene approaches therefore allow the assessment of entire biotic compartments (e.g. zooplankton, benthos), including organisms of various size ranges, providing more comprehensive ecological surveys (Cowart et al. 2015; Drummond et al. 2015; Stefanni et al. 2018; Tedersoo et al. 2016). Because it enables faster community description, metabarcoding has also gained adoption in diverse applied contexts. It is for example increasingly used to identify or detect agricultural pests and pathogens, to detect invasive species, or in the context of wildlife forensics (Hebert et al. 2016a). Moreover, studies have validated its use for assessing environmental impacts (Cordier et al. 2019a; Laroche et al. 2018), and biomonitoring using biotic indices (Aylagas et al. 2014; 2018; Cordier and Pawlowski 2018; Pawlowski et al. 2018) or using bioindicator taxa (Pawlowski et al. 2014; Laroche et al. 2016; Pawlowski et al. 2016b; a).

Finally, evaluating the diversity of life is challenging as the majority of organisms are small (< 1 mm), cryptic, rare, and belong to poorly known groups. Traditional visual inventories remain limited by the difficulty of sampling certain organisms (for e.g. due to behavioural

avoidance), the difficulty of morphologically identifying smaller taxa, and the lack of taxonomic experts (Blaxter 2016; Carugati et al. 2015; Leray and Knowlton 2015). Metabarcoding is thus a very effective approach for detecting diversity of small organisms (bacteria, unicellular eukaryotes, meiofauna), otherwise largely disregarded in visual biodiversity inventories.

I.III. The deep sea, the last frontier on earth

While diversity patterns, their predictors and effects have been relatively well-studied in land-based systems (Loreau et al. 2001), our understanding of global marine diversity and its influence on ecosystem functioning has been limited, although studies have shown strong differences to widely-held terrestrial paradigms (Tittensor et al. 2010; Emmerson et al. 2001; Chaudhary et al. 2016). This contrasts with the fact that most of the world's population is increasingly living in urban areas near the coast (Palmer et al. 2004), bringing marine environments under increased pressure of human activity.

Global studies have shown that virtually no part of the oceans are unaffected by human activity, not even open oceans or deep sea environments, and that up to 41% of ocean areas are heavily impacted by anthropogenic stressors (Halpern et al. 2008; Peng et al. 2020). These stressors range from overfishing, to pollution from land-based or aquatic activities, habitat alteration, or disease spread and biological invasions (Costello et al. 2010b; Ramirez-Llodra et al. 2010; Halpern et al. 2008). Species extinctions in the marine realm have not been as documented as in terrestrial environments. Yet, it has been shown that at regional scales, ecosystems like estuaries, coral reefs, or coastal and oceanic fish communities are rapidly losing populations, species, or entire functional groups (Worm et al. 2006). While marine defaunation seems to be less severe than on land, the current low extinction rates may just be the beginning of a major marine extinction pulse, as the impact of human ocean use grows and global climate change intensifies (McCauley et al. 2015).

Deep-sea habitats, although remote, are under increased threats from a variety of direct and indirect anthropogenic pressures. Indirect threats comprise climate-induced changes in ocean biogeochemistry such as increased sea-surface temperature, increased thermal stratification, or decreased nutrient upwelling due to modifications in water mass circulation (Ramirez-Llodra et al. 2011; Hu et al. 2020; Jorda et al. 2020). Indeed, deep-sea communities rely on primary production in surface waters, and changes in quality and quantity of food supply, i.e. particulate

organic carbon (POC) flux, from the euphotic zone can profoundly affect deep-sea faunal communities (Smith, K. L. et al. 2013). Increased thermal stratification of the upper ocean, combined with changes in water mass circulation could also result in the extension of Oxygen Minimum Zones (OMZs), greatly decreasing abyssal biodiversity. Finally, increased ocean acidification may result in the decline of organisms with calcium carbonate skeletons (e.g. corals, molluscs, echinoderms, or foraminifera) as well as decreasing carbon flux to the abyss by changing plankton assemblages in surface waters (Ramirez-Llodra et al. 2010; Smith, C. R. et al. 2008).

Direct threats are widespread and their impacts remain poorly known. Although disposal of industrial and municipal waste seems to have decreased in the past decades, littering is a recognised environmental problem leading to the accumulation of (micro)plastics, metal, glass, and discarded or lost fishing gear on the deep seafloor (Ramirez-Llodra et al. 2011). Chemical pollutants also accumulate in deep-sea sediments and fauna, including persistent organic pollutants, toxic metals (Hg, Cd, Pb, Ni, isotopic tracers), pesticides, herbicides, and pharmaceuticals (Ramirez-Llodra et al. 2010; 2011).

Resource exploitation, primarily fishing and oil and gas extraction, has considerably increased in the deep-sea since the 1990s, as resources have been depleted in environments that are more accessible. Deep-sea fishing destroys habitats, as trawls are dragged over the seabed in a non-selective manner, generating great amounts of bycatch and leaving the seafloor barren (Ramirez-Llodra et al. 2011). Oil and gas extraction and deep-sea mining are other resource extraction industries that affect the deep seafloor. Deep-sea mining, although still in its infancy, targets three types of mineral resources and thus directly threatens various ecosystems: (1) manganese nodules on abyssal plains, (2) cobalt-rich ferromanganese crusts on seamounts, and (3) massive polymetallic sulphide deposits on hydrothermal vents. Both types of extraction industries are associated with potentially high-levels of habitat destruction and chemical pollution and therefore high impacts on deep-sea biodiversity (Fisher et al. 2016; Ramirez-Llodra et al. 2011) and ecosystem functioning (Zeppilli et al. 2016). Finally, although different anthropogenic impacts have different and potentially localized effects on deep-sea habitats and fauna, synergies between two or more impacts are largely unknown but likely to magnify individual effects (Fig. 5).

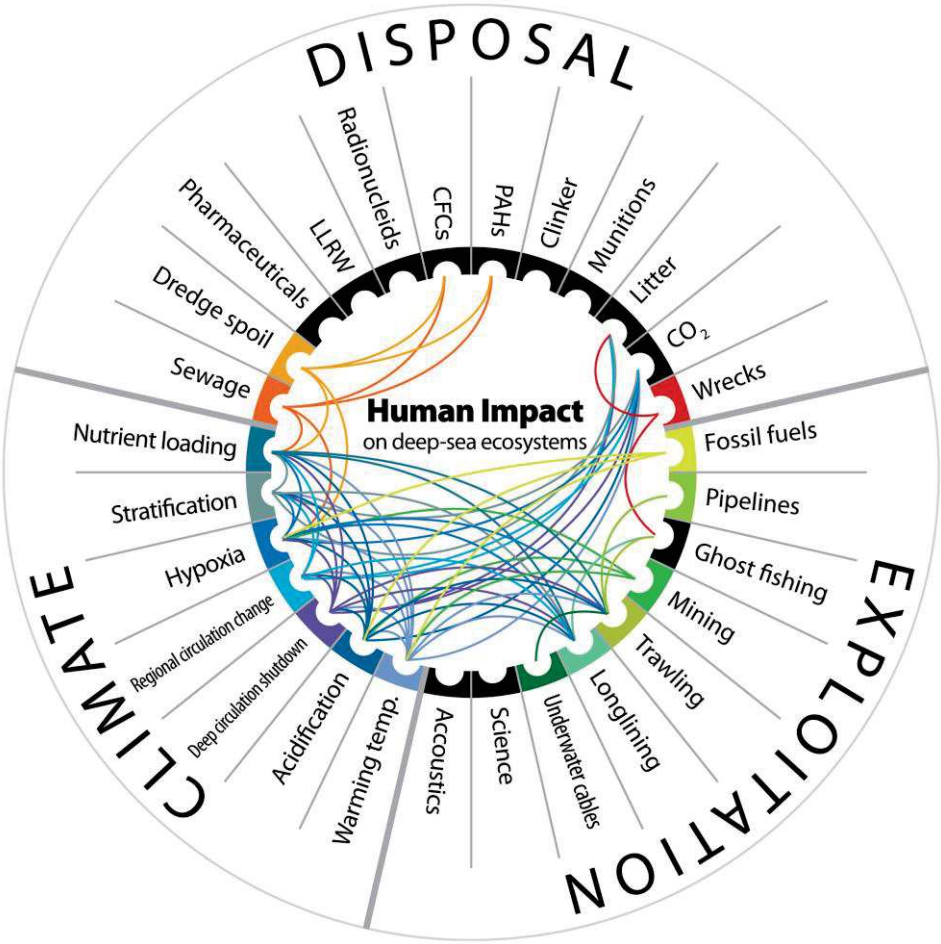


Figure 5. Anthropogenic impacts on deep-sea ecosystems and potential synergies. The lines link impacts that have synergistic effects on habitat or faunal communities. The lines are coloured to indicate the direction of the synergy. LLRW: low-level radioactive waste; CFCs: chlorofluorocarbons; PAHs: polycyclic aromatic hydrocarbons. From Ramirez-Llodra et al. 2011.

Given the magnitude of these impacts and their potentially global consequences, there is crucial need to increase our knowledge and understanding of the patterns and drivers of biodiversity in the marine biome. This is especially urgent for environments that are hard to access but may host a large variety of life forms and perform key roles in global nutrient cycles. The oceans cover 71% of the planet and are on average ~3,700 m deep. Half of all marine waters are below 3,000 m and approximately 90% of the oceans are considered deep sea (Ramirez-Llodra et al. 2010). Marine regions deeper than 2,000 m cover ~60% of the Earth's surface and have been postulated to be both a great reservoir of biodiversity and a source of important ecosystem services (Smith, C. R. et al. 2008; Pawlowski et al. 2011; Sinniger et al. 2016; Smith, K. L. et al. 2009). Yet, human exploration has described more about the surface of the moon and Mars than it has about this enigmatic backyard.

The beginning of deep-sea exploration took place in the Mediterranean in the mid-nineteenth century during the cruise of the *H.M.S Beacon* (1841-1842), where Edward Forbes and his colleagues were dredging in the Aegean Sea down to approximately 500 m. They noticed that biodiversity in sediments decreased with increasing sampling depth, and suggested that no life could be present below 600 m, a hypothesis known today as the "Azoic Theory" (Forbes 1844). This theory was highly debated, especially because evidence of life well below 600 m already existed (Risso 1816; McIntyre et al. 1975), and in the following years, data confirming deep-sea life accumulated (e.g. Sars 1849 and Jenkin 1862 in Ramirez-Llodra et al. 2010). This led to the launch of the *H.M.S Challenger* circumglobal expedition (1872-1876), whose aim was to study the physical, chemical, and biological processes in the deep ocean. This expedition is considered today as the birthmark of modern oceanography and initiated a series of other, primarily descriptive, expeditions in the 100 years to follow (USS *Albatross* Cruises 1882-1921, *Galathea* expedition 1950-1952, …). Since the 1960s, advances in deep-sea technology have permitted the development of deep-sea submersibles, Remotely Operated Vehicles (ROVs), Autonomous Underwater Vehicles (AUVs), and deep-sea permanent observatories. Meanwhile, advances in image capturing and sampling technologies are increasing the capabilities of scientists to explore, observe and experiment in the deep sea. These advances allowed the remarkably numerous findings of the last 30 years, and the description of unique habitats, such as cold-water coral reefs or chemosynthetic ecosystems like hydrothermal vents, cold seeps, and whale falls. It is estimated that since Forbes, twenty-two new habitats have been discovered, making it an average of one new habitat every 8 years. Yet, we have so far explored only 5% of the deep oceans and less than 1% of the deep seafloor, making the world's largest ecosystem the most poorly known biome on Earth (Ramirez-Llodra et al. 2010). It is thus clear that our appraisal of deep-sea habitats and the life they support is still extremely limited and that most of it may well be undiscovered.

I.IV. Oceanic regions and their associated deep-sea ecosystems

There are two broad realms in the oceans: the pelagic and the benthic. Pelagic refers to the open water in which swimming (the nekton) and floating (the plankton) organisms live. Benthic zones are defined as the bottom sediments or surfaces, and organisms living in or on it are called the benthos. Biologists have traditionally divided oceanic regions depending on depth (Fig. 6), although according to a recent meta-analysis of the largest worldwide databases it remains unclear whether depth zonation is ecologically meaningful in deeper waters (Costello and Breyer 2017).

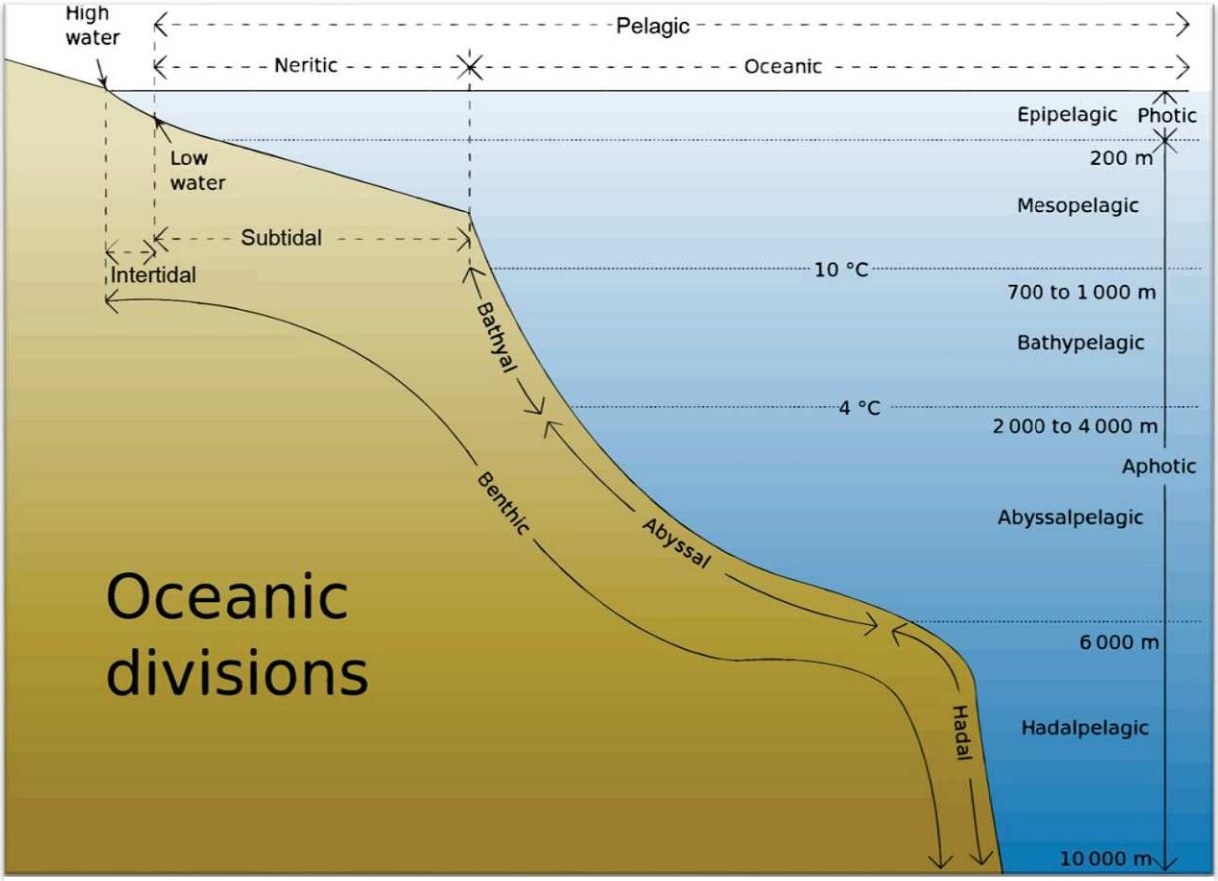


Figure 6. Oceanic divisions in the pelagic and benthic environments. Adapted from commons.wikipedia.org

The epipelagic, or photic zone, comprises the first 200 m of the water column where photosynthesis can take place, leading to high oxygen and low nutrient concentrations. Shallow benthic habitats close to the shore are additionally distinguished by tidal influence: the intertidal (interface between land and sea) hosts distinct communities adapted to air, wave action, and particular kinds of grazing and predation, while the subtidal comprises all the seafloor on continental shelves, to around 200 m depth. Deeper, light is too faint for photosynthesis to take place, but animals use this zone for feeding or avoiding predators. This so-called mesopelagic

or "twilight" zone (200- ~1,000 m) is therefore characterised by lower oxygen concentrations. Deeper still, the bathyal (~1,000 to ~2,000), the abyssal (~2,000-6,000 m), and the hadal (deep trenches below 6,000 m, see Fig. 6) are "true" deep-sea zones, characterised by low environmental variation, low temperatures, no sunlight, high oxygen concentrations, and higher nutrient levels (Costello and Breyer 2017)

The three-dimensionality of the water column and the fact that ~90% of the oceans are considered deep-sea (below 200 m) make the deep pelagic environment the largest biome on Earth, with over 1 billion cubic kilometres ($1x10^9$ km^3) hosting animals, plants, and microbes that grow, feed, and reproduce within the water column (Ramirez-Llodra et al. 2010). The deep seafloor (≥200 m) covers 360 million km^2, equivalent to over 50% of the Earth's surface (Ramirez-Llodra et al. 2011). It consists of a vast network of plains punctuated by specific topographical features (seascapes) such as slopes, mid-ocean ridges, deep-sea faults and trenches, but also canyons, seamounts, hydrothermal vents, methane seeps, mud volcanoes, or cold-water coral reefs (Fig.7, Costello 2009; Costello et al. 2010a). Each of these seascapes hosts specific prokaryotic and eukaryotic fauna. Of the benthic habitats, the abyssal plains represent ~70% of the seafloor, followed by continental margins (~10%) and ridge systems (~9%). Geographically more restricted habitats comprise seamounts, hydrothermal vents, cold seeps, food falls, cold-water coral reefs, and benthic oxygen minimum zones (OMZs), or

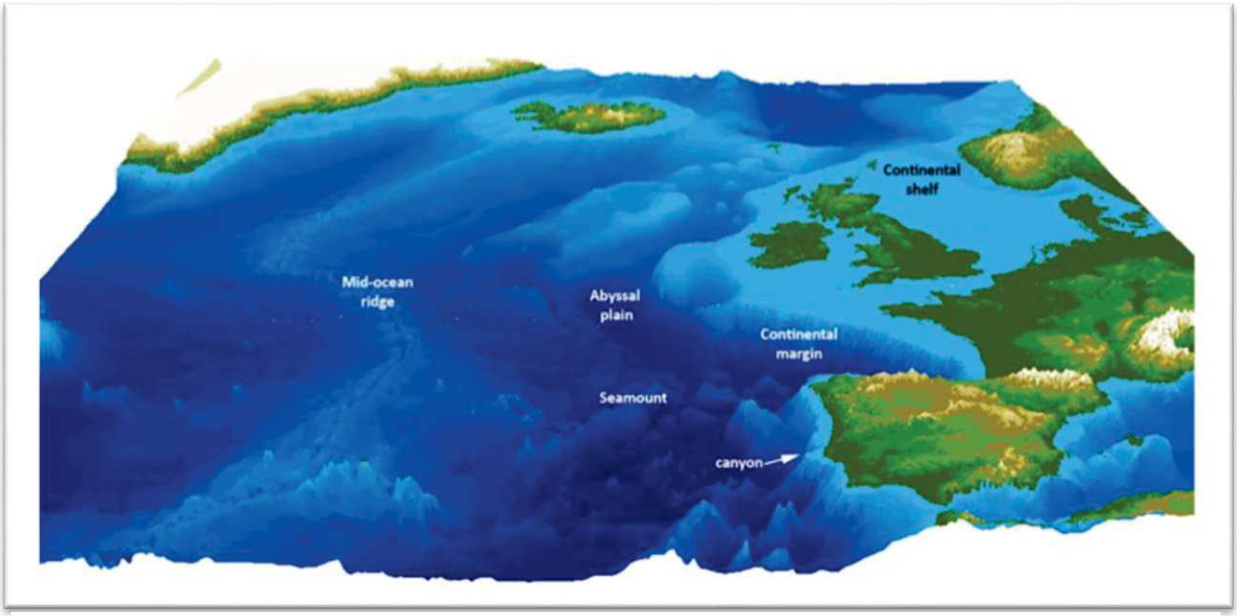


Figure 7. The Northeast Atlantic seafloor showing some distinct deep-sea ecosystems such as continental margins, which can include canyons, cold seeps, and cold-water coral reefs; the abyssal plain, seamounts, and the mid-ocean ridge where hydrothermal vents are found. From Ramirez-Llodra et al. 2010.

"DHABs"—Deep Hypersaline Anoxic Basins (Ramirez-Llodra et al. 2010; Merlino et al. 2018).

Abyssal plains

Lying between continental margins and mid-ocean ridges, abyssal plains are the greatest and least explored expanses on Earth (Fig. 8). They cover over 50% of the planet, are possibly the largest reservoirs of biodiversity and play a major role in important ecosystem services such as carbon cycling or calcium carbonate dissolution (Smith, K. L. et al. 2009; Smith, C. R. et al. 2008). The abyssal seafloor is mostly covered by very fine sediments (clays), termed abyssal mud or "ooze" (mud with a high percentage of organic remains). These sediments originate from the accumulation of pelagic organisms that sink after they die or from terrigenous particles derived from rock weathering on land. Hard substrates, such as manganese nodules, rock outcrops, or fault scarps also occur in many parts of the abyss, and these habitats host faunal assemblages that are different from those found in the surrounding soft sediments (Smith, C. R. et al. 2006). Abyssal plains are also characterized by an absence of *in situ* primary production (except at the spatially rare vents and seeps), well-oxygenated waters (except in OMZs) and low but constant temperatures of -0.5-3.0 °C. Regional differences can exist, such as in the Mediterranean and Red Sea where the average temperatures are higher, i.e. 14°C and 21°C respectively (Ramirez-Llodra et al. 2010; Smith, C. R. et al. 2006). Abyssal seafloor communities are food-limited as their productivity depends on the input of organic material falling down from the surface waters, termed marine snow (Smith, C. R. et al. 2008). Moreover, the abyssal seafloor is a dynamic environment, with regular (tidal currents, bottom currents, seasonal sedimentation) and episodic (benthic storms) disturbances that can affect benthic fauna (Ramirez-Llodra et al. 2010). While this fauna is not as conspicuous as in other deep-sea habitats, the abyssal seafloor is colonized by a great variety of mega- (macrourid fish, holothurians, echinoids…), macro- (crustaceans, polychaete worms, nematodes, gastropods…), and meiofauna (nematodes, harpacticoid copepods, foraminiferans, rotifers and other protists) with potentially high population densities (Smith et al. 2009; Smith et al. 2008; De Broyer et al. 2004).

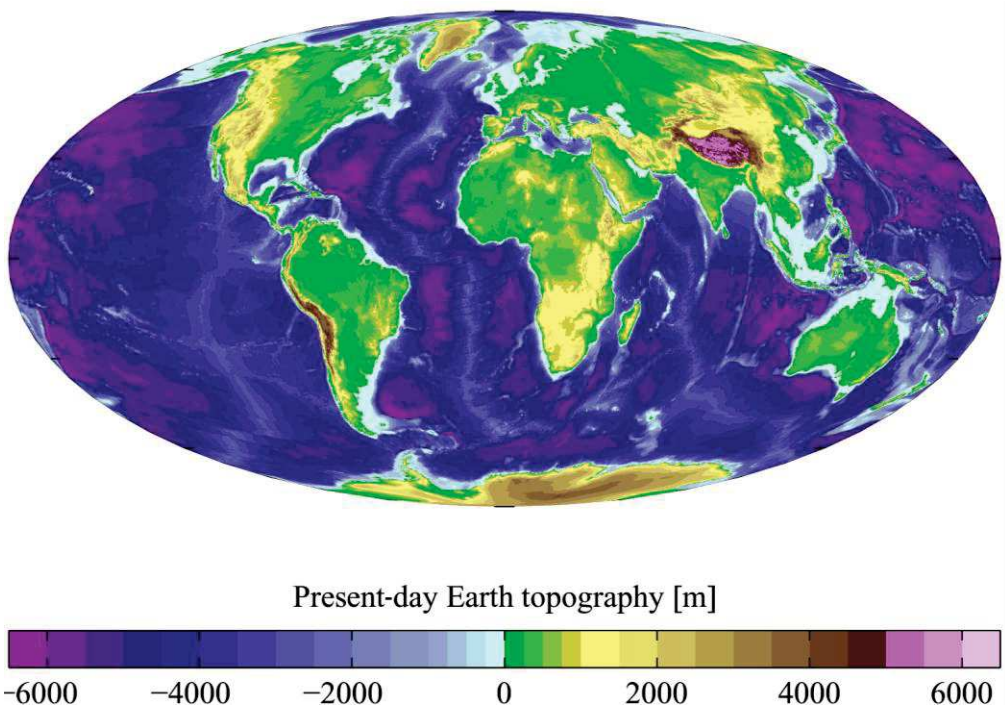


Figure 8. Present-day Earth topography, with abyssal regions (-4,000 to -6,000 m) in dark-blue and purple. From NOAA's NCEI.

Mid-ocean ridges

Mid-ocean ridges are a type of divergent plate boundary; they are deep-sea volcanic chains and the longest mountain ranges on Earth. They extend through all major ocean basins, with a total length over 60,000 km (Fig. 9). They usually occur in the middle part of the oceans (with the exception of the east Pacific rise) and their crests rise around 1,000-3,000 m above the adjacent seafloor (Wilson 2007).

They are so-called ocean spreading centres, as magma constantly emerges onto the seafloor to form new ocean crust. Mid-ocean ridges generate about three km^2 of new seafloor every year, driving continental drift and seafloor spreading at a rate of 1-10 cm/year. The current spreading episode began around 200 million years ago, with the opening of the Atlantic and Indian ocean basins, which are still growing, while the Pacific is decreasing (Ramirez-Llodra et al. 2010; Wilson 2007).

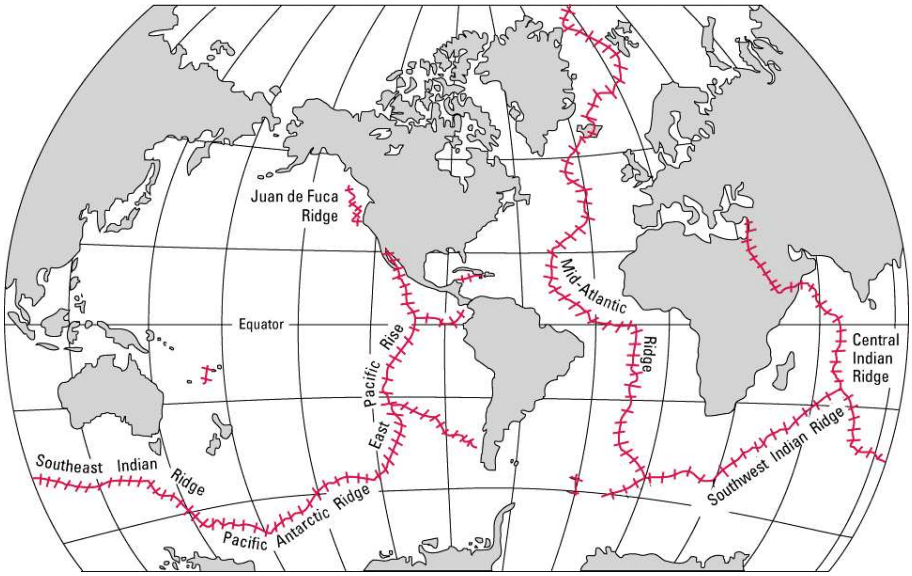


Figure 9. World distribution of mid-ocean ridges (from US Geological Survey).

Mid-ocean ridges offer a high diversity of habitats, from hills and seamounts to axial valleys and fracture zones dropping to more than 4,000 m. The presence of these huge mountain ranges affects the distribution of both pelagic and benthic organisms, as they represent dispersal barriers for species distributed in neighbouring abyssal plains. The substratum present along the ridges is primarily rocky because these areas are too geologically new to have accumulated much sediment. Thus, ridges provide habitat to a variety of sessile fauna, from filter feeders to chemosynthetic organisms, which take advantage of the specific hydrographic conditions produced along the ridge.

In particular, hydrothermal vent ecosystems arise when cold seawater seeps down into the ocean crust and reacts with magma to generate hot (up to 407 °C), chemically-laden fluids (Haase et al. 2009). Chemosynthetic bacteria and Archaea use these reduced minerals exported by the vent fluids as sources of energy to fix inorganic carbon. These primary producers can be found both free living, forming microbial mats, and in symbiosis with many mega-, macro, and meiofauna (Dubilier et al. 2008; Bellec et al. 2018). The latter comprise invertebrates that filter or graze on the microorganisms (e.g. barnacles, limpets), and numerous invertebrate taxa that host the microorganisms as epi- or endosymbionts, such as worms (e.g., siboglinid polychaete tubeworms, flatworm, nemerteans, nematodes), bivalves (e.g., mytilids, vesicomyids, lucinids, and thyasirids), gastropods (e.g., abyssochrysoideans), and many families of decapod crustaceans (e.g., alvinocaridid shrimp or galatheid squat lobsters) (Dover, Van et al. 2002; Ramirez-Llodra et al. 2010; Martin and Haney 2005; Desbruyeres et al. 2006; Dover, Van et

al. 2001; Bellec et al. 2018). Even though these food-rich oases are often space and time limited, they support high-biomass communities that are different from those in/on the surrounding seafloor. The establishment of symbiosis between chemoautotrophic microorganisms and fauna allows the latter to harness abundant chemical energy and explains the success of vent, seep, and food fall communities as well as the high biomass observed, especially in the megafauna. Over 600 vent and seep species have been described since 1977, and more than 50 new species have been recorded from whale falls in the North Pacific alone. Many of these taxa have diversified within these reducing habitats at high taxonomic levels (Dover, Van et al. 2002; Ramirez-Llodra et al. 2010). However, although exhibiting high densities, these ecosystems support a low diversity compared to the surrounding benthos, with communities usually dominated by a few species. It has been suggested that this is due to the extreme conditions (high temperatures, H_2S…) encountered in these environments, which select for a small number of taxa (Ramirez-Llodra et al. 2010).

Active and passive continental margins

Continental margins are the zones of the ocean floor that separate oceanic crust from continental crust. They have very high habitat heterogeneity and are the most geologically diverse components of the seafloor (Ramirez-Llodra et al. 2010). Geologists differentiate geologically active from passive continental margins (Fig. 10).

Active margins are convergent plate boundaries and occur mostly in the Pacific and Indian oceans. They are so-called subduction zones, regions in which the denser oceanic plate sinks under the terrestrial plate, back into the Earth's interior. This creates ocean trenches plunging >10,000 km deep and active magmatism resulting in a great diversity of deep-sea geological formations, from volcanic islands or mountain chains, to submerged volcanoes known as seamounts (island arcs), and back-arc basins that produce similar habitats than mid-ocean ridges (Wilson 2007; Ramirez-Llodra et al. 2010).

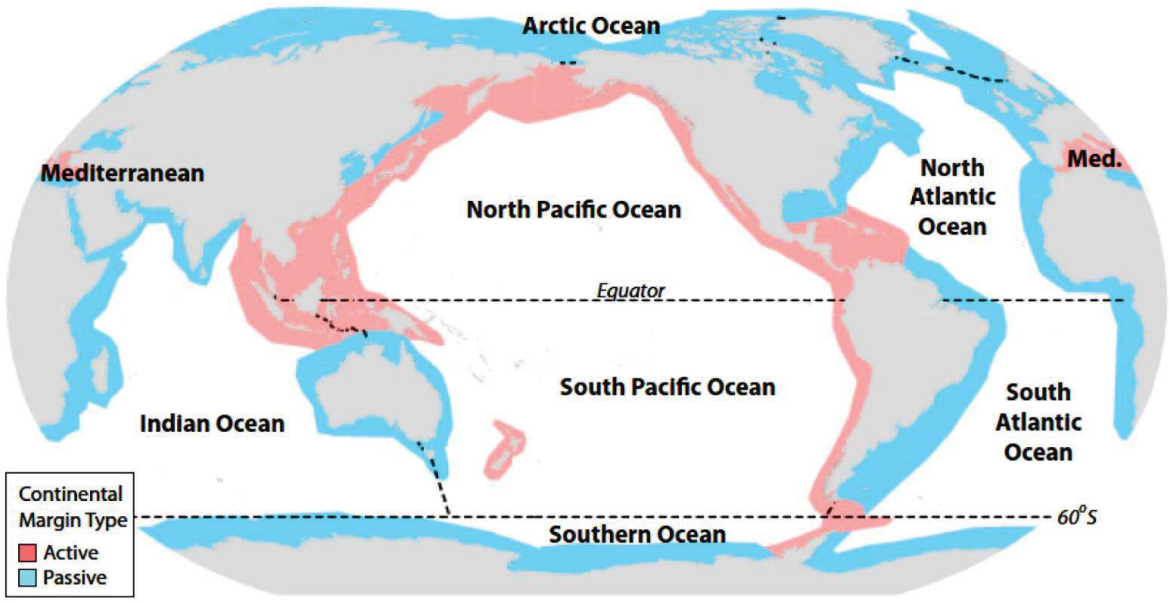


Figure 10. Map showing the locations of active (red) and passive (blue) continental margins in the ocean basins. From www.geologyin.com

Passive margins occur when an ocean rift has split a continent in two (Fig. 10). Sedimentation is the primary driving force of passive margins and all processes affecting sediment input (types of continental rocks, topography of adjacent land masses, and productivity of surface waters…) greatly influence the margin geomorphology. This results in the formation of distinct habitats including sedimentary slopes, submarine canyons or cold-water coral reefs and gardens. Among those, canyons have been shown to be essential habitats for the local fauna, i.e. habitats used by fauna for critical aspects of their life cycle. Also, canyons modify local current regimes and are important conduits for the transport of particles between the continental shelf and the abyss. They harbour diverse habitats, from rocky outcrops on canyon head and walls, which are dominated by sessile filter feeders like cnidarians and sponges, to soft sediment in the canyon axis, with a fauna dominated by deposit feeders, scavengers, and predators like echinoderms, crustaceans, and fish (Ramirez-Llodra et al. 2010). Cold-water coral reefs are another characteristic habitat along passive continental margins and can occur within canyons but were also discovered in many other environments (continental slopes, fjords, seamounts…). They are formed by a heterogeneous group of azooxanthellate cnidarians, with representatives from hydrozoans (Stylasteridae), octocorals (Alcyonaria, Gorgonacea, and Pennatulacea), and hexacorals (Scleractinia and Antipatharia). Deep reefs develop in a much slower process than shallow-water reefs, but have the ability to establish stable, long-lasting and highly diverse ecosystems (Murray Roberts et al. 2006). Cold-water

corals are therefore habitat-forming organisms, which provide shelter for many organisms in the deep-sea. They have been recorded in all oceans and in the Mediterranean Sea, from 50 m down to 6300 m in the Pacific (Ramirez-Llodra et al. 2010).

Along continental margins, sub-seafloor geological processes, like groundwater discharge or organic matter decomposition, also influence the environment and give rise to cold seeps, where hydrocarbon-rich fluids leek out of the ocean floor (oil or gas seeps, brine pools, and mud volcanoes). These geological features produce very specific types of substrata and thus sustain different geochemical and, mainly chemosynthetic, microbial processes (Olu et al. 1996; Sibuet and Olu 1998; Bernardino et al. 2012; Vanreusel et al. 2010b). Cold seeps therefore harbour fauna similar to those found on hydrothermal vents, especially at higher taxonomic levels. Like vents, cold seeps are chemosynthetic systems supporting dense communities of faunal groups such as bivalves (mytilids, vesicomyids, lucinids, thyasirids), siboglinid tubeworms, decapod crustaceans (shrimp and crabs), gastropods, and cladorhizid sponges (Sibuet and Olu 1998; Olu-Le Roy et al. 2004; Olu et al. 2010).

Seamounts

Seamounts are topographically isolated and submerged peaks of volcanic origin, rising more than 1,000 m above the surrounding seabed, although recent definitions include prominences of 100-1,000 m in height (Etnoyer et al. 2010). More than 100,000 seamounts have been revealed by satellite gravimetry data worldwide (Ramirez-Llodra et al. 2010), and estimates range from hundreds of thousands to over 1 million (Staudigel and Clague 2010; Costello et al. 2010a). Long chains of seamounts can also occur, marking the presence of “magmatic hotspots” (like the Hawaiian Islands). Their overall abundance makes them one of the most common but least understood marine biomes on Earth, covering an area at least the size of Europe and Russia combined (Etnoyer et al. 2010). Most seamounts have a complex topography, which modifies surrounding ocean currents, resulting in increased productivity over and around these seascapes. Due to this concentration of organic matter, seamounts can harbour large communities with complex trophic networks, making them hot spots of diversity and nurseries for commercial species. Seamounts also provide a rocky substratum due to their steepness, and therefore offer distinct benthic habitats compared to the surrounding sedimentary ocean floor. Seamounts are thus colonized by a range of mainly epifaunal suspension feeders, dominated by cnidarians (gorgonians, zoanthids, antipatharians, actinians, pennatulids, and

hydroids), while sponges, cirripeds, molluscs, crinoids, ascidians, ophiurids, asteroids, and holothurians can also be found (Ramirez-Llodra et al. 2010).

Hadal trenches

The hadal zone extends from ~6,000 m down to the deepest trenches at almost 11,000 m depth, accounting for 45% of the total ocean depth range. Hadal research has been revived in the past ten years thanks to technological developments, and studies have described 46 distinct and often extremely isolated hadal habitats. They comprise 33 trenches, occurring in tectonic convergence zones and resulting from subduction or faulting, and 13 troughs, hadal basins within abyssal plains, not formed at convergent plate boundaries (Jamieson 2015; 2011). Similarly to abyssal and bathyal ecosystems, hadal environments display low temperatures (1-4 °C) with limited within-trench variability, and low food supply, although the latter can be greater than in neighbouring abyssal habitats, suggesting that trenches may accumulate organic matter due to their steep topography (Glud et al. 2013; Leduc et al. 2016). However, the combination of low temperature, high-pressure (650-1,100 atm), and low food supply makes the hadal zone a unique environment, requiring particular physiological adaptations (Zeppilli et al. 2018). Combined with geographical isolation, this explains the high levels of species endemism reported in hadal habitats (Jamieson 2015; Blankenship-Williams and Levin 2009). Characteristic members of the macrofauna of hadal zones include scavenging amphipods and snail fishes (Jamieson et al. 2010; Linley et al. 2016). Although lower diversity was reported, smaller taxa (<1 cm) are the most abundant members of benthic communities, with densities (~100-~1,000 individuals per 10 cm^{-2}) similar to values in abyssal environments (Zeppilli et al. 2018).

Other known benthic ecosystems: OMZs and organic falls

Oxygen depletion is widespread in the world oceans, and zones of permanent hypoxia are defined as oxygen minimum zones (OMZs), in which oxygen concentrations are below 0.5 $ml.l^{-1}$ or 22 µM (Ramirez-Llodra et al. 2010). They occur at different water depths, from shelf to bathyal areas (10-1,300 m) and usually develop under regions of intense upwelling and surface productivity, due to the consumption of oxygen by aerobic bacteria that degrade dead organisms falling down the water column. When OMZs intercept with continental margins or seamounts, they produce hypoxic or anoxic sediments that are major sites of carbon burial and greatly

influence benthic assemblages (Levin, L. A. 2003; Ramirez-Llodra et al. 2010). OMZs allow the establishment of extensive mats of large sulphide-oxidizing bacteria and high-density, low-diversity protozoan and metazoan communities that have specific adaptations to hypoxia. Adaptations include small, thin bodies, enhanced respiratory surface areas, blood pigments such as haemoglobin, increased numbers of pyruvate oxidoreductases, formation of biogenic support structures for stability in soupy sediment, and the prevalent association to chemosynthetic symbionts similar to those of hydrothermal vents and cold seeps (Levin, L. A. 2003). Dense aggregations of protists and metazoan meiofauna including calcareous foraminifera thrive in OMZs. In contrast, low-diversity macro- and megafauna assemblages are common on the edges of OMZs (Ramirez-Llodra et al. 2010).

As most of the deep seafloor is typically food-limited and highly oligotrophic, sunken wood, cetacean carcasses, or other food falls represent local and temporally fluctuating resources. Providing food, shelter, and substrate, whale and wood falls produce new habitat that is distinct from the surrounding ocean floor. Indeed, cold temperatures, high hydrostatic pressures, and slow decomposition rates allow these organic falls to remain intact, permitting the establishment of complex but localized ecosystems that can last for decades (Smith, C. R. and Baco 2003). Whale fall communities undergo at least three successional stages that are characterized by different faunal assemblages. First a mobile *scavenger stage*, characterized by large animals such as sleeper sharks, hagfish and other invertebrate scavengers, followed by an *opportunistic stage* during which the organically-enriched sediment gets colonized by opportunistic heterotrophic invertebrates (mainly polychaetes and small crustaceans). Finally a *sulfophilic stage* in which the whale fall gets colonized by highly specialized and dense communities of chemosynthesis-driven fauna, including mytilid mussels, vesicomyid clams, polychaete worms, diverse crustaceans (giant isopods, shrimps, lobsters), gastropods, ctenophores, or lancelets (Fujiwara et al. 2007; Goffredi et al. 2004). Although strong differences can exist between the organisms inhabiting vents, seeps, and food falls, the communities of these highly sulphidic environments share many dominant taxa at the family and genus level, suggesting widespread dispersal mechanisms between chemosynthetic habitats (Bernardino et al. 2012; Teixeira et al. 2013; 2012). Moreover, some generalist species even seem to inhabit multiple types of reducing ecosystems, although this may be undermined by cryptic speciation (Dover, Van et al. 2002). It has thus been suggested that large organic falls serve as stepping-stones for the evolution and dispersal of highly specialized chemosynthetic taxa inhabiting hot vents and cold seeps (Bienhold et al. 2013).

I.V. Deep-sea benthic biodiversity patterns

The major part of research in the deep-sea has been directed towards mid-ocean ridges, and their associated chemosynthetic hydrothermal vent ecosystems. Although scientifically interesting, these ecosystems only represent a small area of the ocean floor. Ridges cover 9.2% of the seafloor, and < 1% of the latter are hydrothermal vents. In comparison, abyssal plains represent 75% of deep-sea habitats and < 1% have been investigated (Smith, K. L. et al. 2009; Ramirez-Llodra et al. 2010).

The bathyal and abyssal heterotrophic sedimentary seafloor was until recently believed to be a monotonous and poor ecosystem, interspersed by oases of extremely high productivity and high biomass, where organic material falling down the water column accumulates (e.g., seamounts, canyons, food falls) or where nutrient-rich fluids allow the establishment of chemosynthesis-driven ecosystems. In contrast, deep-sea benthic sedimentary communities were found to harbour high species diversity and high levels of evenness, some authors suggesting that they may be comparable to tropical rainforests (R. Hessler and L. Sanders 1967; Grassle 1989; Smith, C. R. and Snelgrove 2002).

Investigations of deep-sea sedimentary habitats during the 1970s to 1990s were centred on macrofauna of continental shelves and bathyal depths, predominantly along the North American and European margins (Levin, L. A. et al. 2001; Smith, C. R. and Snelgrove 2002). Research on smaller benthic size compartments was geographically restricted (Thiel 1983; Snider et al. 1984; Tietjen 1992; Danovaro et al. 1995; Soltwedel 2000). However, these studies revealed the extreme patchiness of species distributions in the deep-sea, and highlighted that research on species diversity in this biome must include this variability at small (centimetres), local (meters), and large (kilometres) spatial scales (Rex 1981). Patchiness is mostly a result from variations in food availability, and the great diversity and evenness observed in deep sea sediments are in part a response for optimizing the exploitation of the limited food sources, and have positive consequences on the stability and resilience of deep-sea benthic communities (Ramirez-Llodra et al. 2010). Nevertheless, significant regional variations in the relationship between species diversity and abundance with food availability do exist, thought to result from the influence of environmental variation (pressure, temperature, oxygen concentrations, sediment granulometry) and biotic interactions (Levin, L. A. et al. 2001; Rosli et al. 2017).

Research at multiple spatial scales (Fonseca, G. et al. 2010; Danovaro et al. 2013; Gambi and Danovaro 2006; Gaever, Van et al. 2010; Bianchelli et al. 2013), and targeting a diversity

of ecosystems (Danovaro et al. 2009a; Bianchelli et al. 2010; Olu-Le Roy et al. 2004; Zeppilli et al. 2013; Clark et al. 2010; Smet, De et al. 2017) has strongly increased in the past twenty years, with a notable effort on taxa with smaller body sizes. Large international long-term collaborations allowed shedding light on ecosystems or ocean basins at large spatial scales (Danovaro et al. 2010; Vanreusel et al. 2009; Danovaro et al. 2009b), or on very remote ocean regions, such as the southern ocean (Brandt, A. et al. 2007b; a; Brandt, A. and Ebbe 2009) and the arctic (Hasemann and Soltwedel 2011; Górska et al. 2014; Bodil et al. 2011; Renaud et al. 2006). Research in the Pacific has also expanded, with a particular focus on the New-Zealand margin (Leduc et al. 2012a) and pacific hadal trenches (Itoh et al. 2011; Kitahashi et al. 2013; Leduc et al. 2016). More recently, eDNA metabarcoding tools were successfully applied to deep-sea sediments, for e.g., in the Mediterranean (Guardiola et al. 2016; Cordier et al. 2019a) and the Atlantic (Pawlowski et al. 2011; Bik et al. 2012b; Lejzerowicz et al. 2014). Overall, these studies highlight that biogeographic and species distribution patterns in the deep-sea show considerable variability with body size, life history, and taxonomic identity. To achieve a global synthesis of these patterns, deep-sea research must thus include both spatial and biological variability at various scales (Smith, C. R. et al. 2006).

Benthic size classes and their main differences

Deep-sea benthic fauna is divided into four major somewhat overlapping categories, primarily based on their body size, but also habitat, ecological features (e.g. feeding mode), and taxonomy (Rex 1981; Thiel 1983). Assemblages and species ranges have mostly been investigated for larger taxa (Levin, L. A. et al. 2001), probably mostly due to the difficulty in morphological identification of small organisms. The *megafauna* comprises conspicuous epibenthic animals, larger than 2 cm and readily visible on photographs. It includes highly mobile demersal and benthopelagic fishes and amphipods, but also obligate bottom dwellers such as echinoderms (e.g., brittle stars, crinoids, sea stars, and sea cucumbers), arthropods (e.g., decapods, pycnogonid sea spiders), corals, or sponges. Deep-sea *macrofauna* is composed of animals retained on a 250-300 µm sieve, but not readily visible on photographs, thus having body sizes of 1 cm to 2 cm. It includes numerous familiar invertebrate phyla, and is particularly dominated by polychaete worms, peracarid crustaceans (e.g., amphipods, isopods, cumaceans, tanaids), bivalves, and gastropods. The diversity of deep-sea megafauna is much lower than

that of the macrofauna, and megafauna accounts for lower abundance and biomass throughout depth ranges from 0-6,000 m (Rex 1981; Rex et al. 2006).

Meiofauna includes both metazoans as well as some small single-celled protists. The boundaries between meiofauna and macrofauna were defined by mesh sizes of the sieves used for extracting these organisms from the sediments, and as different studies used different mesh sizes, these boundaries could vary widely among researchers. Today a size range of 32 µm to 1 mm seems generally accepted (Soltwedel 2000; Thiel 1983; Snider et al. 1984). As the separation between meio- and macrofauna is biologically speaking artificial, some groups are found in both size fractions. This means that the meiofauna size class may include juveniles or larvae of macrobenthos (e.g., cnidarian polyps, annelids, copepods, or tunicates), also called temporary meiofauna (Giere 2009). Similarly, large nematodes or copepods will be part of the macrofauna. Better-known metazoan taxa that are predominantly in the meiofauna size class comprise nematodes, copepod and ostracod crustaceans, certain malacostracan crustaceans (e.g., members of the Isopoda, Amphipoda, Tanaidacea), but also tardigrades, kinorhynchs, loriciferans or halacaroid mites (Thiel 1983; Giere 2009). However, there are also many smaller and/or soft-bodied taxa, largely disregarded in morphological inventories, probably because their bodies get broken during the sieving process. These include interstitial cnidarians (hydrozoans, scyphozoans, and anthozoans), free-living platyhelminths, the Gnathifera (Gnathostomulida, rotifers, micrognathozoans), the Gastrotricha, the Sipuncula, some chaetognaths, but also brachiopods and bryozoans. Unicellular heterotrophic meiofauna, often neglected by zoologists, are also surprisingly diverse in the meiofauna and comprise members of the Foraminifera, the Heliozoa, the Amoebozoa, or the Ciliophora (Giere 2009). Finally, the *nanofauna* comprises all organisms smaller than 42 µm, and includes some metazoans, but mainly consists of flagellates, ciliates and yeasts (Thiel 1983; Tietjen 1992).

A global-scale analysis of abundance and biomass of major benthic size classes found that all animal size classes (metazoan meio-, macro-, and megafauna) significantly decrease in abundance and biomass with depth, while the values showed no decline with depth for bacteria (Rex et al. 2006). The decrease observed in metazoans was less steep for the smaller meiofauna, than for the macro- and megafauna, indicating that animal sizes in deep-sea communities as a whole decrease with depth. This leads to an increase in the relative abundance of small organisms (meiofauna, bacteria) with increasing water depth, their smaller size allowing them to cope better with low food availability (Thiel 1983; Rosli et al. 2017). Meiofauna are thus an

important component of deep-sea benthic communities due to their high relative abundance and diversity, their close connection to other size compartments of the benthos, and their important role in benthic food webs (Zeppilli et al. 2018; Schratzberger and Ingels 2017). Meiofauna abundances can range from 100 to 1,000 individuals per m^2 (Tietjen 1992), communities being dominated by foraminiferans, nematodes, and copepods (Thiel 1983; Snider et al. 1984; Tietjen 1992). Nematodes generally comprise ~ 90% of metazoan individuals, compared to 3% to 10% for copepods. However, nematodes do not dominate meiobenthic biomass to the same extent that they do abundance, as individual body weights can be larger in other organisms. Nematodes thus constitute 13% to 65 % of meiofaunal biomass in most deep-sea sediments, compared to for e.g. 15% to 75% for copepods (Tietjen 1992). Snider et al. (1984) first highlighted the importance of Foraminifera in meiofaunal communities. The authors showed that Foraminifera comprised ~50% of meiofauna individuals in sediments of the North Pacific, and made up 87% of biomass. The extraordinary numbers and diversity of Foraminifera in deep-sea sediments has been subsequently confirmed by numerous investigations worldwide (Brandt, A. et al. 2007a; Gooday 1999; Gooday et al. 2004). eDNA-based studies have confirmed the great diversity of nematodes, which are usually found to be the most diverse metazoan group (Sinniger et al. 2016; Guardiola et al. 2016). They also highlighted the diversity of less-studied metazoan phyla like the Platyhelminthes, the Nemertea, and the Xenacoelomorpha (Pawlowski et al. 2011; Sinniger et al. 2016; Guardiola et al. 2016), and confirmed the unprecedented abundance of other, mostly unicellular, eukaryotic groups, like the SAR and the Fungi (Pawlowski et al. 2011; Sinniger et al. 2016; Guardiola et al. 2016).

In terms of biomass, smaller size classes replace larger size classes with increasing depth. While mega- and macrofauna dominate biomass at upper bathyal depths (above ~2,000 m), this is reversed in the abyss. It was thus suggested that the bathyal zone (i.e. upper continental slopes), providing higher levels of energy supply, offers more ecological and evolutionary opportunities for adaptive radiation, at least for larger organisms (Rex et al. 2006). Examinations of depth ranges of deep-sea gastropods and bivalves lead to the proposal of the slope-abyss source-sink (SASS) hypothesis for abyssal diversity (Rex et al. 2005). The authors suggested that the abyssal seafloor might constitute a vast sink of larval refugees from upper continental slopes, whose populations are not reproductively self-sustaining. It has however been found that abyssal macrofauna populations are unlikely sustained by bathyal standing

stocks alone, and that local abyssal reproduction has to be considered, especially in high-productivity areas (Hardy et al. 2015).

In terms of distribution ranges, strong differences exist depending on size class, life history, and taxonomic identity. Interestingly, abundant genera seem to be abundant all over the world, so-called "cosmopolitan deep-sea genera". Some taxa of the mega- (e.g. rattail fishes, elasipod holothurians), macro- (e.g., isopods, amphipods, neogastropods), and even meiofauna (e.g., Foraminifera, harpacticoid copepods) exhibit very wide (> 1,000 km) distribution ranges (Smith, C. R. et al. 2006; Menzel et al. 2011; Easton and Thistle 2016; Gooday et al. 2004). For larger fauna, this is explained by their benthopelagic lifestyle, and/or their good dispersal capacities, planktotrophic larvae being able to survive in the water column for months to over a year (Smith, C. R. et al. 2006; Costello and Chaudhary 2017). For the metazoan meiofauna, lacking a planktonic life stage, this is surprising and has been coined the *meiofauna paradox* (Giere 2009; Carugati et al. 2015). Some authors suggest that passive transport by bottom currents after resuspension may enhance dispersal in these small taxa (Menzel et al. 2011), but it is still unclear whether these are enough to explain the observed wide distribution ranges. In addition, molecular studies have revealed that cosmopolitan megafauna and macrofauna species are often complexes of cryptic species that each have much smaller distribution ranges (Teixeira et al. 2013; Havermans et al. 2013). Thus, the generality of wide distribution ranges remains to be confirmed, especially for meiofauna.

Similarly, the strong species turnover observed between sites or regions may simply reflect global under sampling of deep-sea environments. Populations described as different morphospecies due to discrete and distant distribution ranges may be the result of sampling artefacts and in fact be the same species genetically. Consequently, it is still extremely difficult to differentiate between rarity and endemicity, and the high degrees of endemicity as well as the high percentages of new species found may decrease as more information is gathered (Smith, C. R. et al. 2006; Brandt, A. et al. 2007a; Teixeira et al. 2013).

Ecological patterns at regional (~100-10,000 km) spatial scales

Understanding species distribution patterns has been a primary interest of biologists since the beginning of large-scale voyages of scientific exploration in the late 17th century. Indeed, biodiversity is distributed heterogeneously across planet Earth: while some regions appear to

be extremely diverse (rainforest, coral reefs), others *seem* devoid of life (deserts, polar regions) and most are somewhere in between (Gaston 2000). Since the 1970s, a considerable amount of work has tried to explain broad-scale geographical patterns in marine biodiversity. Indeed, large-scale distribution boundaries reveal the importance of global factors that can influence species distribution, such as continental drift, salinity and temperature, sea-level rise, or glaciation (Costello et al. 2017)

Overall, large-scale biogeographic regions do exist on the deep-seafloor both for chemosynthetic and heterotrophic ecosystems and correlate well with the major ocean basins (Dover, Van et al. 2002; Bik et al. 2012b; Moalic et al. 2012; Watling et al. 2013; Costello et al. 2017). However, most studies attempting to delimitate these boundaries focused on megafauna or nanofauna, both considered to have good dispersal abilities, as such these boundaries remain to be confirmed for the (metazoan) meiofauna. Menzies et al. (1973) summarized the distributions of many megafauna as well as isopod crustaceans to delineate five large biogeographic regions in depths > 4,000 m, one in each ocean basin (Pacific, Arctic, Atlantic, Indian, and Antarctic). This work was recently extended by the Global Open Ocean and Deep Seabed (GOODS) classification using high-resolution water mass characteristics (temperature and salinity) and particulate organic-matter flux data to the seafloor. This resulted in the delineation of 14 lower bathyal, 14 abyssal, and 10 hadal geographic provinces within the five biogeographic regions (Watling et al. 2013). The classification into geographic areas may not truly represent biogeographic realms, as it lacks species information. The latest global study and first holistic analysis of the Ocean Biogeographic Information System (OBIS) revealed 18-continental-shelf and 12 offshore deep-sea realms, reflecting the wider distribution ranges currently recorded for many deep-sea species (Costello et al. 2017).

Overall, these studies show that regional geological history can affect diversity as events such as glaciation or isolation can induce higher extinction or speciation rates. Historical events, particularly during the Cenozoic, have resulted in both geological and oceanographic changes (e.g. isthmus closures, opening of ocean basins, sea level rise and fall, periods of deep-sea anoxia). These have been important in defining contemporary biogeography of many deep-sea taxa by controlling larval dispersal and survival (Herrera et al. 2015; Dover, Van et al. 2002; Smith, C. R. et al. 2006; Costello et al. 2017).

Furthermore, large-scale spatial distributions in the marine biome are primarily driven by temperature, salinity, habitat complexity, and food (and oxygen) availability (Tittensor et al. 2010; Smith, C. R. et al. 2008; Fonseca, V. G. et al. 2014; Costello and Chaudhary 2017;

Costello et al. 2017). Latitude is a *proxy* for temperature and solar radiation (including day length and seasonality), which are known to influence primary and secondary productivity. Similar to what has been observed on land, marine species richness thus varies with latitude. Unimodal marine latitudinal gradients in species richness, i.e. diversity increasing from high to low latitudes, was reported in the Atlantic for some deep-sea macrofauna (Rex et al. 1993). However, these patterns are not confirmed globally. Indeed, studies have found increasing nematode diversity from the tropics northwards and very similar meiofaunal taxa richness at all latitudes (Ramirez-Llodra et al. 2010 and references therein). Moreover, the most comprehensive study in the Southern Ocean challenges ideas that deep-sea diversity is lower at higher latitudes, given the extraordinary diversity found in both meio- and macrofauna in the southern ocean (Brandt, A. et al. 2007b). Recent global studies reported bimodal gradients with latitude, with highest species numbers in the subtropics and a dip near the equator (Chaudhary et al. 2016; 2017). They suggested that temperature is the main driver explaining species richness patterns, a statement congruent with what has been observed in euphotic plankton (Sunagawa et al. 2015; Ibarbalz et al. 2019). In contrast, in the deep-sea, where temperatures are uniformly low, biodiversity patterns are primarily driven by food supply. This has been shown by numerous studies at the regional to global scales, for taxa from all size compartments (Woolley et al. 2016; Levin, L. A. et al. 2001). Large-scale studies of meiobenthic diversity even suggested that it is primarily niche-driven, i.e. dependent on contemporary ecology and food supply rather than historical events (Lambshead et al. 2002; Fonseca, V. G. et al. 2014). Food availability, i.e. nutrient supply to the deep sea, varies depending on 1) distance from the coast, 2) depth, and 3) large-scale ocean currents. Thus, any variation in these three parameters will directly influence species abundance and diversity.

Effect of food and oxygen availability

Particulate organic flux towards the abyss varies as a function of primary production in the surface waters. It has been calculated that only 0.5-2.0 % of the net primary production reaches the deep seafloor below 2000 m (Ramirez-Llodra et al. 2010). Deep-sea benthic communities are thus amongst the most food-limited on the globe (Smith, C. R. et al. 2008). The primary productivity of ocean surface waters varies both regionally and seasonally, thus seasonal patterns of diversity occur in the deep-sea benthos (Smith, K. L. et al. 2009; Massana et al. 2015; Guardiola et al. 2016), as well as regional differences, for e.g., between upwelling zones and oligotrophic central gyres (Smith, C. R. et al. 2006). Studies on meiofauna clearly

demonstrate regional differences on a global scale: richer communities are generally found in areas with increased productivity and enhanced organic matter flux to the seafloor (Soltwedel 2000). Ocean thermohaline circulation patterns greatly influence carbon export flux to the deep-sea and this, combined with higher productivity in subtropical surface waters and higher proximity to continental margins, explains why deep-sea species show maximum richness at higher (30–50°) latitudes (Woolley et al. 2016). Diversity can be increased in high productivity regimes, but this is not always the case, as high organism abundance induces high competition levels and low oxygen concentrations (Levin, L. A. et al. 2001; Rosli et al. 2017).

Effect of depth

Studies have indicated depth to be a main factor influencing the distribution of deep-sea organisms, mainly due to the depth-related decrease in productivity (Levin, L. A. et al. 2001; Olu et al. 2010; Bik et al. 2012b).

Qualitative (Rex 1981) and quantitative (Etter and Grassle 1992) studies in the North Atlantic indicated that diversity-depth patterns in the deep-sea benthos are unimodal, with a peak in diversity at intermediate depths (300 – 4,700 m). The depth of the peak was found to decrease with size class, megafauna showing a diversity peak at ~1,900-2,300 m, while metazoan meiofauna diversity peaked at 3,000 m and foraminiferans showed highest diversity at > 4,000 m (Rex 1981). However, unimodal patterns are not universal, vary regionally with environmental gradients or oceanographic conditions, and between taxa (Levin, L. A. et al. 2001; Ramirez-Llodra et al. 2010; Costello and Chaudhary 2017). The latest global study, including data from 243,000 species catalogued in WoRMS confirmed a peak in species richness at 400–500 m depth (Costello and Chaudhary 2017), a figure in accordance with the patterns found for megafauna in the Atlantic, and reflecting the bias of the database towards large size classes.

Ecological patterns at habitat to small spatial scales

Habitat-scale (100 m-100 km) influences

Generally, areas with greater variation in environmental and topographical conditions support more species and thus exhibit higher regional diversity, explaining why macrohabitat heterogeneity contributes significantly to diversity on a global scale (Vanreusel et al. 2010a; Rosli et al. 2017). Current regimes, although generally low in the abyss, can rise locally due to seafloor topography and influence benthic assemblages (Stefanoudis et al. 2016; Levin, L. A.

et al. 2001). Variations in current regimes and topography, including biogenic structures (Buhl-Mortensen et al. 2010), can create favourable environmental conditions supporting enhanced meiofauna abundances and biomasses, by influencing the amount of organic matter accumulating on the seafloor (Zeppilli et al. 2016; Rosli et al. 2017). This partly explains higher species abundance and biomass in canyons and around seamounts. Similarly, hadal trenches also concentrate food particles, and are thus associated with surprisingly high abundance of meio- and nanofauna (Schmidt and Martínez Arbizu 2015; Zeppilli et al. 2018).

These habitat-specific environmental conditions are unique and distinct from adjacent regions of the ocean floor, leading to the presence of specific taxa adapted to these particular environments (Zeppilli et al. 2012; 2013; 2011; Ramirez-Llodra et al. 2010). Subsurface deposit-feeders (e.g., polychaetes, echiurans) dominate on organic-rich margin sediments; surface deposit-feeders prevail on the oligotrophic abyssal seafloor (e.g., holothurians, other polychaetes, asteroids), suspension feeders (e.g., corals, sponges, crinoids, ascidians) dominate in habitats where currents are stronger like on rocky slopes of seamounts, canyons, ridges, and banks. Taxa living exclusively in canyons (tanaids, echinoid larvae) as well as specific morphological adaptations to cope with increased current regimes in foraminiferans (agglutinated *vs*. organic-welled) or in nematodes (Zeppilli et al. 2018; Rosli et al. 2017) have also been reported. Similarly, chemoautotrophy is the main feeding mode in reducing ecosystems (vents, seeps, and food falls), which are thus associated to symbiotic taxa. Kinorhynchs are particularly abundant at cold seeps and other habitats that undergo drastic changes in salinity. Hypoxic or anoxic environments typical of OMZs or DHABs are associated to a decrease of species abundance and biomass, with the exception of nematodes and loriciferans that are particularly well-adapted to low oxygen conditions (Zeppilli et al. 2018).

Finally, sediment granulometry is known to affect deep-sea benthic diversity, and particle size heterogeneity has been shown to be positively correlated to species diversity (Etter and Grassle 1992; Leduc et al. 2012b). Significant differences in meiofauna abundance have been reported between hemipelagic *vs*. turbidite sediments, and these were related to median grain size and percent content of silt-clay particles (Woods and Tietjen 1985).

Influence of local (10 cm- 100 m) to small (1-10 cm) scale factors

Any habitat-scale factors mentioned above that can vary also at a local scale are likely to influence benthic community composition. Primarily, substratum type greatly affects benthic assemblages as it defines species composition and influences spatial variability in species

distribution (Ramirez-Llodra et al. 2010). In addition, high community dissimilarities can occur at local-scale due to the occurrence of different sub-habitats within a sampling site (Gaever, Van et al. 2010). This can also affect local-scale species abundance due to strong food patchiness, particularly in seep and vent habitats (Rosli et al. 2017). In contrast, little local variation has been reported in terms of species abundance and diversity for heterotrophic sediments, particularly in taxa with locomotory abilities (Woods and Tietjen 1985; Rosli et al. 2017).

Energy availability within sediments is positively correlated with sediment-community respiration, rate of organic carbon burial within the sediment, and benthic biomass and abundance (Levin, L. A. et al. 2001). Organic matter input and oxygen levels, sediment particle size and heterogeneity, as well as bioturbation by larger organisms are all factors influencing the amount of nutrients available within the sediment and can therefore affect both community diversity and composition (Lambshead et al. 1995; Rosli et al. 2017). Substantial small-scale horizontal and vertical variation in benthic assemblages have thus long been reported in both macro- and meiofauna size compartments (Rex 1981). Considerable vertical zonation has been reported in meiofauna communities of deep-sea sediments worldwide (Thiel 1983; Danovaro et al. 1995; Gallucci et al. 2009; Rosli et al. 2016). In all studies, most meiofauna organisms were located in the upper 3 cm of sediment, but some found organisms up to 10 cm or 30 cm within the sediment (Snider et al. 1984; Danovaro et al. 1995; Shirayama 1984). Assemblages were found to vary between sediment layers, and this generally reflects the taxa's ability to cope with lower oxygen concentrations. For example, crustaceans, mainly harpacticoid copepods and ostracods, are most abundant in upper sediment layers due to their increased oxygen consumption, while nematodes cope well with low oxygen concentrations and can thus penetrate deeper into the sediment (Thiel 1983; Tietjen 1992). Species-interactions such as avoidance of predators, or competitive exclusion are also thought to play a role in the vertical distribution patterns observed (Lambshead et al. 1995; Steyaert et al. 2003; Gallucci et al. 2008). Some authors even reported that sediment depth had a greater influence on meiofauna communities than horizontal factors such as sampling stations or habitats (Rosli et al. 2016; Górska et al. 2014).

I.VI. Aims and objectives

While ocean exploration is relatively recent, studies in the last decades have started shedding light on biodiversity and biogeography in the deep-sea realm. However, these studies were confronted with the extraordinary vastness of deep-sea ecosystems, the difficulty of sampling in these remote and high-pressure locations, as well as the high costs and time involved in collecting and analysing samples.

Analytical methods based on extrapolation from known samples clearly indicated that deep-sea life is undeniably diverse, although estimates remain highly uncertain, primarily due to under-sampling and to the difficulty of identifying specimens. The large marine databases assembled in recent years include too little information about deep-sea species in order to make extrapolation approaches a useful tool for the estimation of deep-sea biodiversity (Ramirez-Llodra et al. 2010). Experimental approaches also highlight the strong link between surface and deep-ocean regions, showing that benthic deep-sea communities are affected by climate-driven variations in carbon cycles and can therefore directly influence carbon remineralisation and sequestration processes (Smith, K. L. et al. 2009; 2013). However, monitoring these surface-driven changes in deep-sea benthic communities is costly and difficult to sustain over long-term periods.

Deep-sea sedimentary habitats cover more than 50% of the Earth's surface, can host high numbers of organisms (50,000-5 million individuals per square meter), which perform key ecosystem roles such as nutrient cycling, sediment stabilisation and transport, or secondary production (Bik et al. 2012b; Fonseca, V. G. et al. 2010). Despite this, they are under increased threat from a variety of ongoing or forecasted human activities, ranging from climate change-induced indirect threats due to modifications in ocean biogeochemistry to direct threats from activities such as waste disposal, pollution, or resource exploitation (Ramirez-Llodra et al. 2010; Smith, C. R. et al. 2008; Ramirez-Llodra et al. 2011). Better knowledge of deep-sea biodiversity patterns and the development of deep-sea biomonitoring protocols are therefore becoming necessary in order to preserve this vast and elusive backyard. This PhD thus aims at bringing new perspectives to the study of biodiversity in deep-sea sediments to bridge this knowledge gap.

Environmental DNA metabarcoding approaches have revolutionized biodiversity research in the past decade and have already been successfully applied in marine sedimentary habitats (Pawlowski et al. 2011; Bik et al. 2012b; Fonseca, V. G. et al. 2010; 2014; Cowart et al. 2015; Sinniger et al. 2016; Forster et al. 2016; Cordier et al. 2019a). They represent useful tools for

increasing the spatial scale of deep-sea studies, while allowing to target biodiversity of various biological compartments in parallel, including the commonly overlooked meio- and nanofauna.

However, while this tool greatly facilitates the study of remote ecosystems, many challenges remain to be resolved in order to apply eDNA methods on a broad scale (Cristescu & Hebert 2018). In particular, the use of eDNA to assess metazoan biodiversity remains complex due to the difficulty in defining accurate species-level molecular Operational Taxonomic Units (OTUs) and improvements in the bioinformatic data processing are necessary to achieve more accurate and reliable biodiversity inventories. Moreover, the accuracy of protocols based on eDNA in deep sea sediments still needs to be assessed, as analysis outcomes may be biased by ancient (archived) DNA (aDNA), resulting in biodiversity assessments not targeting *live* organisms.

Objectives

The first, primarily technical aims of this thesis are thus to help developing accurate eDNA metabarcoding protocols for the study of deep-sea biodiversity across multiple life compartments, i.e. prokaryotes, unicellular eukaryotes, and metazoans. Using mitochondrial and nuclear marker genes, the eDNA workflow for deep sea sediments was evaluated and optimized on a bioinformatic and molecular processing level:

1. In order to limit the pitfalls regarding the number of molecular entities, the **second chapter** of this thesis thus describes how newly developed bioinformatic tools were assessed and combined in order to get more reliable biodiversity inventories, approaching a 1:1 species-OTU correspondence.

2. The **third chapter** details the assessment of the potential bias of aDNA through 1) the evaluation of the effect of removing short DNA fragments via size-selection or ethanol reconcentration, and 2) the comparison of communities revealed by co-extracted DNA and RNA in five deep-sea sites.

3. The **fourth chapter** assesses sampling techniques for deep-sea sediment and water in order to define optimal ways to achieve most comprehensive biodiversity inventories, and evaluates whether aboveground water and sediment samples yield comparable communities.

Finally, the **fifth chapter** of this thesis shows the application of these optimized eDNA metabarcoding protocols on deep seafloor of the Atlantic-Mediterranean transition zone. The influence of local abiotic factors on deep-sea benthic metazoan OTU richness and community structure are evaluated at the local, habitat, and regional scales, along this west east transect ranging from the Western North Atlantic to the Ionian Sea.

Chapter II. Bioinformatic pipelines combining correction and clustering tools allow for flexible and comprehensive prokaryotic and eukaryotic metabarcoding

Published as:

Brandt MI, Trouche B, Quintric L, Wincker P, Poulain J, and Arnaud-Haond S. A flexible pipeline combining clustering and correction tools for prokaryotic and eukaryotic metabarcoding. BioRxiv, 717355, ver. 3 peer-reviewed and recommended by PCI Ecology. DOI: 10.1101/717355.

In revision at Molecular Ecology Resources:

Brandt MI, Trouche B, Günther B, Quintric L, Wincker P, Poulain J, and Arnaud-Haond S. Bioinformatic pipelines combining correction and clustering tools allow more flexible and comprehensive prokaryotic and eukaryotic metabarcoding.

A significant source of error in molecular biodiversity inventories of metazoans is due to the fact that metazoans are multicellular organisms, and the marker genes targeted for metabarcoding are present in multiple copies per cell (Krehenwinkel et al. 2017). Thus, sequencing errors, amplification errors, and mutations of marker genes within organisms lead to the fact that single species and even single individuals produce several Operational Taxonomic Units (OTUs). As OTUs are used as a *proxy* for species (as defined by morphological criteria), it is essential that this proxy remains valid to maintain the reliability of metabarcoding inventories.

Metabarcoding bioinformatic pipelines have been in constant refinement, and recent advances have produced new Illumina sequence correction (Callahan et al. 2016) and cluster filtering (Frøslev et al. 2017) tools.

Clustering sequences also alleviates the noise originating from errors and intraspecific variation, as it pools similar but not identical sequences. New clustering methods now allow highly scalable and fine-scale clustering (Mahé et al. 2015), avoiding imposing a "universal" clustering threshold on metabarcoding datasets.

In this chapter, we implement these new tools in a bioinformatic pipeline and assess the level of diversity they allow describing by evaluating their performance on mock communities and deep-sea sediment samples.

Question addressed:

Do new bioinformatic tools such as DADA2, LULU, and swarm v2 allow achieving biodiversity inventories at the level of the morphospecies ?

Résumé en français

Le metabarcoding par ADN environnemental (ADNe) est un outil puissant pour étudier la biodiversité. Cependant, les approches bioinformatiques doivent s'adapter à la diversité des compartiments taxonomiques ciblés ainsi qu'aux spécificités de chaque gène marqueur. Nous avons construit et testé un pipeline basé sur la correction de séquences avec DADA2 permettant d'analyser des données de métabarcoding de compartiments de vie procaryotes (16S) et eucaryotes (18S, COI). Nous avons implémenté l'option de regrouper les variants de séquence d'amplicon (ASV) en unités taxonomiques opérationnelles (OTU) avec swarm, un algorithme de clustering basé sur l'analyse des réseaux, et la possibilité de filtrer les ASV / OTU à l'aide de LULU. Enfin, l'assignation taxinomique a été mise en place via le classificateur bayésien du Ribosomal Database Project (RDP) et BLAST. Nous évaluaons ce pipeline avec des marqueurs ribosomaux et mitochondriaux à l'aide de communautés métazoaires connues et de 42 échantillons de sédiments abyssaux. Les résultats montrent que les ASV et les OTU décrivent différents niveaux de diversité biotique, dont le choix dépend des questions de recherche. Ils soulignent les avantages et la complémentarité du clustering et de la filtration avec LULU pour produire des inventaires de la biodiversité métazoaire à un niveau proche de celui obtenu à partir de critères morphologiques. Alors que le clustering supprime la variation intraspécifique, LULU supprime efficacement les unités génétiques erronées, provenant d'erreurs techniques ou de variabilité intragénomique. Le clustering a affecté la diversité alpha et bêta différemment selon le marqueur génétique. Plus précisément, les valeurs de swarm à d > 1 se sont avérées moins appropriées avec 18S pour les métazoaires. De même, augmenter le niveau du *minimum ratio* de LULU s'est avéré essentiel pour éviter de perdre des espèces dans des jeux de données pauvres en échantillons. La comparaison de BLAST et de RDP a souligné que des assignations taxonomiques précises peuvent être obtenues pour les espèces d'eau profonde avec RDP, mais a souligné la nécessité d'un effort concerté pour créer des bases de données complètes et spécifiques à l'écosystème.

Abstract

Environmental metabarcoding is a powerful tool for studying biodiversity. However, bioinformatic approaches need to adjust to the diversity of taxonomic compartments targeted as well as to each barcode gene specificities. We built and tested a pipeline based on read correction with DADA2 allowing analysing metabarcoding data from prokaryotic (16S) and eukaryotic (18S, COI) life compartments. We implemented the option to cluster Amplicon Sequence Variants (ASVs) into Operational Taxonomic Units (OTUs) with swarm, a network-based clustering algorithm, and the option to curate ASVs/OTUs using LULU. Finally, taxonomic assignment was implemented via the Ribosomal Database Project Bayesian classifier (RDP) and BLAST. We validate this pipeline with ribosomal and mitochondrial markers using metazoan mock communities and 42 deep-sea sediment samples. The results show that ASVs and OTUs describe different levels of biotic diversity, the choice of which depends on the research questions. They underline the advantages and complementarity of clustering and LULU-curation for producing metazoan biodiversity inventories at a level approaching the one obtained using morphological criteria. While clustering removes intraspecific variation, LULU effectively removes spurious clusters, originating from errors or intragenomic variability. Swarm clustering affected alpha and beta diversity differently depending on genetic marker. Specifically, *d*-values > 1 appeared to be less appropriate with 18S for metazoans. Similarly, increasing LULU's *minimum ratio* level proved essential to avoid losing species for sample-poor datasets. Comparing BLAST and RDP underlined that accurate assignments of deep-sea species can be obtained with RDP, but highlighted the need for a concerted effort to build comprehensive, ecosystem-specific databases.

1 Introduction

High-throughput sequencing (HTS) technologies are revolutionizing the way we assess biodiversity. By producing millions of DNA sequences per sample, HTS allows broad taxonomic biodiversity surveys through metabarcoding of bulk DNA from complex communities or from environmental DNA (eDNA) directly extracted from soil, water, and air samples. First developed to unravel cryptic and uncultured prokaryotic diversity, metabarcoding methods have been extended to eukaryotes as powerful, non-invasive tools, allowing detection of a wide range of taxa in a rapid, cost-effective way using a variety of sample types (Valentini et al. 2009; Taberlet et al. 2012a; Creer et al. 2016; Stat et al. 2017). In the last decade, these tools have been used to describe past and present biodiversity in terrestrial (Ji et al. 2013; Yoccoz et al. 2012; Yu et al. 2012; Slon et al. 2017; Pansu et al. 2015), freshwater (Valentini et al. 2016; Deiner et al. 2016; Bista et al. 2015; Dejean et al. 2011; Evans, N T et al. 2016), and marine (Fonseca, V. G. et al. 2010; Sinniger et al. 2016; Pawlowski et al. 2011; Massana et al. 2015; Vargas et al. 2015; Salazar et al. 2016; Boussarie et al. 2018; Bik et al. 2012b) environments.

As every new technique brings on new challenges, a number of studies have put considerable effort into delineating critical aspects of metabarcoding protocols to ensure robust and reproducible results (see Fig.1 in Fonseca 2018). Recent studies have addressed many issues regarding sampling methods (Dickie et al. 2018), contamination risks (Goldberg et al. 2016), DNA extraction protocols (Brannock and Halanych 2015; Deiner et al. 2015; Zinger et al. 2016), amplification biases and required PCR replication levels for improved detection probability (Nichols et al. 2018; Alberdi et al. 2017; Ficetola et al. 2015). Similarly, computational pipelines, through which molecular data are transformed into ecological inventories of putative taxa, have also been in constant improvement. PCR-generated errors and sequencing errors are major bioinformatic challenges for metabarcoding pipelines, as they can strongly bias biodiversity estimates (Coissac et al. 2012; Bokulich et al. 2013). A variety of tools have thus been developed for quality-filtering amplicon data to remove erroneous reads and improve the reliability of Illumina-sequenced metabarcoding inventories (Bokulich et al. 2013; Eren et al. 2013; Minoche et al. 2011). Studies that evaluated bioinformatic processing steps have generally found that sequence quality-filtering parameters and clustering thresholds most strongly affect molecular biodiversity inventories, resulting in considerable variation during data analysis (Brannock and Halanych 2015; Clare et al. 2016; Brown et al. 2015; Xiong and Zhan 2018).

There were historically two main reasons for clustering sequences into Operational Taxonomic Units (OTUs). The first was to limit the bias due to PCR, sequencing errors, and intragenomic variability (e.g. pseudogenes) by clustering erroneous sequences with error-free target sequences. The second was to delineate OTUs as clusters of homologous sequences (by grouping the alleles/haplotypes at the same locus) that would best fit a "species level", i.e. the Operational Taxonomic Units defined using a classical phenetic *proxy* (Sokal and Crovello 1970). Recent bioinformatic algorithms alleviate the influence of errors and intraspecific variability in metabarcoding datasets. First, amplicon-specific error correction methods, commonly used to correct sequences produced by pyrosequencing (Coissac et al. 2012), have now become available for Illumina-sequenced data. Introduced in 2016, DADA2 effectively corrects Illumina sequencing errors and has quickly become a widely used tool, particularly in the microbial world, producing more accurate biodiversity inventories and resolving fine-scale genetic variation by defining Amplicon Sequence Variants (ASVs) (Callahan et al. 2016; Nearing et al. 2018). Second, LULU is a recently developed curation algorithm designed to filter out spurious clusters, originating from PCR and sequencing errors, or intra-individual variability (pseudogenes, heteroplasmy), based on their similarity (*minimum match*) and co-occurrence rate (*minimum relative cooccurence*) with more abundant clusters, allowing the acquisition of curated datasets while avoiding arbitrary abundance filters (Frøslev et al. 2017). The authors validated their approach on metabarcoding of plants using ITS2 (nuclear ribosomal internal transcribed spacer region 2) and evaluated it on several pipelines. Their results show that ASV definition with DADA2, subsequent clustering to address intraspecific variation, and final curation with LULU is the safest pathway for producing reliable and accurate metabarcoding data. The authors concluded that their validation on plants is relevant to other organism groups and other markers, while recommending future validation of LULU on mock communities as LULU's *minimum match* parameter may need to be adjusted to less variable marker genes.

The impact of errors being strongly decreased by correction algorithms such as DADA2 and LULU, the relevance of clustering sequences into OTUs is now being debated. Indeed, after presenting their new algorithm on prokaryotic communities, the authors of DADA2 proposed that the reproducibility and comparability of ASVs across studies challenge the need for clustering sequences, as OTUs have the disadvantage of being study-specific and defined using arbitrary thresholds (Callahan et al. 2017). Yet, clustering sequences may still be necessary in metazoan datasets, where very distinct levels of intraspecific polymorphism can exist in the

same gene region among taxa, due to both evolutionary and biological specificity (Bucklin et al. 2011; Phillips et al. 2019). ASV-based inventories will thus be biased in favour of taxa with high levels of intraspecific diversity, even though these are not necessarily the most abundant ones (Bazin et al. 2006). Such bias is magnified with presence-absence data, commonly used for metazoan metabarcoding (Ji et al. 2013). However, as intraspecific polymorphism and interspecific divergence are phylum-specific, imposing a universal clustering threshold on metabarcoding datasets is also introducing bias, penalizing groups with lower polymorphism or divergence levels, while overestimating species diversity in groups with higher interspecific divergence. Universal clustering thresholds can be avoided with tools such as swarm v2, a single-linkage clustering algorithm (Mahé et al. 2015), implemented in recent bioinformatic pipelines, such as FROGS (Escudié et al. 2018) or SLIM (Dufresne et al. 2019). Based on network theory, swarm v2 aggregates sequences iteratively and locally around seed sequences, based on *d*, the number of nucleotide differences, to determine coherent groups of sequences, independent of amplicon input order, allowing highly scalable and fine-scale clustering. Finally, it is widely recognized that homogeneous entities sharing a set of evolutionary and ecological properties, i.e. *species* (Mayr 1942; Queiroz, de 2005), sometimes referred to "ecotypes" for prokaryotes (Cohan 2001; Gevers et al. 2005), represent a fundamental category of biological organization that is the cornerstone of most ecological and evolutionary theories and empirical studies. Maintaining ASV information for feeding databases and cross-comparing studies is not incompatible with their clustering into OTUs, and this choice depends on the purpose of the study, i.e. providing a census of the extent and distribution of genetic polymorphism for a given gene, or a census of biodiversity to be used and manipulated in ecological or evolutionary studies.

Here we evaluate DADA2 and LULU, using them alone and in combination with swarm v2, to assess the performance of these new tools for metabarcoding of metazoan communities. Using both mitochondrial COI (Leray et al. 2013) and the V1-V2 region of 18S ribosomal RNA (rRNA) (Sinniger et al. 2016), we evaluated the need for clustering and the effectiveness of LULU curation to select pipeline parameters delivering the most accurate resolution of two deep-sea mock communities. We then test the different bioinformatic tools on a deep-sea sediment dataset in order to select an optimal trade-off between inflating biodiversity estimates and loosing rare biodiversity. As a baseline for comparison, and in the perspective of the joint study of metazoan and microbial taxa, we also analysed the 16S V4-V5 rRNA barcode (Parada et al. 2016) on these environmental samples.

Our objectives were to (1) discuss the use of ASV *vs* OTU-centred datasets depending on taxonomic compartment and study objectives, and (2) determine the most adequate swarm-clustering and LULU curation thresholds that avoid inflating biodiversity estimates while retaining rare biodiversity.

2 Materials and methods

2.1 Preparation of samples

Mock communities

Two genomic-DNA mass-balanced metazoan mock communities (5 ng/µL) were prepared using standardized 10 ng/µL DNA extracts of ten deep-sea specimens belonging to five taxonomic groups (Polychaeta, Crustacea, Anthozoa, Bivalvia, Gastropoda; Table S1). Specimen DNA was extracted using a CTAB extraction protocol, from muscle tissue or from whole polyps in the case of cnidarians. The mock communities differed in terms of ratios of total genomic DNA from each species, with increased dominance of three species and secondary species DNA input decreasing from 3% to 0.7%. We individually barcoded the species present in the mock communities: PCRs of both target genes were performed using the same primers as the ones used in metabarcoding (see below). The PCR reactions (25 µL final volume) contained 2 µL DNA template with 0.5 µM concentration of each primer, 1X *Phusion* Master Mix, and an additional 1 mM $MgCl_2$ for COI. PCR amplifications (98 °C for 30 s; 40 cycles of 10 s at 98 °C, 45 s at 48 °C (COI) or 57 °C (18S), 30 s at 72 °C; and 72 °C for 5 min) were cleaned up with ExoSAP (Thermo Fisher Scientific, Waltham, MA, USA) and sent to Eurofins (Eurofins Scientific, Luxembourg) for Sanger sequencing. The barcode sequences obtained for all mock specimens were added to the databases used for taxonomic assignments of metabarcoding datasets, and were submitted on Genbank under accession numbers MN826120-MN826130 and MN844176-MN844185.

Environmental DNA

Sediment cores were collected from fourteen deep-sea sites ranging from the Arctic to the Mediterranean during various cruises (Table S2). Sampling was carried out with a multicorer or with a remotely operated vehicle. Three tube cores were taken at each sampling station (GPS coordinates in Table S2). The latter were sliced into depth layers that were transferred into zip-lock bags, homogenised, and frozen at −80°C on board before being shipped on dry ice to the

laboratory. The first layer (0-1 cm) was used in the present study. DNA extractions were performed using approximately 10 g of sediment with the PowerMax Soil DNA Isolation Kit (Qiagen, Hilden, Germany). To increase the DNA yield, the elution buffer was left on the spin filter membrane for 10 min at room temperature before centrifugation. The ~5 mL extract was then split into three parts, one of which was kept in screw-cap tubes for archiving purposes and stored at -80°C. For the four field controls, the first solution of the kit was poured into the control zip-lock bag, before following the usual extraction steps. For the two negative extraction controls, a blank extraction (adding nothing to the bead tube) was performed alongside sample extractions.

2.2 Amplicon library construction and high-throughput sequencing

Two primer pairs were used to amplify the mitochondrial COI and the 18S V1-V2 rRNA barcode genes specifically targeting metazoans, and one pair of primer was used to amplify the prokaryote 16S V4-V5 region. PCR amplifications, library preparation, and sequencing were carried out at Genoscope (Evry, France) as part of the eDNAbyss project. Four (16S), eight (18S), and ten (COI) control PCRs were performed alongside sample PCRs, depending on the amount of trials needed to achieve successful amplification.

Eukaryotic 18S V1-V2 rRNA gene amplicon generation

Amplifications were performed with the *Phusion* High Fidelity PCR Master Mix with GC buffer (Thermo Fisher Scientific, Waltham, MA, USA) and the SSUF04 (5'-GCTTGTCTCAAAGATTAAGCC-3') and SSUR22*mod* (5'- CCTGCTGCCTTCCTTRGA-3') primers (Sinniger et al. 2016), preferentially targeting metazoans, the primary focus of this study. The PCR reactions (25 µL final volume) contained 2.5 ng or less of DNA template with 0.4 µM concentration of each primer, 3% of DMSO, and 1X *Phusion* Master Mix. Three PCR replicates (98 °C for 30 s; 25 cycles of 10 s at 98 °C, 30 s at 45 °C, 30 s at 72 °C; and 72 °C for 10 min) were performed in order to smooth the intra-sample variance while obtaining sufficient amounts of amplicons for Illumina sequencing.

Eukaryotic COI gene amplicon generation

Metazoan COI barcodes were generated using the mlCOIintF (5'-GGWACWGGWTGAACWGTWTAYCCYCC-3') and jgHCO2198 (5'-

TAIACYTCIGGRTGICCRAARAAYCA-3') primers (Leray et al. 2013). Triplicate PCR reactions (20 µl final volume) contained 2.5 ng or less of total DNA template with 0.5 µM final concentration of each primer, 3% of DMSO, 0.175 mM final concentration of dNTPs, and 1X Advantage 2 Polymerase Mix (Takara Bio, Kusatsu, Japan). Cycling conditions included a 10 min denaturation step followed by 16 cycles of 95 °C for 10 s, 30s at 62°C (−1°C per cycle), 68 °C for 60 s, followed by 15 cycles of 95 °C for 10 s, 30s at 46°C, 68 °C for 60 s and a final extension of 68 °C for 7 min.

Prokaryotic 16S rRNA gene amplicon generation

Prokaryotic barcodes were generated using 515F-Y (5'- GTGYCAGCMGCCGCGGTAA-3') and 926R (5'- CCGYCAATTYMTTTRAGTTT-3') 16S-V4V5 primers (Parada et al. 2016). Triplicate PCR reactions were prepared as described above for 18S-V1V2, but cycling conditions included a 30 s denaturation step followed by 25 cycles of 98 °C for 10 s, 53 °C for 30 s, 72 °C for 30 s, and a final extension of 72 °C for 10 min.

Amplicon library preparation

PCR triplicates were pooled and PCR products purified using 1X AMPure XP beads (Beckman Coulter, Brea, CA, USA) clean up. Aliquots of purified amplicons were run on an Agilent Bioanalyzer using the DNA High Sensitivity LabChip kit (Agilent Technologies, Santa Clara, CA, USA) to check their lengths and quantified with a Qubit fluorimeter (Invitrogen, Carlsbad, CA, USA). One hundred nanograms of pooled amplicon triplicates were directly end-repaired, A-tailed and ligated to Illumina adapters on a Biomek FX Laboratory Automation Workstation (Beckman Coulter, Brea, CA, USA). Library amplification was performed using a Kapa Hifi HotStart NGS library Amplification kit (Kapa Biosystems, Wilmington, MA, USA) with the same cycling conditions applied for all metagenomic libraries and purified using 1X AMPure XP beads.

Sequencing library quality control

Amplicon libraries were quantified by Quant-iT dsDNA HS assay kits using a Fluoroskan Ascent microplate fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and then by qPCR with the KAPA Library Quantification Kit for Illumina Libraries (Kapa Biosystems, Wilmington, MA, USA) on an MxPro instrument (Agilent Technologies, Santa Clara, CA,

USA). Library profiles were assessed using a high-throughput microfluidic capillary electrophoresis system (LabChip GX, Perkin Elmer, Waltham, MA, USA).

Sequencing procedures

Library concentrations were normalized to 10 nM by addition of 10 mM Tris-Cl (pH 8.5) and applied to cluster generation according to the Illumina Cbot User Guide (Part # 15006165). Amplicon libraries are characterized by low diversity sequences at the beginning of the reads due to the presence of the primer sequence. Low-diversity libraries can interfere in correct cluster identification, resulting in a drastic loss of data output. Therefore, loading concentrations of libraries were decreased (8–9 pM instead of 12–14 pM for standard libraries) and PhiX DNA spike-in was increased (20% instead of 1%) in order to minimize the impacts on the run quality.

Libraries were sequenced on HiSeq2500 (System User Guide Part # 15035786) instruments (Illumina, San Diego, CA, USA) in a 250 bp paired-end mode.

2.3 Bioinformatic analyses

All bioinformatic analyses were performed using a Unix shell script run on a home-based cluster (DATARMOR, Ifremer). The script is available on Gitlab (https://gitlab.ifremer.fr/abyss-project/) and is based on DADA2 v.1.10 (Callahan et al. 2016) and FROGS (Escudié et al. 2018) as core processing tools. It allows the use of sequence data obtained from libraries produced by double PCR or adaptor ligation methods, as well as having built-in options for using six commonly used metabarcoding primers.

For all analyses, the mock communities were analysed alongside all environmental samples, and used to validate the metabarcoding pipeline in terms of detection of correct species and presence of false-positives. The details of the pipeline, along with specific parameters used for all three metabarcoding markers are listed in Table S3.

Reads preprocessing

Our multiplexing strategy relies on ligation of adapters to amplicon pools, meaning that contrary to libraries produced by double PCR, the reads in each paired sequencing run can be forward or reverse. DADA2 correction is based on error distribution differing between R1 and R2 reads. We thus developed a custom script (*abyss-preprocessing* in abyss-pipeline) allowing separating forward and reverse reads in each paired run and reformatting the outputs to be

compatible with DADA2. Briefly, the script uses cutadapt v1.18 to detect and remove primers, while separating forward and reverse reads in each paired sequence file to produce two pairs of sequence files per sample named R1F/R2R and R2F/R1R. Cutadapt parameters (Table S3) were set to require an overlap over the full length of the primer (default: 3 nt), with 2-4 nt mismatches allowed for ribosomal loci, and 7 nt mismatches allowed for COI (default: 10%). Each identified forward and reverse read is then renamed which the correct extension (/1 and /2 respectively), which is a requirement for DADA2 to recognize the pairs of reads. Each pair of renamed sequence files is then re-paired with BBMAP Repair v38.22 in order to remove singleton reads (non-paired reads). Optionally, sequence file names can also be renamed if necessary using a CSV correspondence file.

Read correction, amplicon cluster generation and taxonomic assignment

Pairs of Illumina reads were corrected with DADA2 following the online tutorial for paired-end HiSeq data (https://benjjneb.github.io/dada2/bigdata_paired.html). Reads containing ambiguous bases removed and trimming lengths were adjusted based on sequence quality profiles, so that Q-scores remained above 30 (truncLen at 220 for 18S and 16S, 200 for COI, maxEE at 2, truncQ at 11, maxN at 0). Error model calculation (for R1F/R2R read pairs and then R2F/R1R read pairs), read correction, and read merging was performed at default settings. Amplicons were filtered by size, with size ranges set to 330-390 bp for the 18S SSU rRNA marker gene, 300-326 bp for the COI marker gene, and 350-390 bp for the 16S rRNA marker gene, based on raw size distributions observed. Chimera removal and taxonomic assignment were performed with default methods implemented in DADA2.

A second taxonomic assignment method was optionally implemented in the pipeline, allowing assigning ASVs using BLAST+ (Basic Local Alignment Search Tool v2.6.0) based on minimum similarity and minimum coverage (-perc_identity 70 and –qcov_hsp 80). An initial test implementing BLAST+ to assign taxonomy only to the COI dataset using a 96% percent identity threshold led to the exclusion of the majority of the clusters. Given observed inter-specific mitochondrial DNA divergence levels of up to 30% within a same polychaete genus (Zanol et al. 2010) or among some closely related deep-sea shrimp species (Shank et al. 1999), and considering our interest in the identities of multiple, largely unknown taxa in poorly characterized communities, more stringent BLAST thresholds were not implemented at this stage. However, additional filters were performed during downstream processing described below, and only clusters with assignments reliable at phylum-level were retained in the analysis.

The Silva132 reference database was used for 16S and 18S SSU rRNA marker genes (Quast et al. 2012), and MIDORI-UNIQUE (Machida et al. 2017) was used for COI. The databases were downloaded from the DADA2 website (https://benjjneb.github.io/dada2/training.html) and from the FROGS website (http://genoweb.toulouse.inra.fr/frogs_databanks/assignation). Finally, to evaluate the effect of swarm clustering, ASV tables were clustered with swarm v2 (Mahé et al. 2015) in FROGS (http://frogs.toulouse.inra.fr/) at *d*-values (i.e. nucleotide differences) ranging from 1 to 13 (*d = 1, 3, 4, 5, 11* for 18S/16S, and *d = 1, 5, 6, 7, 13* for COI), based on settings previously used in the literature (Clare et al. 2016; Atienza et al. 2020; Turon et al. 2020; Djurhuus et al. 2017; Cordier et al. 2019a; Sawaya et al. 2019; Wood et al. 2019; Laroche et al. 2018; Andújar et al. 2018a). Resulting OTUs were chimera-filtered and taxonomically assigned via RDP and BLAST+ with the databases stated above, using standard FROGS procedures.

Molecular clusters were refined in R v.3.5.1 (R Core Team 2018). A blank correction was made using the *decontam* package v.1.2.1 (Davis et al. 2018), removing all clusters that were prevalent (more frequent) in negative control samples. ASV/OTU tables were refined based on their BLAST or RDP taxonomy. For both assignment methods, clusters unassigned at phylum-level were removed. With BLAST, assigned clusters represented 33% of COI data, 76% of 18S data, and 97% of 16S data. With RDP, assigned clusters represented 95-99% of data. Non-target clusters (i.e. either non-metazoan or non-bacterial) were removed. Additionally, for metazoans, clusters with terrestrial assignments (taxonomic groups known to be terrestrial-only, such as Insecta, Arachnida, Diplopoda, Amphibia, terrestrial mammals, Stylommatophora, Aves, Onychophora, Succineidae, Cyclophoridae, Diplommatinidae, Megalomastomatidae, Pupinidae, Veronicellidae) were removed. Samples were checked to ensure that a minimum of 10,000 reads were left after refining. Finally, as tag-switching is to be expected in multiplexed metabarcoding analyses (Schnell et al. 2015), an abundance renormalization was performed to remove spurious positive results due to reads assigned to the wrong sample (Wangensteen and Turon 2016), the original R script being available at https://github.com/metabarpark/R_scripts_metabarpark.

To test LULU curation (Frøslev et al. 2017), refined 18S and COI ASVs/OTUs were curated with LULU v.0.1 following the online tutorial (https://github.com/tobiasgf/lulu). The LULU algorithm detects erroneous clusters by comparing their sequence similarity and co-occurrence rate with more abundant ("parent") clusters. LULU was applied on the full dataset (mock and environmental samples) with a minimum relative co-occurrence of 0.95 (default),

using a minimum similarity threshold (*minimum match*) at 84% (default) and slightly higher at 90%, following recommendations of the authors for less variable loci than ITS. The design of the mock samples was not ideal to test LULU, as some mock species were not occurring (or rarely occurring) in environmental samples, but all species were always co-occurring in the mock samples and this at consistent abundance ratios. With the *minimum ratio* parameter at the default value of 1, this led to the loss of closely related but true mock species for 18S, due to random amplification biases leading to consistent read abundance patterns. In order to remove only errors and avoid losing true mock species, we thus tested *minimum ratio* at 100 and 1000, which allows removing only clusters that are 100/1000 times less abundant than a potential parent OTU.

The vast majority of prokaryotes usually show low levels (< 1%) of intra genomic variability for the 16S SSU rRNA gene (Acinas et al. 2004; Pei et al. 2010). These low intragenomic divergence levels can be efficiently removed with swarm clustering at low *d*-values. Although LULU curation may still be useful to merge redundant phylotypes in specific cases such as haplotype network analyses, this was not tested in this study. Indeed, parallelization not being currently available for LULU curation, the richness of prokaryote communities implied an unrealistic calculation time, even on a powerful cluster (e.g. LULU curation was at 20 - 40% after 4 days of calculation on our cluster).

In order to have reliable BLAST phylum assignments for pipeline comparison, final datasets were taxonomically filtered by retaining only clusters having a minimum hit identity of 86% for rRNA loci and 80% for COI. These values were chosen as they represent approximate minimum identities for reliable phylum assignment (Stefanni et al. 2018).

2.4 Statistical analyses

Data was analysed using R with the packages phyloseq v1.22.3 (McMurdie and Holmes 2013) following guidelines on online tutorials (http://joey711.github.io/phyloseq/tutorials-index.html), and vegan v2.5.2 (Oksanen et al. 2018). The datasets were normalized by rarefaction to their common minimum sequencing depth (COI: 15,575; 18S: 33,916; 16S: 70,474), before analysis of mock communities and environmental samples.

To evaluate the functionality of the bioinformatic tools with the mock communities, taxonomically assigned metazoan clusters were considered as derived from one of the ten species used for the mock communities when the assignment delivered the corresponding

species, genus, family, or class. Clusters not fitting the expected taxa were labelled as 'Others'. These non-target clusters may originate from contamination by external DNA or from DNA of associated microfauna, or gut content in the case of whole polyps used for cnidarians.

Alpha diversity detected using each pipeline in the environmental samples was evaluated with the number of observed clusters in the rarefied datasets via analyses of variance (ANOVA) on generalized linear models based on quasipoisson distribution models. Homogeneity of multivariate dispersions were verified with the *betadisper* function of the *betapart* package v.1.5.1 (Baselga and Orme 2012). The effect of site and LULU curation on community composition was tested by PERMANOVA, using the function *adonis*2 (vegan), with Jaccard incidence dissimilarities for metazoans and Bray-Curtis dissimilarities for prokaryotes, and significance was evaluated by permuting 999 times. Beta-diversity patterns were visualised via non-metric multidimensional scaling (NMDS) using the same dissimilarities stated above.

Finally, BLAST and RDP taxonomic assignments were compared at the most adequate pipeline settings for each locus. BLAST and RDP datasets were compared on ASV-level for prokaryotes, and OTU-level for metazoans (swarm *d=1*, LULU with *minimum match* at 84% and *minimum ratio* at 1 for COI, and 90% and 100 respectively for 18S). As trials on MIDORI-UNIQUE resulted in very poor performance of RDP for COI (assignments belonging mostly to Insecta), the comparison was performed with MIDORI-UNIQUE subsampled to marine taxa only. For the global dataset, full ranges of BLAST hit identities and phylum-level bootstraps were plotted and numbers of clusters left after phylum-level and genus-level quality filtering were calculated, while for evaluation on the mock samples, rarefied data was subsampled to reliable phylum-level assignments (i.e. ≥ 80% / 86% similarity, ≥ 80% phylum-level bootstraps).

3 Results

3.1 Alpha diversity in mock communities

A total of 1.5 million (COI) and 2 million (18S) raw reads were obtained from the two mock communities (Table S4). After refining (decontamination, renormalisation, removal of non-target taxa, and clusters unassigned at phylum-level or with unreliable phylum-level assignments), these numbers were decreased to 0.7 million for COI and 1.3 million for 18S.

All ten mock species were detected in the COI dataset (Table 1), even with minimum relative DNA abundance levels as low as 0.7% (Mock 5). With 18S, seven species were

recovered and the three bivalve species remained unresolved. Taxonomic assignments were correct at the genus-level for six species with COI and three species for 18S, but all mock species produced ASVs/OTUs correctly assigned up to family or class level. Dominant species generally produced more reads in both the clustered and non-clustered datasets, with the notable exception of the gastropod *Paralepetopsis* sp, which was poorly detected with 18S (Table S5).

When ASVs were clustered with swarm v2, this generally led to a reduction in taxonomic recovery: the two bivalves *P. kilmeri* and *C. regab* were taxonomically misidentified with COI at $d \geq 1$ and *Chorocaris* sp. was not detected with 18S at $d > 1$. Clustering ASVs with swarm v2 reduced the number of molecular clusters produced per species, but some species still produced multiple OTUs even at d values as high as $d = 13$ for COI (*D. dianthus, A. muricola, Chorocaris* sp., and *Paralepetopsis* sp.) and $d = 11$ for 18S (*A. arbuscula*, *A. muricola, Munidopsis* sp., and *E. norvegica*).

Curating ASVs/OTUs with LULU allowed reducing the number of clusters produced per species for both loci, and optimal results were obtained in datasets clustered at $d \geq 1$ for COI and $d = 1$ for 18S. The number of unexpected clusters ("Others") was hardly affected by LULU curation (Table 1). In the COI dataset, curating with LULU at 84% or 90% *minimum match* resulted in similar OTU numbers, although 84% performed slightly better in Mock 3 (Table 1). Increasing the *minimum ratio* parameter to 100 or 1000 resulted in the retention of more error OTUs and thus higher OTU numbers in each mock species (data not shown). For 18S, both LULU *minimum match* and *minimum ratio* affected species recovery. LULU curation with *minimum ratio* = 1 led to the loss of the shrimp *Chorocaris* sp. at both *minimum match* levels and the gastropod *Paralepetopsis* sp. at 84% *minimum match* (Table S6). With *minimum ratio* at 100, *Chorocaris* sp. was retained in the dataset at both *minimum match* levels and *Paralepetopsis* sp. with *minimum match* at 90% (Table 1). With *minimum ratio* at 1000, both species were retained at both *minimum match* levels but more OTUs were retained for another species (*Munidopsis* sp., Table S6). As LULU curation with higher *minimum ratio* levels resulted in more accurate species compositions in the mock samples with 18S, we only present LULU curation with *minimum ratio* = 100 for the environmental samples.

Table 1. Number of ASVs/OTUs detected per species in the mock communities using different bioinformatic pipelines. White cells indicate an exact match with the number of OTUs expected (i.e., 1 OTU for each mock species), light grey cells indicate a number of OTUs differing by ±3 from the number expected, dark grey cells indicate a number of OTUs > 3 times the one expected, and black cells a number ≥ 10 times the one expected. Ø indicates absence of expected OTU. Taxonomy is given up to the lowest common rank assigned to OTUs from mock species. "Others" represents unexpected OTUs, i.e. with assignments not related to any species in the mocks. These may represent contamination or symbionts of the mock species. LULU was run at *minimum ratio* = 100 for 18S and *minimum ratio* = 1 for COI.

	COI	DADA2 +	Ø	LULU		swarm					swarm+LULU 90%					swarm+LULU 84%				
		DNA in %		**90%**	**84%**	**d1**	**d5**	**d6**	**d7**	**d13**	**d1**	**d5**	**d6**	**d7**	**d13**	**d1**	**d5**	**d6**	**d7**	**d13**
Mock3	*Acanella arbuscula*	20																		
	Hexacorallia; *D.dianthus*	3																		
	Alvinocaris ; *A. muricola*	3																		
	Chorocaris sp.	3																		
	Munidopsis sp.	3																		
	Gastropoda; *Paralepetopsis* sp.	20																		
	Bivalvia; *C. regab°*	3																		
	Phreagena kilmeri°	3																		
	Vesicomya gigas	3																		
	Polychaeta; *E.norvegica*	40																		
	Others	0																		
Mock5	*Acanella arbuscula*	10																		
	Hexacorallia; *D.dianthus*	0.7																		
	Alvinocaris ; *A. muricola*	0.7																		
	Chorocaris sp.	0.7																		
	Munidopsis sp.	0.7																		
	Gastropoda; *Paralepetopsis* sp.	5																		
	Bivalvia; *C. regab°*	0.7																		
	Phreagena kilmeri°	0.7																		
	Vesicomya gigas	0.7																		
	Polychaeta; *E.norvegica*	80																		
	Others	0																		
	18S			**90%**	**84%**	**d1**	**d3**	**d4**	**d5**	**d11**	**d1**	**d3**	**d4**	**d5**	**d11**	**d1**	**d3**	**d4**	**d5**	**d11**
Mock3	Alcyonacea; *A.arbuscula*	20																		
	Caryophylliidae; *D.dianthus*	3																		
	Alvinocaris muricola	3																		
	Chorocaris sp.	3					Ø	Ø	Ø	Ø		Ø	Ø	Ø	Ø		Ø	Ø	Ø	Ø
	Munidopsis sp.	3																		
	Gastropoda; *Paralepetopsis* sp.	20			Ø											Ø	Ø	Ø	Ø	Ø
	Vesicomyidae; *P. kilmeri/C. regab/V. gigas**	9																		
	Polychaeta; *E.norvegica*	40																		
	Others	0																		
Mock5	Alcyonacea; *A.arbuscula*	10																		
	Caryophylliidae; *D.dianthus*	0.7																		
	Alvinocaris muricola	0.7																		
	Chorocaris sp.	0.7					Ø	Ø	Ø	Ø		Ø	Ø	Ø	Ø		Ø	Ø	Ø	Ø
	Munidopsis sp.	0.7																		
	Gastropoda; *Paralepetopsis* sp.	5			Ø		Ø		Ø	Ø						Ø	Ø	Ø	Ø	Ø
	Vesicomyidae; *P. kilmeri/C. regab/V. gigas**	2.1																		
	Polychaeta; *E.norvegica*	80																		
	Others	0																		

°Bivalvia was common rank for OTUs of *P. kilmeri* and *C. regab* for all pipelines with swarm clustering

*Bivalvia was common rank for all pipelines with *d* > 1

3.2 Alpha-diversity patterns in environmental samples

High-throughput sequencing results

A total of 44 million (18S), 33 million (COI) and 16 million (16S) reads were obtained from 42 sediment samples, 4 field controls, 2 extraction blanks, and 4 (16S), 8 (18S), and 10 (COI) PCR blanks (Table S4). The final datasets contained ~5 million (COI) to ~8 million (18S) marine metazoan target reads and ~7 million prokaryotic 16S reads (Table S4). COI reads produced 13,397 ASVs, 3,518 – 5,563 OTUs after swarm clustering (*d* = 1-13), and 1,758 – 10,028 OTUs after LULU curation (Table S7). Final 18S reads comprised 8,280 ASVs, 1,869 – 6,015 OTUs after swarm clustering (*d* = 1-11), and 1,469 – 6,909 OTUs after LULU curation. The prokaryote dataset produced 53,815 target ASVs and 12,800 – 38,972 OTUs after swarm clustering (*d* = 1-11).

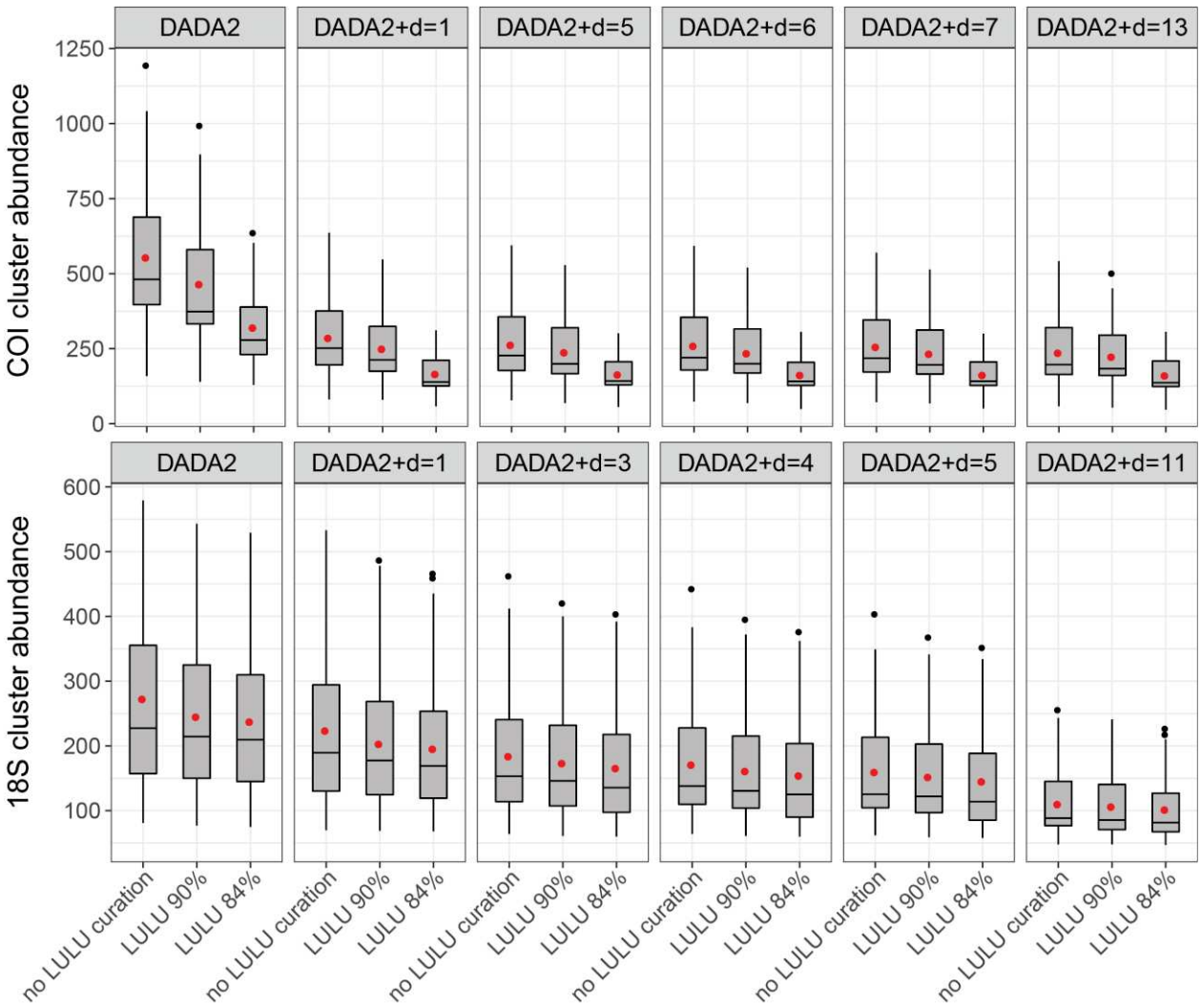
Number of clusters among pipelines

The number of metazoan clusters detected in the deep-sea sediment samples varied significantly with bioinformatic pipeline and site (Table 2). The pipeline effect was consistent across sites (Table 2), although mean cluster numbers detected per sample spanned a wide range in all loci (50 - 500 for 18S, 100 – 1,000 for COI, and 1,500 – 4,000 for 16S, Fig. 1).

As expected, clustering significantly reduced the number of detected molecular clusters per sample for all loci. Consistent to results observed in mock communities, clustering at $d = 1\text{-}13$ resulted in comparable OTU numbers for COI, while significantly higher OTU numbers were obtained at $d = 1$ than with $d > 1$ for ribosomal loci (Fig. 1, Table 2). DADA2 detected on average 555 (SE = 42) metazoan COI ASVs per sample, and clustering reduced this number to around 250, regardless the *d*-value. For ribosomal loci, clustering at $d = 3\text{-}5$ reduced OTU numbers of around ~30% compared to without clustering, while at $d = 11$, cluster numbers were more than halved.

Table 2. Effect of pipeline and site on the number of metazoan and prokaryote clusters. Results of the analysis of variance (ANOVA) of the rarefied cluster richness for the three genes studied. Pairwise comparisons were performed with Tukey's HSD tests. DS: Dada2+swarm; DSL: Dada2+swarm+LULU; d: swarm *d-value*. LULU curation was performed with *minimum match* at 84% and 90%, and with *minimum ratio* = 100 for 18S and *minimum ratio* = 1 for COI. Significance codes: ***: p<0.001; **: p<0.01; *: p<0.05.

LOCUS	**F-value**	***p-value***	**Significant pairwise comparisons**
COI			
Pipeline	135.2	p < 0.001	Dada2 > DS***
Site	226.7	p < 0.001	DS(d=1) > DS(d=13)**; DS > DSL84%***; D(S)L90% > D(S)L84%***
Pipeline x Site	0.15	p > 0.05	Dada2 > DL***; DL90% > DS(L)***; DL84% > DS(d=5-13)***; DL > DSL***
18S V1-V2			
Pipeline	67.2	p < 0.001	Dada2 > DS***
Site	263.1	p < 0.001	DS(d=1) > DS(d=3-11)***; DS(d=11) < DS(d=1-5)***
Pipeline x Site	0.3	p > 0.05	Dada2 > DL84%*; DL > DS(d=3-11)***; DL > DSL***
16S V4-V5			
Pipeline	188.7	p < 0.001	Dada2 > DS***
Site	18.3	p < 0.001	DS(d=1) > DS(d=3-11)***; DS(d=3) > DS(d=5)***; DS(d=11) < DS(d=1-5)***
Pipeline x Site	0.06	p > 0.05	



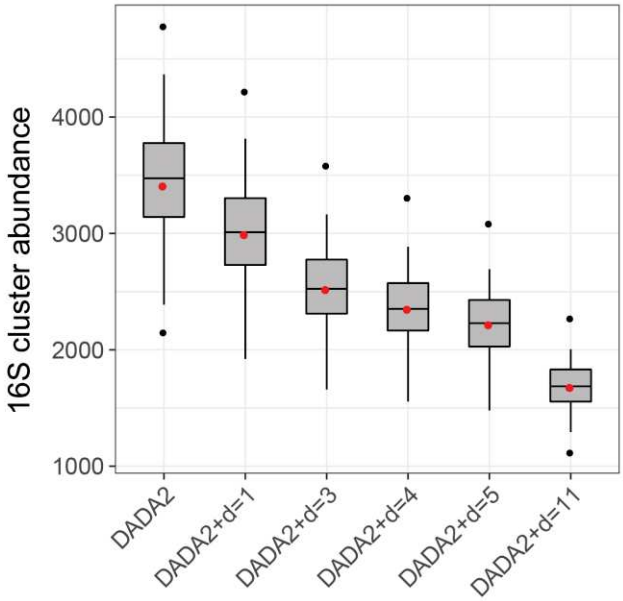


Figure 1. Number of metazoan (COI, 18S) and prokaryote (16S) clusters detected in sediment of 14 deep-sea sites with ASV vs OTU-centred datasets. ASVs were obtained with the DADA2 metabarcoding pipeline, and clustered with swarm at different d values. Metazoan ASVs and OTUs were curated with LULU at 84% and 90% minimum match. LULU curation was performed with minimum ratio = 100 for 18S and minimum ratio = 1 for COI. Cluster abundances were obtained from datasets rarefied to same sequencing depth. Boxplots represent medians with first and third quartiles. Red dots indicate means.

LULU curation of ASVs or OTUs decreased the number of COI and 18S clusters detected (Fig. 1). This decrease was significant for both ASVs and OTUs with COI, but less marked for 18S as LULU's *minimum ratio* was set to 100 (Table 2). For COI, where LULU curation was performed with *minimum ratio* = 1, the *minimum match* parameter had a strong influence on alpha diversity. Indeed, LULU curation of ASVs or OTUS with *minimum match* at 90% resulted in significantly more clusters than at 84% (Table 2). In contrast, the magnitude of the *minimum match* parameter did not significantly affect the number of clusters for 18S, where LULU curation was performed with *minimum ratio* = 100. LULU curation of ASVs resulted in more OTUs than swarm clustering for both loci, with both *minimum match* levels tested (Fig. 1, Table 2). Similarly, LULU curation of ASVs resulted in significantly more clusters than LULU curation of OTUs produced with any *d*-value (Fig. 1, Table 2).

Looking at mean ASV and OTU numbers detected per phylum with each pipeline showed consistent effects of swarm clustering and LULU curation, but highlighted strong differences in the amount of intragenomic variation between taxonomic groups. For all loci investigated, some taxa displayed high ASV to OTU ratios, while others were hardly affected by clustering or LULU curation in terms of numbers of clusters detected (Fig S1).

3.3 Patterns of beta-diversity between pipelines

PERMANOVAs confirmed that sites differed significantly in terms of community structure, accounting from 46% to 89% of variation in data. Evaluating the effect of LULU curation for metazoans showed that LULU-curated data resolved similar community compositions than non-curated data, accounting for < 1% of variation in data (Fig. 2).

Although ASV and OTU datasets detected similar amounts of variation due to sites in PERMANOVAs, clustering levels affected the ecological patterns resolved by ordinations in rRNA loci (Fig 2). Metazoan 18S ASVs showed strong segregation by ocean basin, with samples grouped by depth within each basin, and prokaryote ASVs showed both strong segregation by ocean basin and depth (Fig. 2). Clustering at *d*-values > 1 decreased differences among deep sites (> 1,000 m) across ocean basins, emphasizing the depth effect over the basin effect. This change in ecological pattern occurred consistently with *d*-values from 3 to 11 (Fig. 2, Fig. S2).

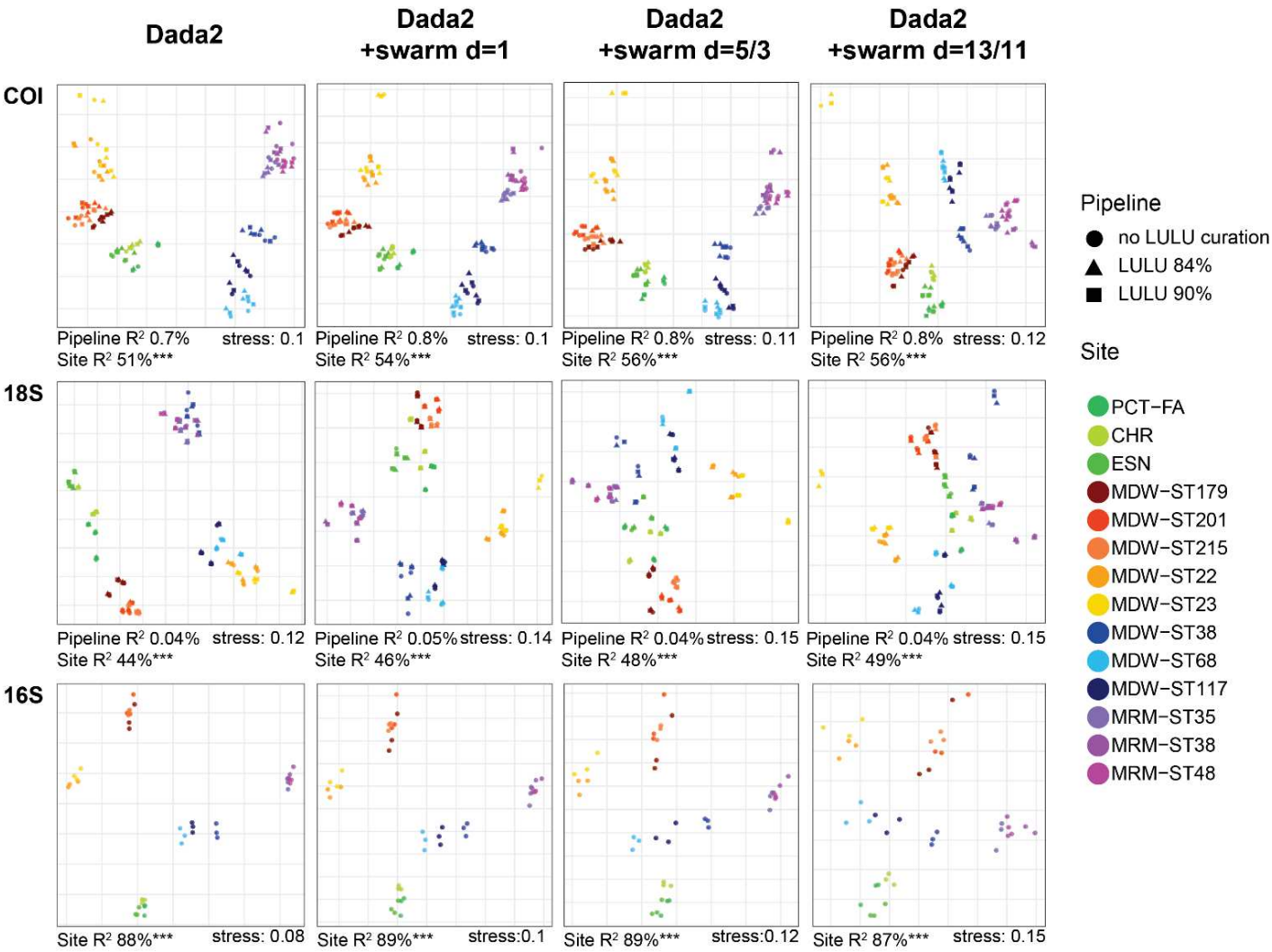


Figure 2. Metazoan (COI, 18S) and prokaryote (16S) beta-diversity patterns in ASV and OTU-centred datasets. Nonmetric multidimensional scaling (NMDS) ordinations showing community differentiation observed between sites with different clustering scenarios. ASVs were obtained with the DADA2 metabarcoding pipeline, and clustered with swarm at d = 1, 5, and 13 (COI) and d = 1, 3, 11 (18S, 16S). Metazoan ASVs and OTUs were curated with LULU at 84% and 90% *minimum match*. LULU curation was performed with *minimum ratio* = 100 for 18S and *minimum ratio* = 1 for COI. R^2 values and associated p-values obtained in PERMANOVAs are shown under the ordination plots. Significance codes: ***: p<0.001; **: p<0.01; *: p<0.05. Site colour codes: Green: Mediterranean > 1,000 m; Red: Mediterranean Gibraltar Strait 300-1,000 m; Yellow: Atlantic Gibraltar Strait 300-1,000 m; Blue: North Atlantic > 1,000 m; Purple: Arctic > 1,000 m.

3.4 Taxonomic assignment quality

Assigning with BLAST resulted in mock community assignments comparable to described above. With COI, eight of the ten species produced one single OTU, with six correctly assigned at genus-level, and two species were taxonomically correctly assigned only to class-level and produced 2-3 OTUs (Fig S3). With 18S, seven species were recovered (4 correctly assigned at genus-level), with two producing more than one OTU, and the three vesicomyid bivalve species were taxonomically unresolved and assigned up to family-level while generating 2 OTUs. Assigning the COI dataset with RDP using the MIDORI-UNIQUE database resulted in assignments of the mock samples that did not match the expected taxa and were mostly belonging to arthropods, a problem not observed with BLAST (data not shown). When the database was reduced to marine-only taxa, RDP results were comparable to BLAST, with seven species correctly assigned at genus-level. Assigning the 18S dataset with RDP produced results comparable to BLAST, although taxonomic assignments were less accurate for two species.

BLAST and RDP assigned similar amounts of OTUs in the prokaryote dataset, but BLAST assigned 20% (18S) and 70% (COI) less OTUs at phylum-level than RDP in the metazoan datasets, even at minimum hit identity of 70% (Table S8). BLAST hit identities of the overall datasets varied strongly depending on phyla and marker gene (Fig. 3). For 18S, 90% of metazoan OTUs had assignment identities ≥ 86%, corresponding roughly to accurate phylum-level (Stefanni et al. 2018; Edgar 2017a). Only 34% had reliable genus-level assignments, i.e. with > 95% similarity (Table S8). For COI, only 30% of metazoan assignments were reliable at phylum-level (≥ 80%), and only 1% at genus-level (> 93%). BLAST hit identity was much higher for prokaryotes, with 98% of ASVs assigned with ≥ 86% similarity to sequences in databases, and 65% had reliable genus-level assignments (> 95% similarity). With RDP, 77% of metazoan 18S OTUs and 96% of prokaryote 16S ASVs had phylum-level bootstraps ≥ 80%, and 59% and 76% also had genus-level bootstraps ≥ 80%, respectively. For COI, applying a minimum phylum-level bootstrap of 80% resulted in an unviable decrease in the number of target OTUs, as only 242 metazoan OTUs (~1%) remained after filtering, and only 112 (0.3%) with acceptable genus-level bootstraps (Table S8). Indeed, most OTUs, primarily assigned to arthropods, cnidarians, molluscs, vertebrates, and poriferans still had phylum-level bootstraps < 60% (Fig. 3).

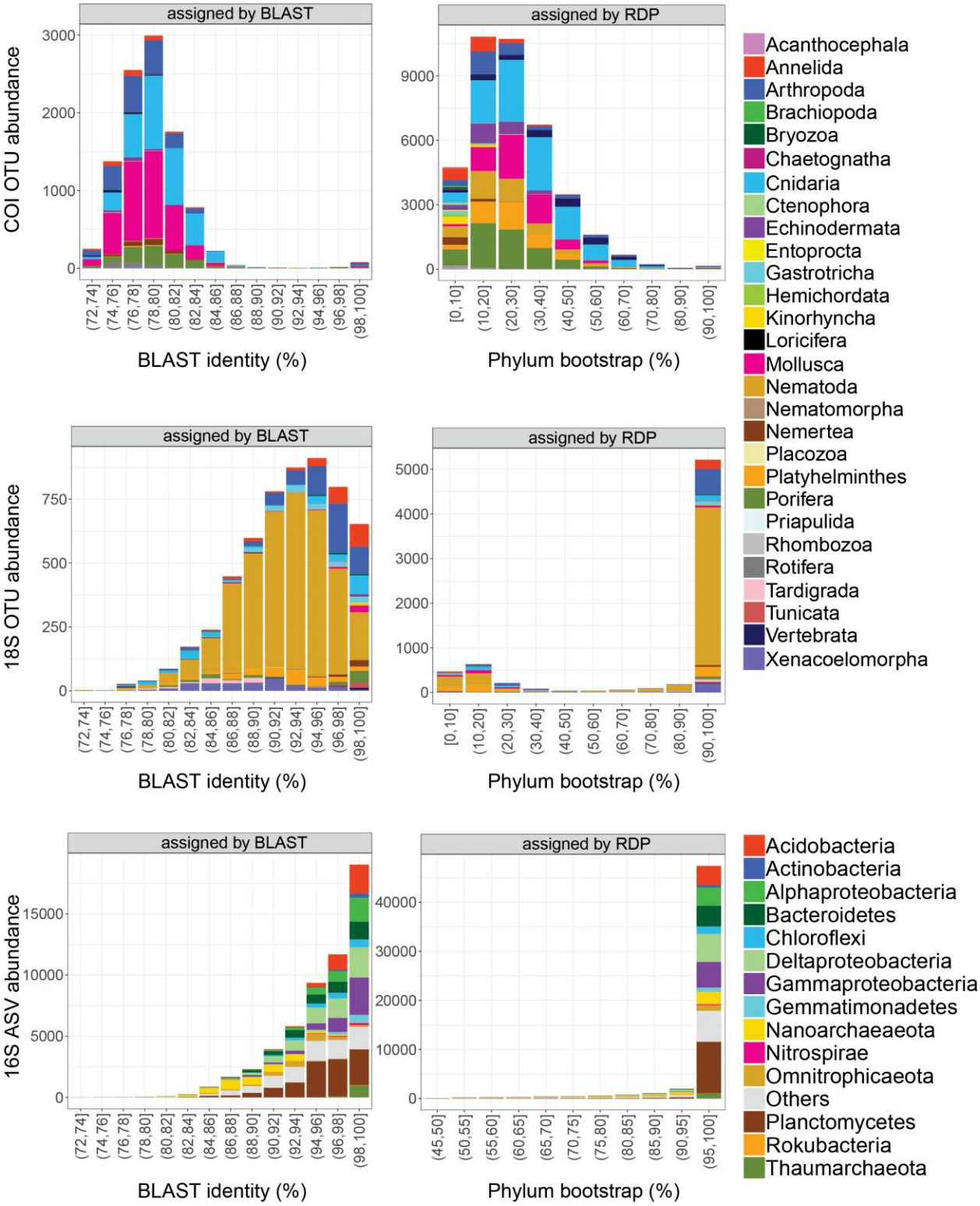


Figure 3. Taxonomic assignment quality of BLAST and RDP methods on metazoan (COI, 18S) and prokaryote (16S) metabarcoding datasets of 14 deep-sea sites. Metazoan data was clustered with swarm at *d=1* and curated with LULU at 90% (*minimum ratio* = 100) for 18S and 84% (*minimum ratio* = 1) for COI. Taxonomic assignments were performed on the Silva132 database for 18S and 16S, and on the MIDORI-UNIQUE database subsampled to marine taxa for COI.

4 Discussion

4.1 ASVs *vs* OTUs: a choice depending on taxon of interest and research question

ASVs have recently been advocated to replace OTUs "*as the standard unit of marker-gene analysis and reporting*" (Callahan et al. 2017): an advice for microbiologists that may not apply when studying metazoans. Life histories of organisms, together with intrinsic properties of marker genes, determine the level of intragenomic and intraspecific diversity. Metazoans are well known to exhibit variable and sometimes very high intraspecific polymorphism. This intraspecific variation is a recognised problem in metabarcoding, known to generate spurious clusters (Brown et al. 2015), especially in the COI barcode marker. Indeed, this gene region has increased intragenomic variation due to its high evolutionary rate (Machida and Knowlton 2012; Machida et al. 2012), but also due to heteroplasmy and the abundance of pseudogenes, such as NUMTs, playing an important part of the supernumerary OTU richness in COI-metabarcoding (Bensasson et al. 2001; Song et al. 2008). Concerted evolution, a common feature of SSU rRNA markers such as 16S (Hashimoto et al. 2003; Klappenbach et al. 2001) and 18S (Carranza et al. 1996), limits the amount of intragenomic polymorphism. In metazoans, a lower level of diversity is thus expected for 18S than for COI. This is reflected in the lower ASV (DADA2) to OTU (DADA2+swarm) ratios observed here for 18S (1.4 – 2.5) compared to COI (2.3 – 3.2), at clustering *d*-values comprised between one and seven (Table S7), underlining the different influence – and importance – of clustering on these loci, and the need for a versatile, marker by marker choice for clustering parameters.

The results on the mock samples showed that even single individuals produced very different numbers of ASVs, suggesting that ASV-centred datasets do not accurately reflect species composition in metazoans. Intragenomic and intraspecific polymorphism are highly variable across taxa (Plouviez et al. 2009; Teixeira et al. 2013), as confirmed by the very variable decrease in cluster numbers observed with clustering in this study for different phyla (Fig. S1). The taxonomic compositions of samples based on ASVs may thus reflect genetic rather than species diversity. This distinction is important to keep in mind, as the *species*, i.e. "*a lineage or group of connected lineages with a distinct role*" (Freudenstein et al. 2017), constitutes the core of biodiversity inventories for biological and ecological studies. The species is a core concept in ecology and evolution that helps organizing agriculture, trade, and industry (e.g. species used for the production of biomaterial), as well as measuring the impact of human activity on Earth's ecosystems (e.g. biomarker taxa and pathogenic or invasive species). While

biotic diversity can be valued and assessed at various levels, including that of the individual organism and the genetic locus, many theoretical and applied developments in ecology are deeply rooted in the species concept, and species richness, while not perfect, remains an essential metric (Freudenstein et al. 2017).

Clustering ASVs into OTUs alleviated the numerical inflation in the mock samples, but some species still produced more than one OTU, even at *d*-values as high as *d = 11-13*. While clustering improved numerical results in the mock communities, it led to poorer taxonomic assignments, for e.g., the vesicomyid bivalves only being identified up to class-level in clustered datasets with both loci. With 18S, clustering at *d*-values > 1 even led to the loss of the shrimp species *Chorocaris* sp., which was merged to the closely related *A. muricola* (Table 1). Similarly, a *d*-value at 11 led to significantly lower OTU numbers than any other tested *d*-value for both ribosomal loci (Table 2), explaining the much higher ASV to OTU ratios observed (4.1 – 4.4, Table S7). When studying natural habitats, very likely to harbour closely related co-occurring species, clustering at *d*-values higher than 1 is thus likely to lead to the loss of true species diversity, particularly in taxa known to be poorly resolved (e.g. cnidarians with COI, Hebert et al. 2003), and in general with markers having lower taxonomic resolution such as 18S.

The reproductive mode and pace of selection in microbial populations may lead to locally lower levels of intraspecific variation than those expected for metazoans. Prokaryotic alpha diversity was however also affected by the clustering of ASVs (Fig. 1), supporting the estimation of a 2.5-fold greater number of 16S rRNA variants than the actual number of bacterial “species” (Acinas et al. 2004). The significant decrease in the number of OTUs after clustering at *d = 1* (Table 2, Fig. 1, decrease of ~30%) suggests the occurrence of very closely related 16S rRNA sequences, possibly belonging to the same ecotype/species. Such entities may still be important to define in studies aiming for example at identifying species associations (i.e. symbiotic relationships) across large distances and ecosystems, where drift or selection can lead to slightly different ASVs in space and time, with their function and association remaining stable.

Finally, apart from alpha diversity estimates, clustering also affected the resolution of ecological patterns in ribosomal loci when *d*-values were higher than 1 (Fig. 2). This can be explained by the fact that clustering gives more weight to large distinct OTUs compared to many small (i.e. with low read numbers) ASVs. The deep Atlantic and Mediterranean sites, segregating at the ASV-level (possibly due to vicariance by distance), thus appeared more

similar at high *d*-values, revealing the occurrence of distinct ASVs belonging to many shared OTUs and thus suggesting an ecological signal in fine-scale sequence variants. This is in accordance with other studies reporting differences in beta diversity patterns in ASV *vs* OTU datasets for ribosomal loci, when large divergence thresholds were used for clustering (Xiong and Zhan 2018; Bokulich et al. 2013). This also reveals the interdependence of alpha and beta diversity components, so that clustering ASVs into OTUs and thereby reducing alpha diversity, leaves more space for beta diversity to be expressed, as observed in both population genetics (Jost 2008; Beaumont and Nichols 1996) and community analysis (Jost 2007). Overall, these results confirm the advantage of combining error-correction tools with clustering and post-clustering curation tools, as this allows access to both interspecies and intraspecies information (Turon et al. 2020).

4.2 Importance of parameter adjustment for LULU curation

LULU curation proved effective in limiting the number of multiple clusters produced by single individuals in the mock samples, confirming its efficiency to correct for intragenomic diversity (Table 1). Moreover, the fact that the number of unexpected clusters ("Others", Table 1) was not affected by LULU curation also shows that LULU specifically removes spurious OTUs and not true species diversity. However, careful adjustment of LULU parameters was needed, particularly for the *minimum ratio*, as at default level (1) it led to the loss of up to two mock species with 18S. This need for relaxed *minimum ratio* values can be explained by the non-ideal design of the mock samples. Indeed, LULU should be applied on datasets containing as many samples as possible, which should have compositional similarities (i.e. overlapping species lists). If this is not the case, LULU will work as a pure clustering algorithm, at defined *minimum match* levels. Here, all species were co-occurring in the mock samples at consistent abundance ratios and some mock species were not occurring (or rarely) in environmental samples. For those, random amplification biases leading to consistently low read numbers in both mock samples resulted in LULU merging them to closely related mock species. Increasing the *minimum ratio*, i.e. the expected minimal abundance ratio between a true OTU and an associated spurious sequence, allowed detecting all mock species with 18S. With *minimum ratio* at 100, one mock species (the gastropod *Paralepetopsis* sp) was still lost when *minimum match* was at 84%, which could indicate that *minimum match* at 90% is more appropriate for 18S. However, as all mock species were retained at both *minimum match* levels with *minimum*

ratio at 1000, the loss of that species at 84% may also just reflect the non-ideal mock design (*Paralepetopsis* sp. being very poorly amplified by 18S, it got merged to a bivalve OTU as their similarity was greater than 84%). Given the fact that 18S is evolving much slower than COI, this marker is taxonomically much less resolutive and phylum-level similarity is at ~86% (Stefanni et al. 2018). As error OTUs are produced within each individual, it is reasonable to think that their similarity to their parent OTUs will be greater than phylum-level similarity, justifying the use of 90% *minimum match*. This increased *minimum match* also has the added benefit to decrease calculation time on large datasets. For COI, although results in the mock samples showed the best performance at *minimum ratio* of 1 and little effect of the *minimum match* parameter (90% *vs* 84%), both *minimum match* levels resulted in significantly different OTU numbers in the environmental samples (Table 2, Fig. 1). This was not the case for 18S, where both 84% and 90% *minimum match* resulted in similar numbers of OTUs in the environmental samples (*minimum ratio* at 100). Thus, increasing the *minimum ratio* parameter is essential for not losing species in sample-poor datasets, and will be more correct than adjusting the *minimum match*.

The mock communities used in this study, apart from being taxonomically limited to just 10 species, did unfortunately not contain several haplotypes of the same species (intraspecific variation). This could explain the comparable results obtained with LULU curation of ASVs and LULU curation of OTUs in the mock samples, and lead to the hasty conclusion of a limited effect of clustering. Communities detected in environmental samples are much more complex, likely comprising many different haplotypes of the same species. However, LULU curation of ASVs cannot substitute clustering algorithms to account for natural haplotype diversity. Indeed, not all haplotypes co-occur and when they do so, they may vary in proportion and dominance relationships, making clustering the best tool to account for natural haplotypic diversity. This is in line with LULU developers (Frøslev et al. 2017), who recommend clustering ASVs for addressing the average intraspecific variation of the target group, and subsequent curation with LULU. In the environmental samples, LULU curation of the ASV datasets led to significantly more OTUs than LULU curation of swarm-clustered OTUs with both metazoan loci (Table 2). This indicates that LULU curation merges less ASVs than the amount grouped through clustering, and highlights the different purposes of both tools, LULU effectively removing spurious OTUs, while clustering allows removing haplotype diversity.

4.3 Taxonomic resolution and assignment quality

The COI locus allowed the detection of all ten species present in the mock samples, compared to seven in the 18S dataset (Table 1). This locus also provided much more accurate assignments, most of them correct at the genus (and species) level, confirming that COI uncovers more metazoan species and offers a better taxonomic resolution than 18S (Tang et al. 2012; Clarke et al. 2017; Andújar et al. 2018b). Our results also support approaches combining nuclear and mitochondrial markers to achieve more comprehensive biodiversity inventories (Cowart et al. 2015; Drummond et al. 2015; Zhan et al. 2014). Indeed, strong differences exist in amplification success among taxa (Bhadury et al. 2006; Carugati et al. 2015), exemplified by nematodes, which are well detected with 18S but not with COI (Bucklin et al. 2011). The 18S barcode marker performed better in the detection of nematodes, annelids, platyhelminths, and xenacoelomorphs while COI mostly detected cnidarians, molluscs, and poriferans (Fig. 3, Fig. S1), highlighting the complementarity of these two loci. This high complementarity of COI and 18S in terms of targeted taxa also supports the approach taken by Stefanni et al. (2018), indeed subsampling each gene dataset for its "best targeted phyla" and subsequently combining both seems to be a very efficient way to produce comprehensive and non-redundant biodiversity inventories.

Finally, compared to BLAST assignments, similar taxonomic assignments were observed using the RDP Bayesian Classifier on the mock samples for 18S and for COI when using the MIDORI-UNIQUE marine-only database (Fig. S3). Poor performance of RDP using the full MIDORI database is likely due to the size of the database, and to its low coverage of deep-sea species. Indeed, small databases, taxonomically similar to the targeted communities, and with sequences of the same length as the DNA fragment of interest, are known to maximise accurate identification (Macheriotou et al. 2019; Ritari et al. 2015). The problem of underrepresentation of deep-sea taxa is especially apparent with the BLAST assignments, which generally displayed low identities to sequences in databases, especially for COI (Fig. 3). Minimum similarities of 80% for COI and 86% for 18S as cut-off values for metazoans have been used to improve the taxonomic quality of metazoan metabarcoding datasets (Stefanni et al. 2018). However, phylogenies of marine invertebrates are characterised by high levels of species divergence (up to ~30%), even within genera (Zanol et al. 2010). Studies on deep-sea taxa have found that some invertebrate species had COI sequences diverging more than 20% from any other species present in molecular databases (Shank et al. 1999; Herrera et al. 2015). At present, it thus seems difficult to work at taxonomic levels beyond phylum-level with deep-sea metabarcoding data

when using large public databases. When using the reduced marine-only COI database, RDP provided the most accurate assignments in the mock samples (Fig. S3). However, filtering to accurate phylum-level bootstraps (≥ 80) drastically reduced the number of OTUs in the overall dataset (1% of OTUs left, Table S8). The development of custom-built marine RDP training sets, without overrepresentation of terrestrial species, is therefore needed for this Bayesian assignment method to be effective on deep-sea datasets. With reduced and more specific databases, removing clusters with phylum-level bootstraps < 80 should be an efficient way to increase taxonomic quality of deep-sea metabarcoding datasets. At present, if accurate taxonomic assignments are sought while using universal primers, we advocate assigning taxonomy in two steps: first, using BLAST and a large database including all phyla amplifiable by the primer set as BLAST performs better than RDP in terms of speed. The clusters belonging to the groups of interest can then be extracted and re-assigned using RDP and a smaller, taxon-specific database.

5 Conclusions and perspectives

Using mock communities and environmental samples, we evaluate several recent algorithms and assess their capacity to improve the quality of molecular biodiversity inventories of metazoans and prokaryotes. Our results support the fact that ASV data should be produced and communicated for reusability and reproducibility following the recommendations of Callahan et al. (2017). This is especially useful in large projects spanning wide geographic zones and time scales, as different ASV datasets can easily be merged *a posteriori,* and clustered if necessary afterwards. However, our results confirm that both ASVs and OTUs describe relevant, yet different levels of biotic diversity. ASVs comprehensively describe genetic diversity (incl. intraspecies) while OTUs more accurately reflect interspecies diversity. Considering 16S polymorphism observed in prokaryotic species (Acinas et al. 2004) and the possible geographic segregation of their populations, using OTUs may also be suitable in prokaryotic datasets, for example in studies screening for species associations, as symbionts may be prone to differential fixation through enhanced drift (Shapiro et al. 2016).

This study emphasized that swarm clustering needs to be adapted to each genetic marker and taxonomic compartment, in order to identify an optimal balance between the correction for spurious clusters and the loss of species. Specifically, *d*-values > 1 appeared to be less appropriate with 18S for metazoans. Our results also demonstrated that LULU effectively curates metazoan biodiversity inventories obtained through metabarcoding. They underline the

need to adapt parameters for LULU curation, in particular the *minimum ratio* level in the case of sample-poor datasets, where co-occurrence and abundance patterns may be distorted.

Finally, this study also showed that accurate taxonomic assignments of deep-sea species can be obtained with the RDP Bayesian Classifier, but only with reduced databases containing ecosystem-specific sequences.

Data accessibility

The data for this work can be accessed in the European Nucleotide Archive (ENA) database (project: PRJEB33873), please refer to the metadat excel sheet for ENA file names. The data set, including raw sequences, databases, and ASV/OTU tables, has been deposited on https://doi.org/10.12770/0b5d250b-8418-4dda-b39c-960c4481df93. Bioinformatic scripts, config files, and R scripts are available on Gitlab (https://gitlab.ifremer.fr/abyss-project/).

Acknowledgements

This work is part of the "*Pourquoi Pas les Abysses?*" project, funded by Ifremer and the eDNAbyss (AP2016 -228) project, funded by France Génomique (ANR-10-INBS-09) and Genoscope-CEA. We wish to thank Caroline Dussart for her contribution to the early developments of bioinformatic scripts,Patrick Durand for bioinformatic support, and Cathy Liautard-Haag for her help with lab management. We also wish to express our gratitude to the participants and mission chiefs of the Essnaut (Marie-Anne Cambon Bonavita), MarMine (Eva Ramirez Llodra, project 247626/O30 (NRC-BI), PEACETIME (Cécile Guieu and Christian Tamburini), CanHROV (Marie Claire Fabri) and MEDWAVES (Covadonga Orejas) cruises. The MEDWAVES cruise was organised in the framework of the ATLAS Project, supported by the European Union 2020 Research and Innovation Programme under grant agreement number: 678760. We thank Stefaniya Kamenova, Tiago Pereira, and four anonymous reviewers for their comments on previous versions of this manuscript. An earlier version of this manuscript has been peer-reviewed and recommended by Peer Community In Ecology (https://dx.doi.org/10.24072/pci.ecology.100043).

Conflict of interest disclosure

The authors declare that they have no financial conflict of interest with the content of this article. Sophie Arnaud-Haond is one of the PCI Ecology recommenders.

Author contributions

MIB and SAH designed the study, MIB and JP carried out the laboratory and molecular work; MIB and BT performed the bioinformatic and statistical analyses. LQ led the bioinformatic development and participated in the study design. MIB, BG, and SAH wrote the manuscript. All authors contributed to the final manuscript.

Chapter III.
An assessment of environmental metabarcoding protocols aiming at favouring contemporary biodiversity in inventories of deep-sea communities

Published as:

Brandt MI, Trouche B, Henry N, Liautard-Haag C, Maignien L, de Vargas C, Wincker P, Poulain J, Zeppilli D, and Arnaud-Haond S. An assessment of environmental metabarcoding protocols aiming at favouring contemporary biodiversity in inventories of deep-sea communities. 2020. *Frontiers in Marine Science*. DOI: 10.1101/836080

The ability to capture all taxa representative of a given community with a minimal set of barcoding primers is a particular challenge in marine sediments due to the high abundance of taxa that are difficult to amplify by PCR (especially the meiofaunal nematodes). To attempt minimizing the influence of primer bias, a nuclear and a mitochondrial primer set were selected to capture the broadest range of metazoan taxa possible (mitochondrial COI and the 18S V1-V2 rRNA marker genes), and combined with two additional nuclear primers targeting unicellular eukaryotes (18S V4 rRNA marker gene) and prokaryotes (16S V4-V5 rRNA marker gene).

However, extracellular DNA is abundant in marine sediments and has been estimated to account for up to 50% of the total DNA pool. This DNA (benthic and potentially non-benthic) may be *archived* in deep-sea sediments due to lower degradation rates and thus significantly bias biodiversity inventories towards describing past, rather than present communities (Torti et al. 2015; Sinniger et al. 2016; Corinaldesi et al. 2018). To evaluate the bias produced by this ancient DNA (aDNA), biodiversity inventories produced by five molecular methods were compared at five sites from various habitats (seamount, mud volcano, hydrothermal vent). Using commercially available extraction kits, we investigated the impacts of two methods aiming at removing small extracellular DNA fragments on biodiversity estimates of pro- and eukaryotes from deep-sea sediments, and compared inventories obtained from co-extracted DNA and RNA in order to evaluate whether RNA may be more suitable for describing the *live* community compartment.

Question addressed:

Does extracellular DNA archived in deep-sea sediments lead to significant bias in molecular biodiversity inventories of prokaryotes, unicellular eukaryotes, and metazoans?

Résumé en français

Les fonds abyssaux couvrent plus de 50% de la planète Terre et représentent un grand réservoir de biodiversité encore largement méconnu . Malgré cette méconnaissance, ils sont de plus en plus sous la menace d'activités anthropiques. Dans ces écosystèmes vastes et difficiles d'accès, le métabarcoding par ADN environnemental (ADNe) est un outil utile et efficace pour étudier la biodiversité et mettre en œuvre des programmes d'évaluation d'impact. Pourtant, son application sur des sédiments profonds est potentiellement biaisée par la présence d'ADN archivé provenant d'organismes morts. Or, l'inclusion de cet ADN ancien (ADNa) aboutirait à des inventaires de biodiversité passée plutôt que présente.

À l'aide de kits d'extractions d'ADN disponibles dans le commerce, nous avons étudié les impacts de cinq méthodes de traitement moléculaire sur les inventaires de la biodiversité produits par métabarcoding, ciblant les procaryotes (16S-V4V5), les eucaryotes unicellulaires (18S-V4) et les métazoaires (18S-V1, COI). Dans un premier temps, des inventaires basés sur l'ADN furent comparés à ceux révélés par l'ARN. En effet, ce-dernier, étant produit uniquement par des organismes vivants, a été présenté comme une approche plus pratique pour cibler la partie active des communautés. Parallèlement, l'ADN ancien étant principalement constitué de petits fragments, nous avons aussi évalué l'effet de l'élimination de fragments d'ADN courts par sélection de taille et reconcentration par éthanol.

Les résultats montrent que l'élimination de fragments d'ADN courts n'affecte pas les estimations de la diversité alpha et bêta dans aucun des compartiments biologiques étudiés. Les résultats confirment également les doutes quant à la possibilité de mieux décrire les communautés vivantes en utilisant l'ARN environnemental (ARNe). Sur les marqueurs ribosomaux, l'ARN, tout en résolvant des schémas spatiaux similaires à l'ADN co-extrait, a entraîné des estimations de richesse significativement plus élevées, soutenant les hypothèses de persistance accrue de l'ARN ribosomal (ARNr) dans l'environnement, et l'existence d'un biais additionnel et non mesuré en raison de la surabondance d'ARNr dans l'environnement et d'ARN sécrété à taux variables en fonction de l'activité métabolique des organismes. Sur le locus mitochondrial, l'ARN a détecté une richesse métazoaire inférieure tout en résolvant moins de différences écologiques que l'ADN co-extrait, reflétant la grande labilité de l'ARN messager. Les résultats soulignent également l'importance d'utiliser de grandes quantités de sédiments (≥ 10 g) pour étudier avec précision la diversité eucaryote. Nous concluons donc que l'ADN est plus pertinent que l'ARN pour des études logistiquement réalistes, reproductibles, et fiables. Nous confirmons aussi que des quantités de sédiments plus grandes (≥ 10 g) fournissent des

évaluations plus complètes et précises de la biodiversité eucaryote benthique et qu'il faut favoriser l'augmentation du nombre de réplicas biologiques plutôt que techniques pour déduire des patrons écologiques fiables.

Abstract

The abyssal seafloor covers more than 50% of planet Earth and is a large reservoir of still mostly undescribed biodiversity. It is increasingly targeted by resource-extraction industries and yet is drastically understudied. In such remote and hard-to-access ecosystems, environmental DNA (eDNA) metabarcoding is a useful and efficient tool for studying biodiversity and implementing environmental impact assessments. Yet, eDNA analysis outcomes may be biased towards describing past rather than present communities as sediments contain both contemporary and ancient DNA.

Using commercially available kits, we investigated the impacts of five molecular processing methods on eDNA metabarcoding biodiversity inventories targeting prokaryotes (16S), unicellular eukaryotes (18S-V4), and metazoans (18S-V1, COI). As the size distribution of ancient DNA is skewed towards small fragments, we evaluated the effect of removing short DNA fragments via size-selection and ethanol reconcentration using eDNA extracted from ~10 g of sediment at five deep-sea sites. We also compare communities revealed by eDNA and environmental RNA (eRNA) co-extracted from ~2 g of sediment at the same sites.

Results show that removing short DNA fragments does not affect alpha and beta diversity estimates in any of the biological compartments investigated. Results also confirm doubts regarding the possibility to better describe live communities using eRNA. With ribosomal loci, eRNA, while resolving similar spatial patterns than co-extracted eDNA, resulted in significantly higher richness estimates, supporting hypotheses of increased persistence of ribosomal RNA (rRNA) in the environment and unmeasured bias due to over-abundance of rRNA and RNA release. With the mitochondrial locus, eRNA detected lower metazoan richness and resolved fewer spatial patterns than co-extracted eDNA, reflecting high messenger RNA lability. Results also highlight the importance of using large amounts of sediment (≥10 g) for accurately surveying eukaryotic diversity.

We conclude that eDNA should be favoured over eRNA for logistically realistic, repeatable, and reliable surveys, and confirm that large sediment samples (≥10 g) deliver more complete and accurate assessments of benthic eukaryotic biodiversity and that increasing the number of biological rather than technical replicates is important to infer robust ecological patterns.

1 Introduction

Environmental DNA (eDNA) metabarcoding is an increasingly used tool for biodiversity inventories and ecological surveys. Using high-throughput sequencing (HTS) and bioinformatic processing, it allows the detection or the inventory of target organisms using their DNA directly extracted from soil, water, or air samples (Taberlet et al. 2012a). As it does not require specimen isolation, it represents a practical and efficient tool in large and hard-to-access ecosystems, such as the marine realm. Besides allowing studying various biological compartments simultaneously, metabarcoding is also very effective for detecting diversity of small organisms (micro-organisms, meiofauna) largely disregarded in visual biodiversity inventories due to the difficulty of their identification based on morphological features (Carugati et al. 2015) .

The deep sea, covering more than 50% of Planet Earth, remains critically understudied, despite being increasingly impacted by anthropogenic activities and targeted by resource-extraction industries (Ramirez-Llodra et al. 2011). The abyssal seafloor is mostly composed of sedimentary habitats containing high numbers of small (< 1 mm) organisms, and characterized by high local and regional diversity (Grassle and Maciolek 1992; Smith and Snelgrove 2002). Given the increased time-efficiency offered by eDNA metabarcoding and its wide taxonomic applicability, this tool is a good candidate for large-scale biodiversity surveys and Environmental Impact Assessments (EIAs) in the deep-sea biome.

eDNA is a complex mixture of genomic DNA present in living cells, extra-organismal DNA, and extracellular DNA originating from the degradation of organic material and biological secretions (Torti et al. 2015). Extracellular DNA has been shown to be very abundant in marine sediments, representing 50-90% of the total DNA pool (Corinaldesi, Tangherlini, Manea, & Dell'Anno, 2018; Dell'Anno & Danovaro, 2005). However, this extracellular DNA compartment may not only contain DNA from contemporary communities. Indeed, nucleic acids can persist in marine sediments as their degradation rate decreases due to adsorption onto the sediment matrix (Corinaldesi, Beolchini, & Dell'Anno, 2008; Torti et al., 2015). Low temperatures, high salt concentrations, and the absence of UV light are additional factors enhancing long-term archiving of DNA in deep-sea sediments (Torti et al. 2015; Nagler et al. 2018). Decreased rates of abiotic DNA decay can thus allow DNA persistence over millennial timescales. Indeed, up to 125,000-year-old ancient DNA (aDNA) has been reported in oxic and anoxic marine sediments at various depths (Boere et al. 2011; Lejzerowicz et al. 2013a; Coolen et al. 2013). As extracellular DNA fragment size depends on its state of degradation (Nagler et

al., 2018 report overall size ranges from 80 to over 20,000 bp), aDNA fragments have generally been reported to be <1,000 bp long (Lennon et al. 2018; Boere et al. 2011; Lejzerowicz et al. 2013a; Coolen et al. 2013). Restricting molecular biodiversity assessments to large DNA fragments may thus allow avoiding the bias of aDNA in biodiversity assessments aiming at describing contemporary communities using eDNA metabarcoding.

Environmental RNA (eRNA) has been viewed as a way to avoid the problem of aDNA in eDNA biodiversity inventories because RNA is only produced by living organisms and quickly degrades when released in the environment, due to spontaneous hydrolysis and the abundance of RNases (Torti et al. 2015). Few studies have investigated this in the deep-sea, with contrasting results. Investigating foraminiferal assemblages, Lejzerowicz, Voltsky, & Pawlowski (2013) found similar taxonomic compositions with DNA and RNA, although highlighting that RNA is more appropriate for targeting the active community component. Contrastingly, Guardiola et al. (2016) detected marked differences between RNA and DNA inventories for most eukaryotic groups, but found that both biomolecules detected similar patterns of ecological differentiation, concluding that “dead” DNA did not blur patterns of community structure. Laroche and coworkers (2018, 2017) found stronger responses to environmental impact in alpha diversity measured with eRNA, while eDNA was better at detecting effects on community composition. Finally, long-term archived and even fossil RNA were also reported in sediment and soil (Orsi et al. 2013; Cristescu 2019), casting doubts as to its advantage over DNA to inventory contemporary biodiversity.

The design of a sound environmental metabarcoding protocol to inventory biodiversity on the deep seafloor relies on a better understanding of the potential influence of aDNA on the different taxonomic compartments targeted. Using commercially available kits based on 2 g and 10 g of sediment, we studied samples from five deep-sea sites encompassing three different habitats and spanning wide geographic ranges, in order to select an optimal protocol to survey contemporary benthic deep-sea communities spanning the tree of life. We analyse eDNA and eRNA extracts via metabarcoding, targeting the V4-V5 regions of the 16S ribosomal RNA (rRNA) barcode (Parada et al. 2016) for prokaryotes, the 18S-V4 rRNA barcode region for micro-eukaryotes (Stoeck et al. 2010), and the 18S-V1V2 rRNA (thereafter 18S-V1) and Cytochrome c Oxidase I (COI) barcode markers for metazoans (Sinniger et al. 2016; Leray et al. 2013).

Our objectives were threefold:

1) Evaluate the effect of removing short DNA fragments from DNA extracts obtained using a 10 g extraction kit;

2) Compare eDNA and eRNA inventories resulting from the same samples via a 2 g joint extraction kit,

3) Assess the aforementioned kits in terms of repeatability and suitability for different taxonomic compartments.

2 Materials and methods

2.1 Collection of samples

Sediment cores were collected from five deep-sea sites from various habitats (mud volcano, seamounts, and an area close to hydrothermal vents, Table S1). Triplicate tube cores were collected with a multicorer or with a remotely operated vehicle at each sampling site. The sediment cores were sliced into layers, which were transferred into zip-lock bags, homogenised, and frozen at −80°C on board before being shipped on dry ice to the laboratory. The first layer (0-1 cm) was used for the present analysis. In each sampling series, an empty bag was kept as a field control processed through DNA extraction and sequencing.

2.2 Nucleic acid extractions and molecular treatments

eDNA with the 10 g-PowerMax kit

DNA extractions were performed using ~10 g of sediment with the PowerMax Soil DNA Isolation Kit (MO BIO Laboratories, Inc.; Qiagen, Hilden, Germany). To increase the DNA yield, the elution buffer was left on the spin filter membrane for 10 min at room temperature before centrifugation. For field controls, the first solution of the kit was poured into the control zip lock, before following the usual extraction steps. DNA extracts were stored at -80°C.

Size-selection of eDNA extracts

Size-selection of total eDNA extracted as detailed above from ~10 g of sediment was carried out to remove small DNA fragments. NucleoMag NGS Clean-up and Size Select beads (Macherey-Nagel, Düren, Germany) were used at a ratio of 0.5X for removing DNA fragments < 1,000 bp from 500 µL of extracted eDNA. The target fragments were eluted from the beads with 100 µL elution buffer, and successful size-selection verified by electrophoresis on an Agilent TapeStation using the Genomic DNA High ScreenTape kit (Agilent Technologies, Santa Clara, CA, USA).

Ethanol reconcentration of eDNA extracts

A 3.5 mL aliquot of eDNA extracted from ~10 g of sediment was reconcentrated with 7 mL of 96% ethanol (EtOH) and 200 µl of 5 M sodium chloride (NaCl), according to the guidelines in the *Hints and Troubleshooting Guide* of the PowerMax Soil DNA Isolation Kit. As this protocol does not include any incubation time, it favours large DNA fragments. The

DNA pellet was washed with 1 mL 70% EtOH, centrifuged again for 15 min at 2500 x g, and air-dried before being resuspended in 450 µL elution buffer.

Joint environmental DNA/RNA with the 2 g- RNeasy PowerSoil kit

Joint RNA/DNA extractions were performed with the RNA PowerSoil Total RNA Isolation Kit combined with the RNeasy PowerSoil DNA elution kit (MO BIO Laboratories, Inc.; Qiagen, Hilden, Germany). Between 3 and 5 g of wet and frozen sediment were used, following the manufacturer's suggestions for marine sediments (Table S2). Extraction controls were performed alongside sample extractions. The RNA pellet was resuspended in 60 µL of RNase/DNase-free water. Extracted RNA was then transcribed to first-strand complementary DNA (cDNA) using the iScript Select cDNA synthesis kit (Bio-Rad laboratories, CA, USA) with its proprietary random primer mix. Quality control 16S-V4V5, 18S-V1, and COI PCRs were performed on the RNA extracts to test for potential DNA contamination.

2.3 PCR amplification and sequencing

Nucleic acid extracts were normalised to 0.25 ng/µL and 10 µL of standardized samples were used in PCR. Four primer pairs were used to amplify one mitochondrial and three rRNA barcode loci targeting metazoans (COI, 18S-V1), micro-eukaryotes (18S-V4) and prokaryotes (16S-V4V5, Table S3). Two metazoan mock communities (detailed in Brandt et al. 2020) were included for 18S-V1 and COI. For each sample and marker, triplicate amplicon libraries (see Supporting Information for amplification details) were prepared by ligation of Illumina adapters on 100 ng of amplicons following the Kapa Hifi HotStart NGS library Amplification kit (Kapa Biosystems, Wilmington, MA, USA). After quantification and quality control, library concentrations were normalized to 10 nM, and 8–9 pM of each library containing a 20% PhiX spike-in were sequenced on a HiSeq2500 (System User Guide Part # 15035786) instruments in a 250 bp paired-end mode.

2.4 Bioinformatic analyses

All bioinformatic analyses were performed using a Unix shell script (Brandt, M. I. et al. 2020), available on Gitlab (https://gitlab.ifremer.fr/abyss-project/), on a home-based cluster (DATARMOR, Ifremer). The details of the pipeline, along with specific parameters used for all metabarcoding markers, are given in Table S4 and in Brandt et al. (2020). Pairs of Illumina

reads were corrected with DADA2 v.1.10 (Callahan et al. 2016) following the online tutorial for paired-end data (https://benjjneb.github.io/dada2/tutorial.html) and delivered inventories of Amplicon Sequence Variants (ASVs). Metazoan data was further clustered into OTUs with swarm v2, a single-linkage clustering algorithm (Mahé et al. 2015) that aggregates sequences iteratively and locally around seed sequences based on *d*, the number of nucleotide differences, to determine coherent groups of sequences, independent of amplicon input order, allowing highly scalable and fine-scale clustering. ASVs were swarm clustered at *d* values of 4 for 18S-V1 and 6 for COI, using the FROGS pipeline (Escudié et al. 2018).

We chose to evaluate micro-eukaryote and prokaryote diversity at the ASV level due to its increasing use in the literature (Callahan et al. 2017). Although the use of OTUs may also be justified for microbial diversity depending on study objectives (Brandt, M. I. et al. 2020), we did not expect an alteration of alpha and beta diversity patterns between ASV and OTU levels for the different molecular treatments investigated. ASVs and OTUs were taxonomically assigned via BLAST+ (v2.6.0) based on minimum similarity and minimum coverage (-perc_identity 70 and –qcov_hsp 80). For ASVs, sequences obtained with DADA2 were subsequently assigned with *blastn*. For OTUs, BLAST assignment in FROGS was performed using the *affiliation_OTU.py* command. It is not uncommon for deep-sea taxa to have closest relatives in databases (even congenerics) exhibiting nucleotide divergence exceeding 20% (Shank et al. 1999; Herrera et al. 2015). Considering our interest in diverse and poorly characterized communities, more stringent BLAST thresholds were thus not implemented at this stage. However, additional filters were performed during downstream bioinformatic processing described below, and taxonomic information was used at phylum-level, only when the assignment was deemed reliable at this taxonomic level. The Silva132 reference database was used for taxonomic assignment of rRNA marker genes (Quast et al. 2012), and MIDORI-UNIQUE (Machida et al. 2017) was used for COI.

Molecular inventories were refined in R v.3.5.1 (R Core Team 2018). A blank correction was made using the *decontam* package v.1.2.1 (Davis et al. 2018), removing all clusters that were more prevalent in negative control samples than in true or mock samples. Unassigned and non-target clusters were removed. Additionally, for metazoan loci, all clusters with a terrestrial assignment (groups known to be terrestrial-only) were removed. Samples with fewer than 10,000 target reads were discarded. We performed an abundance renormalization to remove spurious ASVs/OTUs due to random tag switching (Wangensteen and Turon 2016). The COI OTU table was further curated with LULU v.0.1 (Frøslev et al. 2017) to limit the bias due to

pseudogenes, using a minimum co-occurrence of 0.93 and a minimum similarity threshold of 84%.

2.5 Statistical analyses

Sequence tables were analysed using R with the packages phyloseq v1.22.3 (McMurdie and Holmes 2013), following guidelines in online tutorials (http://joey711.github.io/phyloseq/tutorials-index.html), and vegan v2.5.2 (Oksanen et al. 2018). Alpha diversity between molecular processing methods was estimated with the number of observed target clusters in rarefied datasets. Cluster abundances were compared via analyses of deviances (ANODEV) on generalized linear mixed models using negative binomial distributions, as the data were overdispersed. Pairwise post-hoc comparisons were performed via Tukey HSD tests using the *emmeans* package.

Homogeneity of multivariate dispersions were evaluated with the *betapart* package v.1.5.1 (Baselga and Orme 2012), and statistical tests performed on balanced datasets for COI as dispersions were different between 2 g and 10g datasets (Table S5). Data were rarefied for metazoans and Hellinger-normalised for microbial data.

Differences in community compositions resulting from molecular processing were evaluated with Mantel tests (Jaccard and Bray-Curtis dissimilarities for metazoan and microbial taxa respectively; Pearson's product–moment correlation; 1000 permutations). Permutational multivariate analysis of variance (PERMANOVA) was performed on normalised datasets to evaluate the effect of molecular processing and site on community compositions, using the function *adonis2* (vegan) with Jaccard dissimilarities (presence/absence) for metazoan and Bray-Curtis dissimilarities for prokaryotes and micro-eukaryotes. The rationale behind this choice is that metazoans are multicellular organisms of extremely varying numbers of cells, organelles, or ribosomal repeats in their genomes, and can also be detected through a diversity of remains. The number of reads can thus not be expected to reflect the abundance of detected OTUs. Significance was evaluated via marginal effects of terms, using 10,000 permutations with site as a blocking factor. Pairwise post-hoc comparisons were performed via the *pairwiseAdonis* package, with site as a blocking factor. Differences between samples were visualized via Principal Coordinates Analyses (PCoA) based on abovementioned dissimilarities.

Finally, taxonomic compositions in terms of cluster and read abundance were compared between molecular processing methods. In order to compare accurately phylum-level

taxonomic compositions, datasets were subsampled to clusters having a minimum hit identity of 86% for rRNA loci and 80% for COI. These values were chosen as they represent approximate minimum identity for reliable phylum assignment (Stefanni et al. 2018).

3 Results

3.1 High-throughput sequencing results

A total of 70 million 18S-V1 reads, 61 million COI reads, 30 million 18S-V4 reads, and 45 million 16S-V4V5 reads were obtained from four Illumina HiSeq runs of pooled amplicon libraries built from triplicate PCR replicates of 75 sediment samples, 2 mock communities (for 18S-V1 and COI), 3 extraction blanks, and 2-4 PCR negative controls(Table S6). One to seven sediment samples failed amplification in each dataset. These were always coming from the same sampling sites (MDW-ST117 and MDW-ST38), and predominantly comprised RNA samples (Table S6). After bioinformatic processing, read numbers were reduced to 44 million for 18S-V1, 45 million for COI, 16 million for 18S-V4, and 24 million for 16S-V4V5 (Table S6). For eukaryote markers, fewer reads were retained in negative controls (2-64%) than in true or mock samples (49-83%), while the opposite was observed for prokaryotes with 16S-V4V5 (62% of reads retained in control samples against 49-57% in true samples). Negative control samples (extraction and PCR blanks) contained 0.001-0.6% of total processed reads, compared to 1.3-1.5% in a true samples.

DNA extracts obtained from the joint DNA/RNA protocol based on the 2-g kit produced fewer eukaryotic reads than DNA extracts from the 10-g kit, while similar yields were obtained for prokaryotes. RNA extracts produced more reads than DNA extracts with the ribosomal loci, while they produced fewer reads with the mitochondrial COI locus (Table S6).

After data refining, abundance renormalisation (Wangensteen and Turon 2016), and LULU curation for COI, the final datasets comprised between 8.6 and 16.2 million target reads for eukaryotes and 21.7 million prokaryote reads. Target reads delivered 4,333 and 6,031 metazoan OTUs for COI and 18S-V1 respectively, 40,868 micro-eukaryote 18S-V4 ASVs, and 138,478 prokaryote 16S-V4V5 ASVs (Table S6).

3.2 Alpha diversity between processing methods

Rarefaction curves showed a plateau was reached for all samples, suggesting an overall sequencing depth adequate to capture the diversity present (Fig. S1). Processing methods significantly affected the number of recovered eukaryote and prokaryote clusters, and significant variability among sites was detected for 18S-V1 and 18S-V4 (Table 1, Fig. S2).

Molecular processing designed to remove small DNA fragments (i.e. size-selection of DNA to remove fragments < 1,000 bp and ethanol reconcentration) did not significantly affect recovered cluster numbers obtained from eDNA extracted from 10 g of sediment, for any of the loci investigated (Fig. 1, Table 1, Tukey's HSD multiple comparisons tests, p>0.9).

Extracts based on the 2-g kit resulted in more variability, reflected by greater standard errors in mean recovered cluster numbers (15-26% of the mean for eukaryotes, 7-9% for prokaryotes) than in DNA extracts based on 10 g of sediment (8-11% for eukaryotes, 3-6% for prokaryotes).

DNA extracted using the 2-g kit recovered significantly fewer eukaryotic clusters than extracts based on ~10 g of sediment (Fig. 1, Table 1), a trend consistent across most taxa (Fig. 2). DNA-2g extracts recovered an average of 110±16 18S-V1 and 113±27 COI metazoan OTUs per sample, compared to 264±26 (18S-V1) and 222±23 (COI) in the DNA-10g extracts. Similarly, DNA-10g extracts recovered on average 1,117±100 protistan 18S-V4 ASVs per sample, compared to 595±109 detected in DNA from the 2-g kit. Contrastingly to eukaryotes, all DNA methods, whether based on ~2 g or ~10 g of sediment, resulted in comparable prokaryote ASV numbers detected (Figs. 1-2, Table 1, p>0.8), ranging from 5,330 ±199 to 5,810 ±170 per sample on average.

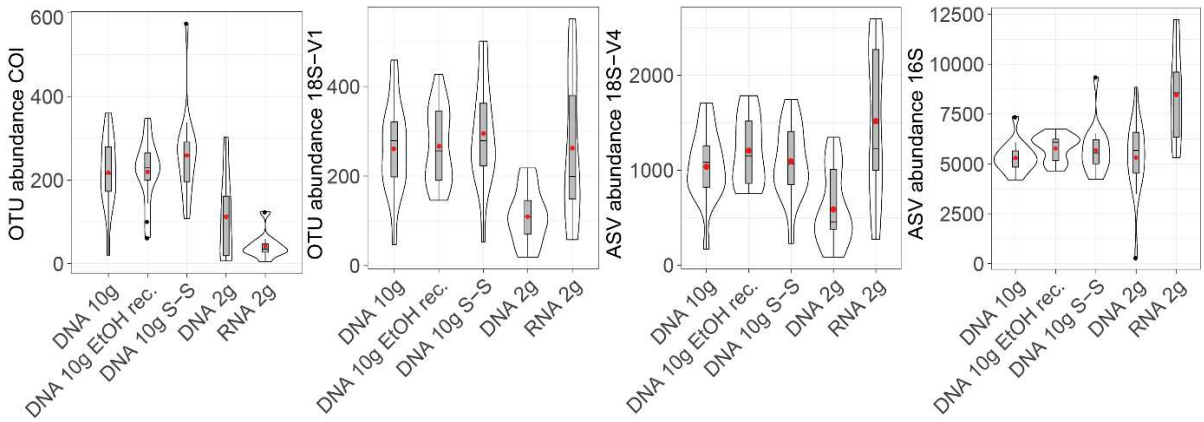


Figure 1. Violin plot showing detected numbers of metazoan OTUs (COI, 18S-V1), micro-eukaryote (18S-V4) ASVs, and prokaryote (16S) ASVs recovered by the five molecular processing methods evaluated in this study (DNA 10g: crude DNA extracts from ~10 g of sediment with the PowerMax Soil kit; DNA 10g EtOH rec.: ethanol reconcentrated 10g DNA extracts; DNA 10g S-S: size-selected 10g DNA extracts; DNA/RNA 2g: crude DNA/RNA extracts from ~2g of sediment with the RNeasy PowerSoil kit). Cluster abundances were calculated on rarefied datasets. Boxplots show medians with interquartile ranges. Red dots indicate mean values.

The joint RNA/DNA extracts shared 15% (COI) to 25% (18S-V1) of metazoan OTUs, 14% of protistan 18S-V4 ASVs, and 25% of prokaryotic 16S ASVs (Fig. S3). With COI, most unique OTUs were present in DNA extracts (74%), and RNA detected significantly fewer metazoan OTUs than co-extracted DNA (Fig. 1, mean of 44±12 vs. 113±27 respectively), a trend observed in most detected metazoan phyla (Fig. 2). Contrastingly, with ribosomal loci, most clusters were unique to RNA (56% for 18S-V1, 63% for 18S-V4, 45% for 16S, Fig. S3), which recovered significantly more clusters than co-extracted DNA (Fig. 1, Table 1). For prokaryotes, RNA extracts even detected significantly more ASVs than DNA extracts based on 10 g of sediment (Table 1, Fig. 1), a pattern observed in most prokaryotic clades, except for the Actinobacteria, Nanoarchaeaeota, Omnitrophicaeota, and the Thaumarchaeota (Fig. 2). For 18S-V4 and 18S-V1, RNA detected a cluster richness comparable to DNA-10 g extracts (Tukey's HSD multiple comparisons tests, p>0.16), yet, average cluster numbers per sample were higher in RNA than in DNA-10g extracts in numerous groups (Fig. 2).

Table 1. Changes in cluster richness and community structures with molecular processing method (DNA 10g: DNA extracts from ~10 g of sediment with the PowerMax Soil kit; DNA/RNA 2g: DNA/RNA extracts from ~2g of sediment with the RNeasy PowerSoil kit) and site, for the four studied genes. ANODEVs were performed on mixed models with negative binomial distributions using rarefied datasets. PERMANOVAs were calculated on normalised datasets by permuting 10,000 times with Site as a blocking factor, using Jaccard dissimilarities for 18S-V1 and COI, and Bray-Curtis dissimilarities for 18S-V4 and 16S. Significant p values are in bold, and significance codes are p<0.001: '***'; p<0.01: '**'; p<0.05: '*'.

LOCUS	Cluster richness			Community differentiation			
	Chi-square	***p-value***	**significant parwise comparisons**		**R^2**	***p-value***	**significant parwise comparisons**
18S-V1V2							
Molecular processing	50.3	**< 0.001**	DNA 2g < RNA 2g *** DNA 10g > DNA 2g ***	Molecular processing	0.06	**< 0.001**	DNA 2g / RNA 2g *** DNA 10g / DNA 2g *** DNA 10g / RNA 2g ***
Site (random effect)	16.2	**< 0.001**		Site	0.23	**< 0.001**	
				Molecular processing x Site	0.19	0.16	
COI							
Molecular processing	57.3	**< 0.001**	DNA 2g > RNA 2g ** DNA 10g > DNA 2g * DNA 10g > RNA 2g ***	Molecular processing	0.09	**< 0.001**	DNA 2g / RNA 2g ** DNA 10g / DNA 2g * DNA 10g / RNA 2g **
Site (random effect)	2.2	0.14		Site	0.20	**< 0.001**	
				Molecular processing x Site	0.17	**0.0013**	
18S-V4							
Molecular processing	38.3	**< 0.001**	DNA 2g < RNA 2g *** DNA 10g > DNA 2g **	Molecular processing	0.08	**< 0.001**	DNA 2g / RNA 2g ** DNA 10g / DNA 2g ** DNA 10g / RNA 2g **
Site (random effect)	15.9	**< 0.001**		Site	0.35	**< 0.001**	
				Molecular processing x Site	0.20	**< 0.001**	
16S-V4V5							
Molecular processing	55.0	**< 0.001**	DNA 2g < RNA 2g *** DNA 10g < RNA 2g ***	Molecular processing	0.06	**< 0.001**	DNA 2g / RNA 2g *** DNA 10g / DNA 2g ** DNA 10g / RNA 2g ***
Site (random effect)	3.4	0.07		Site	0.57	**< 0.001**	
				Molecular processing x Site	0.14	**< 0.001**	

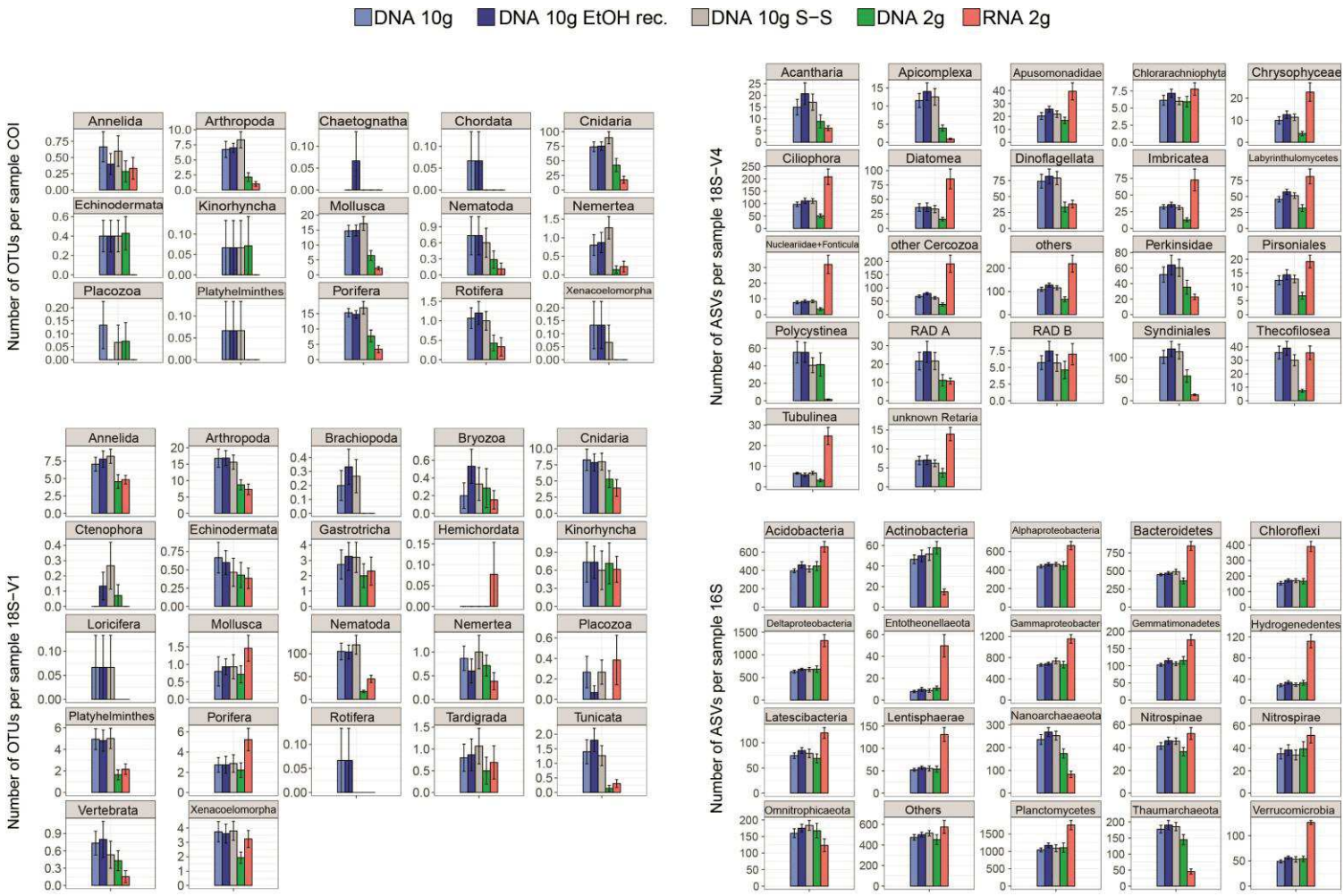


Figure 2. Mean number of metazoan OTUs (COI, 18S-V1), protist ASVs (18S-V4), and prokaryote ASVs (16S) detected per sample for each of the five processing methods (DNA 10g: crude DNA extracts from ~10 g of sediment with the PowerMax Soil kit; DNA 10g EtOH rec.: ethanol reconcentrated 10g DNA extracts; DNA 10g S-S: size-selected 10g DNA extracts; DNA/RNA 2g: crude DNA/RNA extracts from ~2g of sediment with the RNeasy PowerSoil kit). Cluster numbers were calculated on rarefied datasets. Error bars represent standard errors.

3.3 Effect of molecular processing methods on beta-diversity patterns

PERMANOVA showed that, although site was the main source of variation among samples (accounting for 20 to 57% of variability), significant differences existed among molecular methods in terms of community structure for all loci investigated over and above any variation due to site (Table 1). Pairwise comparisons indicated no significant effect of small DNA fragment removal on revealed community composition (Table 1), and high and significant correlations in Mantel tests (*r*: 0.92-1.0, p=0.001) confirmed the minor effect of size-selection and ethanol reconcentration. Based on these results, the size-selected and ethanol-reconcentrated DNA data were removed from further analyses, and community structures of the DNA-10g extracts were compared with those derived from co-extracted DNA/RNA using the 2g kit.

Pairwise comparisons showed significant differences in community structures between RNA and DNA for all markers analysed (Table 1). Ordinations, confirmed the predominant effect of site as the first two PCoA axes mostly resolved spatial effects (Fig. S4), but also revealed that communities detected by RNA differed from those detected by DNA (co-extracted DNA and DNA-10g), the level of differentiation varying among sites (Fig. 3). Pairwise comparisons also indicated significant differences in community structure between DNA extracts from the 2 g and 10 g kits (Table 1), possibly due to higher variability among replicate cores in the DNA-2g method as seen in ordinations (Fig. 3).

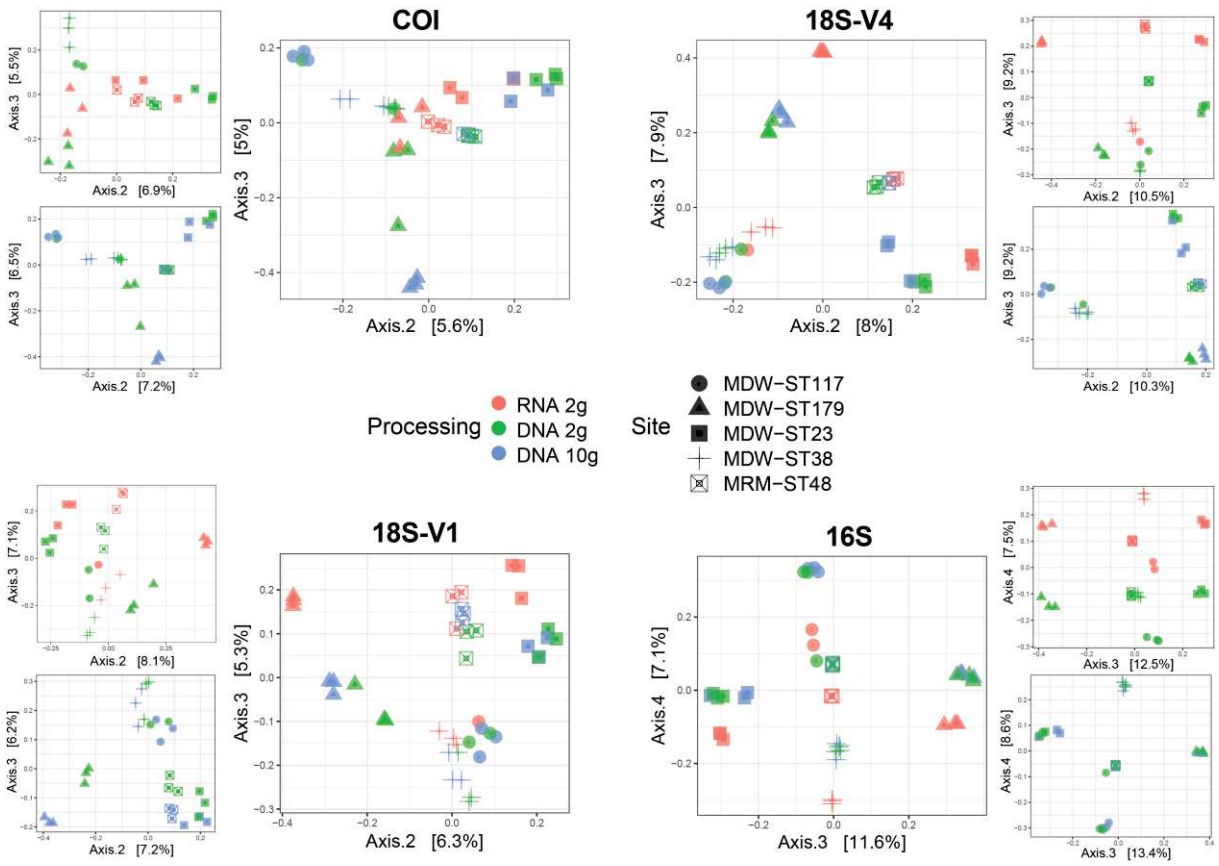


Figure 3. PCoA ordinations showing community differences between RNA and DNA molecular processing methods, using either RNA/DNA extracted jointly from ~2 g of sediment (RNA 2g/DNA 2g) or DNA extracted from ~10g of sediment (DNA 10g) in five deep-sea sites using four barcode markers targeting metazoans (COI, 18S-V1), micro-eukaryotes (18S-V4), and prokaryotes (16S). PCoAs were calculated using Jaccard dissimilarities for metazoans and Bray-Curtis dissimilarities for unicellular organisms. Inserts show pairwise PCoAs.

3.4 Extraction kit *vs* nature of nucleic acid

PERMANOVA of the dataset containing DNA-10g, DNA-2 g, and RNA-2g extracts confirmed that site was the predominant effect, explaining ~20% of variation for metazoans, 33% of variation for micro-eukaryotes, and 54% of variation for prokaryotes. The analysis also indicated that the differences observed between processing methods were predominantly due to the type of nucleic acid rather than the kit used for extraction. Nucleic acid nature (DNA *vs* RNA) led to significant differences among assemblages for all loci, while DNA extraction kit resulted in significant differences only for 18S-V1 and 18S-V4 (Table S7).

This supported observations in relative taxonomic compositions, which were more similar between samples based on DNA (Fig. 4), a pattern consistent across cores within each site (Fig. S5). Expectedly, when looking at read numbers, resolved taxonomic structures were also more similar among DNA-based methods (Fig. S6). Comparing read and cluster abundances revealed that relative taxonomic compositions based on read numbers (Fig. S6) were comparable to those based on cluster numbers (Fig. 4) for micro-eukaryotes and prokaryotes, and confirmed that this was not the case for metazoans.

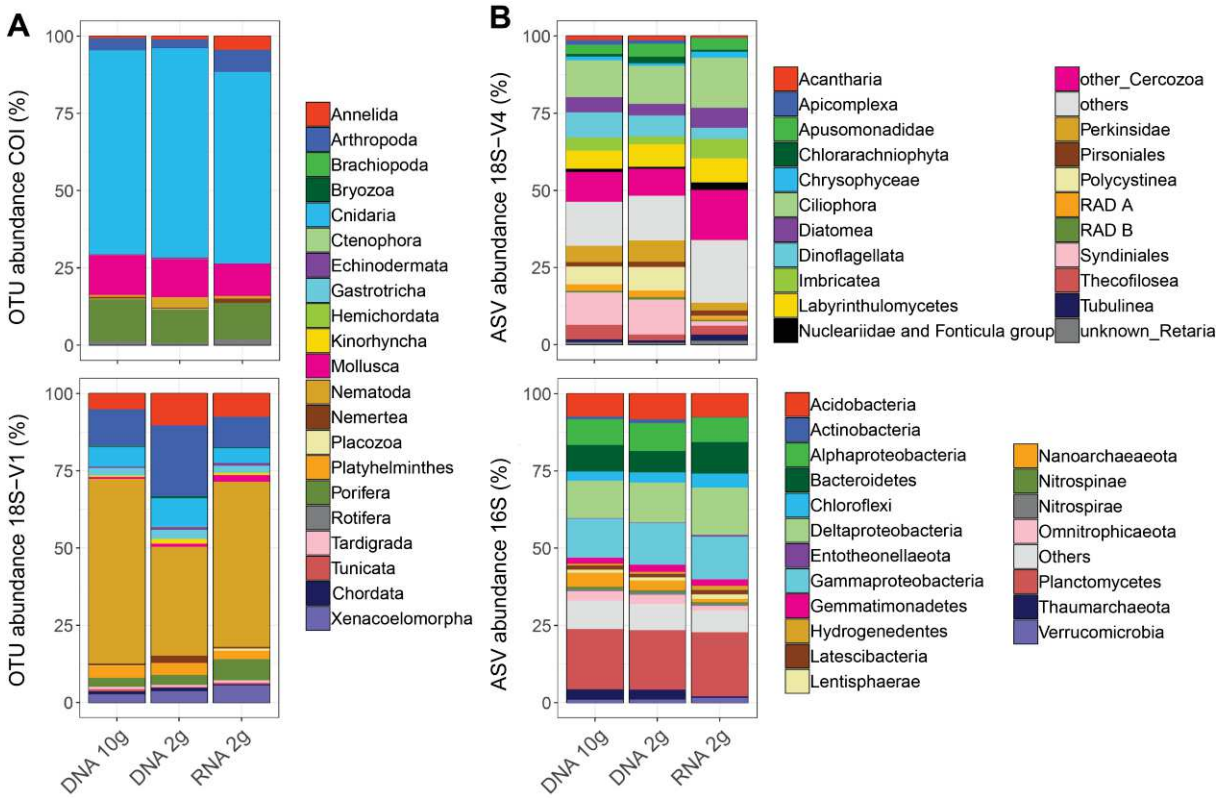


Figure 4. Patterns of relative cluster abundance resolved by metabarcoding of sediment RNA and DNA from five deep-sea sites, using either RNA/DNA extracted jointly from ~2 g of sediment (RNA 2g/DNA 2g) or DNA extracted from ~10g of sediment (DNA 10g), and using four barcode markers targeting metazoans (**A**: COI, 18S-V1), micro-eukaryotes (**B**: 18S-V4), and prokaryotes (**B**: 16S). Values were calculated on balanced datasets.

4 Discussion

The aim of this study was to evaluate different molecular methods in order to select the most appropriate eDNA metabarcoding protocol to inventory contemporary deep-sea communities, with the lowest possible bias due to aDNA.

Using RNA rather than DNA to inventory contemporaneous communities has been suggested as a means of avoiding the bias due to long-term persistence of DNA in marine sediments. Indeed, RNA is only produced by living organisms and is thought to quickly degrade when released in the environment, due to spontaneous hydrolysis and the abundance of RNases (Torti et al. 2015). Expectedly, in our COI dataset, RNA resulted in fewer OTUs (Fig. 1) and detected fewer phyla (Fig. 2) than co-extracted DNA. Contrastingly, for ribosomal loci, RNA detected higher cluster numbers than co-extracted DNA (Fig. 1), resulting in more clusters per sample for most of the taxonomic groups detected (Fig. 2). In these joint datasets, 45-63% of clusters were unique to RNA (Fig. S2). These unique clusters were not singleton clusters as only up to 2.2% of them had fewer than three reads, even if 5-28% had fewer than ten reads (data not shown). Although proportions vary strongly among investigations, other studies using ribosomal loci have also reported increased recovery of OTUs in RNA datasets as well as considerable amounts of unshared OTUs between joint RNA and DNA data (Guardiola et al., 2016; Laroche et al., 2017 and references therein).

This difference observed here between COI and ribosomal loci is likely related to the nature of the targeted RNA molecule. The rapid hydrolysis of RNA mostly applies to random coils (like messenger RNA), while helical conformations (including most types of RNA, such as ribosomal RNA, transfer RNA, viral genomic RNA, or ribozymes) are less prone to hydrolysis by water molecules (Torti et al. 2015). The degradation of rRNA is thus likely to be much slower than that of messenger RNA, which, combined with decreased digestion by RNases due to adsorption onto sediment particles (Torti et al. 2015), makes long-term persistence of rRNA possible, and observed in sediments and even in fossils (Orsi et al. 2013; Cristescu 2019).. Finally, the great abundance of RNA over DNA in living organisms (e.g. 20.5% vs 3.1% in *E. coli*) may also favour its persistence in the environment. This is especially true for rRNA, which is represented in a cell's RNA pool as many times as there are ribosomes, while only being present in a few copies (10-150) in the genome (Torti et al. 2015).

While RNA has been reported as an effective way to depict the active community compartment (Baldrian et al. 2012; Lejzerowicz et al. 2013b; Pawlowski et al. 2014), variation in activity levels between taxonomic groups as well as differences in life histories, life

strategies, and non-growth activities may confound this interpretation and generate taxonomic bias (Blazewicz et al. 2013). Instead, DNA/RNA ratios might reflect different genomic architectures (variation in rDNA copy number) among taxonomic groups, rather than different relative activities (Massana et al. 2015). Thus, environmental RNA data need to be interpreted with caution, as some molecular clusters could be overrepresented due to increased cellular activities (Pochon et al. 2017). This could explain the higher cluster numbers detected here for ribosomal loci with eRNA compared to eDNA for several taxa (Fig. 2).

Moreover, many of the unique RNA ASVs/OTUs may be artefacts from the reverse transcription of RNA to cDNA, a process known to generate errors that are difficult to measure and detect in bioinformatic analyses (Laroche et al. 2017), but highlighted by the greater amounts of chimeras detected in RNA extracts with ribosomal loci (Table S6). This overestimation of RNA-based data will affect non-clustered data more than clustered datasets, in line with the results observed here for microbial ASVs and metazoan OTUs.

In terms of beta diversity patterns, although RNA and DNA detected significantly different communities (Table 1), DNA and RNA samples resolved similar spatial configurations, with samples clustering by site (Fig. 3). This is consistent with Guardiola et al. (2016), who also reported similar patterns of ecological differentiation between DNA and RNA in deep-sea sites although both datasets resolved different communities. Although the comparative study performed here targeted only the first 1 cm layer of sediment, the comparable results obtained by Guardiola et al (2016) on 5 cm suggest these findings may be expanded to deeper layers of sediments. However, spatial variation was more pronounced with DNA samples for eukaryotes, which is congruent with Laroche et al. (2017), who suggested that eDNA may be more reliable for assessing differences in community composition.

Thus, due to its suspected persistence in the environment, and the unknown but potentially additional sources of bias suspected here, using eRNA for metabarcoding of deep-sea sediments does not seem to effectively address the problem of aDNA, and even less so for ribosomal loci. Other studies suggested that a more efficient way to deal with aDNA may be to use joint RNA and DNA datasets, and trim for shared OTUs (Laroche et al. 2017; Pochon et al. 2017). This is however particularly stringent (given the low shared OTU proportions observed in this and other studies), and may result in a substantial number of false negatives. With COI, while mRNA may be more effectively targeting living organisms, the approach remains confronted with the taxonomic bias mentioned above, combined with higher *in vitro* lability of mRNA

making it more challenging to work with (highlighted by the increased failure of RNA extracts in this study, Table S6).

Removing small DNA fragments via size-selection (removing fragments < 1,000 bp) or ethanol reconcentration did not affect recovered cluster numbers in any of the biological compartments investigated (Fig. 1). The methods also did not result in any significant difference in community structures (Table 1), suggesting that small, likely ancient, DNA fragments have a negligible impact on biodiversity inventories produced through eDNA metabarcoding. This finding is in line with results from the deep-sea (Guardiola et al. 2016; Ramírez et al. 2018) and various other habitats (Lennon et al. 2018), which showed no evidence that spatial patterns were blurred by ‘‘dead’’ DNA persistence, and suggested a minimal effect of extracellular DNA on estimates of taxonomic and phylogenetic diversity.

None of the methods evaluated in the present study remove DNA not enclosed in *living* cells (e.g. DNA in organelles, DNA from dead cells…). It is still unclear how long DNA can remain intracellular after cell death or within organelles. Future research quantifying the rate at which “dead” intracellular DNA becomes extracellular and degraded, and investigation of deeper layers of sediment, will be valuable to estimate the potential bias of archived intracellular DNA in eDNA metabarcoding inventories of extant communities. However, there is increasing evidence that DNA from non-living cells is mostly contemporary (Lennon et al. 2018). This ability to detect extant taxa that were not present in the sample at the time of collection highlights the capacity of eDNA metabarcoding to detect local presence of organisms even from their remains or excretions, and even with a small amount of environmental material.

It remains to be elucidated whether more cost and time effective extraction protocols specifically targeting extracellular DNA offer similar ecological resolution as total DNA kits. This is suggested to be the case for terrestrials soils (Zinger et al. 2016; Taberlet et al. 2012b), although authors have highlighted that conclusions from these studies should be interpreted with caution as results might be influenced by actively released and ancient DNA (Nagler et al. 2018). The only available study testing this in the deep-sea showed that richness patterns were strikingly different in several metazoan phyla between extracellular DNA and total DNA. The authors suggested this to be the result of activity bias: sponges and cnidarians were overrepresented in the extracellular DNA pool because they continuously expel DNA, while nematodes were underrepresented as their cuticles shield DNA (Guardiola et al. 2016). As this comparison was performed on samples collected in two consecutive years, differences observed

may partly result from temporal variation. However, another study of shallow and mesobenthic macroinvertebrates showed that targeting solely the extracellular eDNA compartment of marine sediments led to the detection of more than 100 taxa fewer than bulk metabarcoding or morphology, suggesting that extracellular DNA may not be adequate for marine sediments (Aylagas et al. 2016).

Larger amounts of sediment (≥10g) allowed detecting significantly more eukaryotic clusters. This was not true for prokaryotes, for which both ~2 and ~10 g of sediment detected similar numbers of ASVs (Table 1, Fig. 1). It may be suggested that in the joint RNA/DNA kit, DNA elution occurring after RNA elution induces partial DNA loss. However, such effect would be expected to equally affect eu- and prokaryotes, which was not the case here, supporting the fact that quantity of starting material significantly affects results for eukaryotes. The importance of adjusting the amount of starting material to the biological compartment investigated has already been documented (Creer et al. 2016; Dopheide et al. 2019), and this study confirms that while 2-5 g of deep-sea sediment may be enough to capture prokaryote diversity, microbial eukaryotes and metazoans are more effectively surveyed with larger sediment volumes.

Finally, the ~2 g protocols were generally associated to higher variability among replicate cores for all loci investigated (Fig. 1, Fig. 3). This variability increases confidence intervals, reduces statistical power, and increases the risk of not identifying differences among communities, and thus impacts in EIA studies (Type II errors). Small-scale (cm to metres) patchiness has often been reported in the deep-sea (Grassle and Maciolek 1992; Smith, C. R. and Snelgrove 2002; Lejzerowicz et al. 2014). While technical (PCR) replicates allow increasing taxon detection probability (decrease false positives), this within-site variability can only be mitigated by collecting more biological replicates per sampling station, and using a sufficiently high amount of starting material to extract nucleic acids.

Acknowledgements

This work is part of the "*Pourquoi Pas les Abysses?*" project funded by Ifremer, and the project eDNAbyss (AP2016 -228) funded by France Génomique (ANR-10-INBS-09) and Genoscope-CEA. This work also received funding from the European Union's Horizon 2020 research and innovation programme under grant the agreement No 678760 (ATLAS), and the 'Investissements d'Avenir' program OCEANOMICS (ANR-11-BTBR-0008; NH and CV).

This output reflects only the authors' view and the European Union cannot be held responsible for any use that may be made of the information contained therein. We wish to thank Laure Quintric, Patrick Durand, Caroline Belser, and Stéphane Pesant for bioinformatic and archiving support, as well as Jan Pawlowski and Eva Ramirez Llodra for useful advice on this work. We also wish to express our gratitude to the crew, participants, and mission chiefs of the MarMine cruise (Eva Ramirez Llodra, project 247626/O30, funded by the Research Council of Norway and associated industrial partners) and the MEDWAVES cruise supported by the ATLAS project and the Spanish Ministry of Economy, Industry and Competitivity (Covadonga Orejas and all the crew from the Sarmiento de Gamboa), as well as to all the people who helped collecting samples (Perregrino Cambeiro, Juancho Movilla, Maria Rakka, Joana Boavida, Anna Addamo). This manuscript has been released as a Pre-Print at https://doi.org/10.1101/836080.

Author contributions

MB, CL-H, and SA-H designed the study, MB, JP, and CL-H carried out the laboratory work. MB and BT performed the bioinformatic analyses. MB, BT, and NH performed the statistical analyses. MB and SA-H wrote the manuscript. All authors contributed to the final manuscript.

Data accessibility

The data for this work has been submitted to the European Nucleotide Archive (ENA) under the following project: PRJEB33873. . Please refer to the sample metadata in Supplementary Material to download samples from ENA. Additionally, the full dataset, including raw sequences, databases, ASV/OTU tables, and scripts (Gitlab link) are available through https://doi.org/10.12770/cf00aa7b-67e7-49c4-8939-038c4a9d887f.

Conflicts of Interest

The authors declare no competing interests.

Chapter IV. Evaluating sediment and water sampling methods for the estimation of deep-sea biodiversity using environmental DNA

Manuscript accepted with minor revisions at Scientific Reports:

Brandt MI, Pradillon F, Trouche B, Henry N, Liautard-Haag C, Cambon-Bonavita MA, Wincker P, Poulain J, Arnaud-Haond S, and Zeppilli D. 2021. Evaluating sediment and water sampling methods for the estimation of deep-sea biodiversity using environmental DNA.

Size-class sorting such as sieving or elutriation is usually performed on sediment samples in order to split the organisms by size and facilitate morphological characterization of meiofauna and macrofauna. For metabarcoding approaches, it also has the advantage of limiting the over dominance of large organisms in DNA extracts. However, sieving requires larger volumes of sediment and is very time-consuming, and studies have found that the use of non-sieved material does not significantly alter metazoan diversity patterns (Sinniger et al. 2016), suggesting DNA dominance of large bodied taxa does not result in very important biases. Besides, for logistic reasons, the analysis of non-sieved samples is preferable to (1) minimize on-board processing for the team involved, (2) minimize risks of contaminations, and (3) keep the extracellular DNA for potential future studies.

Application of eDNA metabarcoding on deep-sea aboveground water could be useful to evaluate dispersal capacities of benthic organsims as well as benthopelagic diversity. However, sampled water volume, a crucial aspect for efficient species detection, has been variable among studies and it remains unclear whether small volumes (1-2 L) are sufficient for species detection in the deep-sea.

Questions addressed:

Does sieving sediment allow achieving more reliable biodiversity inventories in the deep-sea? How does the SALSA *in situ* pump compare with traditional water sampling devices for biodiversity detection of aboveground water samples?

Résumé en français

Bien qu'elle représente l'un des plus grands biomes du monde, la biodiversité des grands fonds marins est encore mal connue. Le métabarcoding sur ADN environnemental offre des perspectives inédites pour des inventaires et des études d'impact rapides, mais nécessite des méthodes d'échantillonnage standardisées et un choix judicieux de substrat environnemental. Ici, nous avons cherché à optimiser l'évaluation génétique des communautés procaryotes (16S), protistes (18S V4) et métazoaires (18S V1-V2, COI), en évaluant des stratégies d'échantillonnage pour les sédiments et les eaux profondes affleurant le sédiment, déployées simultanément à un site abyssal. Pour les sédiments, alors que le tri des classes de taille par tamisage n'a eu aucun effet sur la diversité alpha totale détectée et a résolu des compositions taxonomiques similaires au niveau du phylum pour tous les marqueurs étudiés, il a effectivement augmenté la détection des phylums de la méiofaune. Pour l'eau, de grands volumes obtenus à partir d'une pompe *in situ* (~ 6000 L) ont détecté beaucoup plus de diversité métazoaire que 7,5 L collectés dans des boîtes d'échantillonnage. Cependant, la pompe étant limitée par des mailles plus grandes (> 20 µm), ne capturait qu'une fraction de la diversité microbienne, tandis que des boîtes d'échantillonnage permettaient d'accéder au pico- et au nanoplancton. Plus important encore, les communautés caractérisées par les échantillons d'eau affleurante différaient significativement de celles caractérisées par des sédiments, quel que soit le volume utilisé, et les deux types d'échantillons ne partageaient qu'entre 5% et 10% des unités moléculaires. Ensemble, ces résultats soulignent que le tamisage peut être recommandé pour cibler la méiofaune, et que les eaux affleurantes ne représentent pas une alternative à l'échantillonnage des sédiments pour les inventaires de la diversité benthique.

Abstract

Despite representing one of the largest biomes on earth, biodiversity of the deep seafloor is still poorly known. Environmental DNA metabarcoding offers prospects for fast inventories and surveys, yet requires standardized sampling approaches and careful choice of environmental substrate. Here, we aimed to optimize the genetic assessment of prokaryote (16S), protistan (18S V4), and metazoan (18S V1-V2, COI) communities, by evaluating sampling strategies for sediment and aboveground water, deployed simultaneously at one deep-sea site.

For sediment, while size-class sorting through sieving had no effect on total detected alpha diversity and resolved similar taxonomic compositions at the phylum level for all markers studied, it effectively increased the detection of meiofauna phyla. For water, large volumes obtained from an *in situ* pump (~6000 L) detected significantly more metazoan diversity than 7.5 L collected in sampling boxes. However, the pump being limited by larger mesh sizes (> 20 µm), only captured a fraction of microbial diversity, while sampling boxes allowed access to the pico- and nanoplankton. More importantly, communities characterized by aboveground water samples significantly differed from those characterized by sediment, whatever volume used, and both sample types only shared between 5% and 10% of molecular units.

Together, these results underline that sieving may be recommended when targeting meiofauna, and aboveground water does not represent an alternative to sediment sampling for inventories of benthic diversity.

1 Introduction

Environmental DNA (eDNA) metabarcoding is an increasingly used tool for non-invasive and rapid biodiversity surveys and impact assessments. Using high-throughput sequencing (HTS) and bioinformatic processing, target organisms are detected using their DNA directly extracted from soil, water, or air samples (Taberlet et al. 2012a). Covering more than 50% of Planet Earth, the deep seafloor is mostly comprised of sedimentary habitats, characterised by a predominance of small organisms (Rex et al. 2006; Snelgrove 1999) difficult to identify based on morphological features (Carugati et al. 2015), and by high local and regional diversity (Grassle and Maciolek 1992; Smith, C. R. and Snelgrove 2002; Hauquier et al. 2019). Given its increased time-efficiency and its wide taxonomic applicability, eDNA metabarcoding is thus a good candidate for large-scale biodiversity surveys and Environmental Impact Assessments in the deep-sea biome.

Size-class sorting such as sieving or elutriation is usually performed on sediment samples in order to split the organisms by size and facilitate morphological characterization of meiofauna and macrofauna. For metabarcoding approaches, it also has the advantage of limiting the over dominance of large organisms, which may produce higher amounts of DNA template, resulting in an incomplete detection of small and abundant taxa. However, sieving requires large volumes of sediment, is very time-consuming, and previous studies have found that the use of non-sieved material does not significantly alter metazoan diversity patterns (Sinniger et al. 2016), suggesting that dominance of large (and often rare) taxa in the DNA extract does not result in important biases. Besides, for logistic reasons, the use of non-sieved sediment samples is preferable to (1) minimize on-board processing time, (2) minimize risks of contamination, and (3) allow other future applications (e.g., characterization of microbial communities, RNA sequencing, and investigation of extracellular DNA).

Finally, studies from various marine habitats have reported that benthic taxa could be found in aboveground water (overlying water layer to 6.5 m above seafloor), possibly due to sediment resuspension and transport, but also to active dispersal (Boeckner et al. 2009; Klunder et al. 2020; Zhao et al. 2020). Application of eDNA metabarcoding on deep-sea aboveground water could thus be a convenient alternative to surface sediment collection, as it involves simplified sample processing and shipping, while additionally allowing investigating benthopelagic diversity and dispersal capacities of benthic organisms. However, distance above seafloor has been variable (0.5 m – 6.5 m) among studies (Boeckner et al. 2009; Klunder et al. 2020; Zhao et al. 2020), and so has the water volume sampled (12 L - 1,000 L). As the latter is a crucial

aspect for efficient species detection (Cantera et al. 2019), it remains unclear whether small volumes (< 10 L) are sufficient to obtain comprehensive species inventories in the deep-sea.

To evaluate the effect of sampling strategy on eDNA metabarcoding inventories targeting prokaryotes (16S V4-V5), unicellular eukaryotes (18S V4), and metazoans (18S V1-V2, COI) from deep-sea sediment and aboveground water, we compared biodiversity inventories resulting from 1) sieved *vs.* unsieved sediment and 2) on-board filtration of ~7.5 L of water collected with a sterile sampling box *vs. in situ* filtration of large volumes (~6,000 L) using a newly-developed pump.

2 Materials and methods

2.1 Sample collection

Sediment cores and water samples were collected from a continental slope site during the EssNaut16 cruise in the Mediterranean in April 2016 (Table S1). Sampling was carried out with a human operated vehicle (Nautile, Ifremer). Two sediment sampling methods were compared on triplicate tube cores, using the upper first centimetre sediment layer. The sediment samples were either 1) transferred into zip-lock bags and frozen at −80°C on board or 2) sieved through five different mesh sizes (1,000 µm, 500 µm, 250 µm, 40 µm, and 20 µm) in order to concentrate organisms and separate them by size-class. Sieving was performed with cold surface water filtered at 0.2 µm. Each mesh concentrate was subsequently stored in a separate zip-lock bag and frozen at −80°C. All samples were shipped on dry ice to the laboratory.

Two different aboveground water-sampling methods were evaluated during EssNaut16 to target microbial and metazoan taxa. All water samples were collected at most 1 m above the seafloor. Water was collected with a newly developed *in situ* pump, the Serial Autonomous Larval Sampler (SALSA), i.e. a McLane WTS-LV sampler adapted by Ifremer, Brest, France to allow replicated sampling. SALSA pumps up to 30,000 L of seawater through a 20-µm nylon mesh, concentrating this water into five 2.8 L sampling bowls that can be used as biological replicates, each representing ~6,000 L of concentrated seawater targeting the > 20 µm size fraction. Two deployments were performed at the study site (PL07, PL11) and one deployment within the same habitat but at shallower depth due to technical reasons impeding deployment at the original site (PL09). Analyses were performed with and without PL09, and as no significant difference was observed between deployments, results from PL09 were included in the study. Two replicates per SALSA deployment were used in this work. Each replicate was filtered on board through polycarbonate membrane filters with 2-µm mesh size (Millipore, Burlington, MA, USA, ref. TTTP04700) to concentrate all retained particles on the filter membrane. Water was also collected using two ~7.5 L Nautile-deployed sterile and watertight sampling boxes (Roussel et al. 2011). These samples were filtered on board successively through membrane filters with 20 µm, 2 µm, and 0.2 µm mesh size (Millipore, Burlington, MA, USA, refs. NY2004700, TTTP04700, GTTP04700), generating three size fractions (>20 µm, 2-20 µm, and 0.2-2 µm). Each water filter was stored in an individual Petri dish, frozen at −80°C, and shipped on dry ice to the laboratory.

2.2 Nucleic acid extractions

For sediment, DNA extractions were performed using 2-10 g of sediment with the PowerMax Soil DNA Isolation Kit (MOBIO Laboratories Inc.; Qiagen, Hilden, Germany). All DNA extracts were stored at −80°C. For sieved sediment, DNA was extracted from each size fraction separately, and an equimolar pool of the DNA extracts of each size fraction was prepared for PCR and sequencing. Water DNA extractions were carried out by Genoscope (Évry, France) using the same protocol as described by Alberti et al. (2017) for Tara Oceans water samples. The protocol is based on cryogenic grinding of membrane filters, followed by nucleic acid extraction with NucleoSpin RNA kits combined with the NucleoSpin DNA buffer set (Macherey-Nagel, Düren, Germany). A negative extraction control was performed alongside sample extractions for both water and sediment samples (adding nothing instead of sample).

2.3 PCR amplification and sequencing

DNA extracts were normalised to 0.25 ng/µL and 10 µL of standardized sample were used for PCR (see Supporting Information for amplification details). Four primer pairs were used to amplify one mitochondrial and three ribosomal RNA (rRNA) barcode loci preferentially targeting metazoans (COI, 18S V1-V2), unicellular eukaryotes (18S V4), and prokaryotes (16S V4-V5, Table S2). PCR amplifications for each locus were carried out in triplicate in order to smooth intra-sample variance while obtaining sufficient amounts of amplicons for Illumina sequencing. PCR triplicates were pooled and amplicon libraries were prepared for sequencing by ligation of Illumina adapters on 100 ng of amplicons, following the Kapa Hifi HotStart NGS library Amplification kit (Kapa Biosystems, Wilmington, MA, USA). After quantification and quality control, library concentrations were normalized to 10 nM, and 8–9 pM of each library containing a 20% PhiX spike-in were sequenced on a HiSeq2500 (System User Guide Part # 15035786) instruments in a 250 bp paired-end mode. For sediment samples, this procedure was carried out on two DNA aliquots, leading to two triplicate amplicon libraries per sample. For water samples collected with the sampling box, the three size fractions were processed separately but, expectedly due to the differential size of micro- and macroorganisms, not all could be successfully amplified or sequenced for each locus. For metazoans, the data thus

comprise the 2-20 µm and > 20 µm size fractions, while for microbial loci the data comprise the 0.2-2 µm and 2-20 µm size fractions.

2.4 Bioinformatic analyses

All bioinformatic analyses were performed using a Unix shell script, available on Gitlab (https://gitlab.ifremer.fr/abyss-project/), on a home-based cluster (DATARMOR, Ifremer), and the samples of the present study were analysed in parallel with 12 to 28 other deep-sea water samples for more accurate error correction and LULU filtering. The details of the pipeline, along with specific parameters used for all metabarcoding markers, are given in Table S3.

Pairs of Illumina reads were corrected with DADA2 v.1.10 (Callahan et al. 2016), following the online tutorial for paired-end data (https://benjjneb.github.io/dada2/tutorial.html), delivering inventories of Amplicon Sequence Variants (ASVs). Data from COI and 18SV1-V2, preferentially targeting metazoans, were further clustered into Operational Taxonomic Units (OTUs) with swarm v2 (Mahé et al. 2015) using the FROGS pipeline (Escudié et al. 2018). Swarm v2 is a single-linkage clustering algorithm that aggregates sequences iteratively and locally around seed sequences based on *d*, the number of nucleotide differences, to determine coherent groups of sequences. This avoids a universal clustering threshold, which is particularly useful in highly biodiverse samples such as those analysed in this study. Metazoan ASVs were swarm-clustered at $d=3$ for 18S V1-V2 and $d=6$ for COI, which has been shown to be appropriate for evaluating species diversity in samples (Brandt, M. I. et al. 2020). We chose to evaluate unicellular eukaryote and prokaryote diversity at the ASV level due to their reproducibility and increasing use in the literature (Callahan et al. 2017). Although the use of OTUs may be justified for microbial diversity depending on study objectives (Brandt, M. I. et al. 2020), we did not expect a significant alteration of alpha and beta diversity patterns between ASV and OTU levels for the different sampling methods investigated.

Clusters were taxonomically assigned with BLAST+ (v2.6.0) based on minimum similarity (70%) and minimum coverage (80%). The Silva132 reference database was used for taxonomic assignment of the 16S V4-V5 and 18S V1-V2 rRNA marker genes (Quast et al. 2012), PR2 v4.11 (Guillou et al. 2013) was used for 18S V4, and MIDORI-UNIQUE (Machida et al. 2017) reduced to marine taxa only was used for COI. Considering our interest in diverse and poorly characterized communities, more stringent BLAST thresholds were not implemented at this

stage. Indeed, it is not uncommon for deep-sea taxa to have closest relatives in databases (even congenerics) exhibiting nucleotide divergence exceeding 20% (Shank et al. 1999; Herrera et al. 2015). However, additional filters were performed during downstream bioinformatic processing described below, and only clusters with assignments reliable at phylum-level were retained in the analysis.

Molecular inventories were refined in R v.3.5.1 (R Core Team 2018). A blank correction was made using the *decontam* package v.1.2.1 (Davis et al. 2018), removing all clusters that were more prevalent in negative control samples (PCR and extraction controls) than in true samples. After comparison, results from the technical duplicates produced for sediment samples were merged and read counts were summed for identical OTUs. Clusters unassigned at phylum-level and non-target clusters were removed. Additionally, for metazoan loci, all clusters with a terrestrial assignment (groups known to be terrestrial-only) were removed. Samples were checked to ensure they had more than 10,000 target reads. Metazoan OTU tables were further curated with LULU v.0.1 (Frøslev et al. 2017) to limit bias due to intraspecific variation and pseudogenes, using a minimum co-occurrence of 0.90 and a minimum similarity threshold of 84% for COI and 90% for 18S V1-V2.

Finally, refined datasets were taxonomically filtered by retaining only clusters having a minimum hit identity of 86% for rRNA loci and 80% for COI. These values were chosen as they represent approximate minimum identity for reliable phylum assignment (Stefanni et al. 2018).

2.5 Statistical analyses

Data were analysed using R with the packages phyloseq v1.22.3 (McMurdie and Holmes 2013), following guidelines in online tutorials (http://joey711.github.io/phyloseq/tutorials-index.html), and vegan v2.5.2 (Oksanen et al. 2018). Read and cluster abundances were evaluated via analyses of variance (ANOVA) on generalised linear models using quasipoisson distributions. Pairwise post-hoc comparisons were performed via Tukey HSD tests using the *emmeans* package. Alpha and beta diversity were compared among sampling methods using datasets rarefied to the minimum sequencing depth (COI: 60,242; 18S V1: 118,401; 18S V4: 33,037; 16S: 100,205). Differences in community composition were assessed with Venn diagrams (computed using the *venn* function in the venn package) and with permutational multivariate analysis of variance (PERMANOVA). The latter were performed using the

adonis2 function (vegan) and significance was evaluated using 1,000 permutations. Incidence-based Jaccard dissimilarities were used for metazoans, while Bray-Curtis dissimilarities were used for prokaryotes and unicellular eukaryotes. The rationale behind this choice is that metazoans are multicellular organisms of extremely varying numbers of cells, organelles, or ribosomal repeats in their genomes, and can also be detected through a diversity of remains. The number of reads can thus not be expected to reliably reflect the abundance of detected OTUs. Pairwise post-hoc comparisons among sampling methods were performed with the *pairwiseAdonis* package. Differences among samples were visualized via Principal Coordinates Analyses (PCoA) based on abovementioned dissimilarities. Finally, taxonomic compositions in terms of cluster abundance were compared among processing methods.

3 Results

3.1 High-throughput sequencing results

A total of 19 million raw 18S V1-V2 reads, 26 million COI reads, 14 million 18S V4 reads, and 17 million 16S V4-V5 reads were obtained from Illumina HiSeq runs of amplicon libraries built from pooled triplicate PCRs of 22 environmental samples, 2 extraction blanks, and 4-6 PCR negative controls (Table S4). The *in situ* pump yielded less raw reads for COI and 16S (Fig. S1, F = 4.02-14.4, p = 0.0003-0.03), while more raw reads were recovered from both water sampling methods with 18S V4 (F = 6.4, p = 0.007). Water samples generally yielded fewer raw clusters (F = 5.2-35.1, p = 3.2×10^{-6}-0.02), except for 18S V4 where numbers were comparable across sample types (Fig. S1).

Bioinformatic processing (quality filtering, error correction, chimera removal, and clustering for metazoans) reduced read numbers to 12 million for 18S V1-V2, 20 million for COI, 11 million for 18S V4, and 10 million for 16S V4-V5, resulting in 17,009 and 10,350 raw OTUs for 18S V1-V2 and COI respectively, 35,538 raw 18S V4 ASVs, and 62,646 raw 16S ASVs (Table S4). For eukaryote markers, less reads were retained in PCR banks (17-55%) than in true samples (52-87%) or extraction blanks (50-75%). Negative controls contained less processed reads (1-2% per sample) than true samples (4-5% per sample). In contrast, more 16S V4-V5 reads were retained in control samples (67-87%) than in true samples (29-73%), and control samples generated more processed reads than true samples (5% per blank *vs.* 3% per true sample). After data refining (decontamination, removal of all control samples and of all

unassigned or non-target clusters), rarefaction curves showed a plateau was reached for all samples except sediment samples with 18S V4, suggesting that not all sediment protist diversity was captured at this sequencing depth (Fig. S1). LULU curation (only for metazoan data) and taxonomic refinement (removal of clusters with assignments not reliable at phylum-level, i.e. < 86% BLAST identity for rRNA loci, and < 80% for COI), resulted in final datasets that comprised between 4.6 and 5.8 million target reads for eukaryotes and 7 million 16S V4-V5 for prokaryotes. Target reads delivered 405 (18S V1-V2) and 507 (COI) metazoan OTUs, 7,081 protist ASVs (18S V4), and 38,816 prokaryote 16S ASVs (Table S4).

3.2 Alpha diversity between sampling methods

Significantly fewer molecular clusters were detected in water samples than in sediment samples for all loci except 18S V4 where both sample types recovered similar levels of diversity (Table 1, Fig. 1). However, this trend was not consistent across taxonomic groups, as recovered diversity in each sample type strongly differed depending on phylum (Fig. 2).

For metazoans, water samples led to the detection of a significantly higher number of OTUs for Arthropoda (COI and 18S V1-V2), Rotifera (COI), and Ctenophora (18S V1-V2, t-tests, p = 0.002-0.04) than sediment samples, and some phyla like Chordata, Echinodermata, Gastrotricha, or Brachiopoda were only detected in water samples (Fig. 2). In contrast, phyla such as Cnidaria, Mollusca, Platyhelminthes, Porifera, Nematoda, or Xenacoelomorpha produced significantly more OTUs in sediment than water samples (t-tests, $p = 4x10^{-5}$- 0.02), and kinorhynchs or tardigrades were only recovered with sediment samples (Fig. 2).

Similarly, some protistan groups, such as the Acantharea, Chlorophyta, Dinophyceae, Haptophyta, and Syndiniales (Fig. S2), were predominant in water samples (t-tests, p = 0.002 – 0.04), while others were significantly more diverse in sediment (e.g., Filosa groups, Ciliophora, Labyrinthulea, t-tests, p = 0.02-0.04).

For prokaryotes, most lineages were predominant in sediment (t-tests, $p = 4.4x10^{-7}$-0.02 e.g., Archaea, Acidobacteria, Actinobacteria, Bacteroidetes, Chloroflexi, Dadabacteria, Delta-, Gammaproteobacteria, Gemmatimonadetes, Lentisphaerae Nitrospirae, Patescibacteria, Planctomycetes), and only Cyanobacteria were significantly more diverse in water samples (t-test, p = 0.0009).

Sieved and unsieved sediment resulted in comparable total cluster numbers in all loci investigated (Table 1, Fig. 1). However, recovered levels of alpha diversity varied by phyla and organism size class (Fig. 2). For metazoans, more OTUs were detected from sieved then from unsieved sediment in meiofauna phyla (Kinorhyncha, Nematoda, Platyhelminthes, Rotifera, Tardigrada, Xenacoelomorpha), although this difference was only significant for Platyhelminthes with 18S V1-V2 (paired t-tests, p = 0.02). Sieved and unsieved sediment

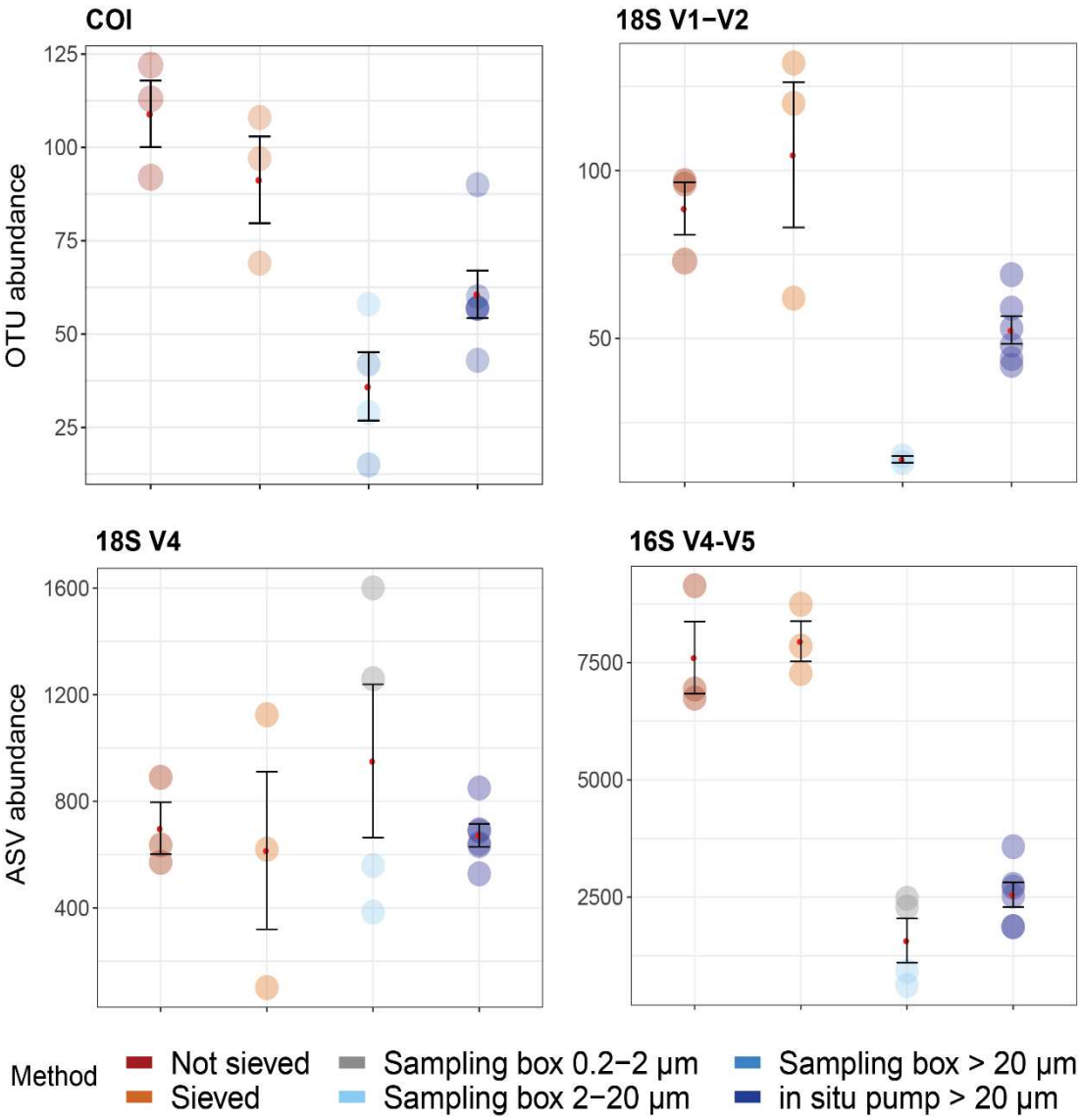


Figure 1. Numbers of metazoan OTUs (COI, 18S V1-V2,), unicellular eukaryote (18S V4) and prokaryote (16S V4-V5) ASVs recovered by deep-sea sediment (brown) and aboveground water (blue), with two sampling methods for each sample type. Sediment was either sieved through 5 mesh sizes to size-sort organisms prior DNA extraction, or DNA was extracted directly from crude sediment samples. Water was collected with a 7.5 L sampling box, allowing recovery of up to two size classes per taxonomic compartment, or sampled in large volumes with an in situ pump. Cluster abundances were calculated on rarefied datasets. Red dots indicate mean values. Bars represent standard errors.

detected comparable ASV numbers in most microbial groups, except the Actinobacteria, Cyanobacteria, Gammaproteobacteria, Nanoarchaeaeota (Fig. S2, paired t-tests, $p = 0.02$-0.04).

The water sampling box and the *in situ* pump recovered similar total OTU/ASV numbers for metazoans (COI), unicellular eukaryotes (18S V4), and prokaryotes (16S V4-V5, Table 1). However, considerable variation in detected cluster numbers was observed between size fractions of the sampling box, as underlined by the amplification failure of the smallest size fraction (0.2-2 µm) with primers targeting metazoans and the largest size fraction (>20 µm) with primers targeting microbial communities. Thus, and as expected, larger size fractions better detected metazoan taxa, while smaller size fractions better detected microbial taxa (Fig. 1). For metazoans resolved with 18S V1-V2, for which only the 2-20 µm size fraction from the sampling box samples was successfully sequenced, significantly fewer OTUs were detected with the sampling box compared to the *in situ* pump (Fig. 1, Table 1).

Water sampling methods strongly differed in terms of recovered alpha diversity depending on taxonomic compartment. The *in situ* pump generally detected more metazoan diversity than the sampling box (Fig. 2), and this difference was significant for Arthropoda, Rotifera (COI and 18S V1-V2), Annelida, Ctenophora, Mollusca, Nematoda, and Vertebrata (18S V1-V2, t-tests, $p = 0.0001$-0.04). For protists and prokaryotes, the *in situ* pump detected significantly more ASVs compared to the sampling box only in some taxonomic groups (i.e., Bacillariophyta, Phaeodarea, Acidobacteria, Bacteroidetes, Delta-, Gammaproteobacteria, Lentisphaerae, Omnitrophicaeota, Planctomycetes, and Patescibacteria, Fig. S2, t-tests, $p = 4.3\text{x}10^{-5}$-0.03). Other clades were significantly more diverse in the sampling box (e.g., Haptophyta, Telonemia, and Cyanobacteria, t-tests, $p = 0.001$-0.02). With the sampling box, the smallest size fraction (0.2-2 µm) allowed recovering more alpha diversity in all microbial groups than the larger size fraction (2-20 µm). This difference was significant only for Labyrinthulea and Chloroflexi (paired t-tests, $p = 0.02$-0.03), although non-significant comparisons may result from the limited sample sizes available in this comparison. The two size fractions available with the sampling box for COI (2-20 µm, > 20 µm) did not reveal differences in diversity recovery with size class, as most phyla were detected equally well in both (Fig. 2).

Table 1. Effect of sampling method on cluster richness and community structure for the 4 studied genes. ANOVAs were performed on models with quasipoisson distributions using on rarefied datasets. PERMANOVAs were calculated on rarefied datasets by permuting 1,000 times, using Jaccard distances for metazoans and Bray-Curtis distances for 18S V4 and 16S V4-V5. Significant *p* values are in bold. For pairwise comparisons significance codes are *p*<0.001: '***'; *p*<0.01: '**'; *p*<0.05: '*'.

LOCUS	Cluster richness			Community differentiation			
	F-value	***p-value***	**pairwise comparisons**		**R^2**	***p-value***	**pairwise comparisons**
COI							
Sample processing	10.6	**0.001**	Water < Not sieved ** Sampling box < Sieved *** Not sieved > Sieved Sampling box < in situ pump	Sample processing	0.42	**< 0.001**	Water / Not sieved * Water / Sieved * Not sieved / Sieved Sampling box / in situ pump **
18S V1-V2							
Sample processing	17.7	**< 0.001**	Water < Not sieved * Water < Sieved *** Not sieved < Sieved Sampling box < in situ pump **	Sample processing	0.47	**< 0.001**	Sampling box / Sediment in situ pump / Sediment * Not sieved / Sieved Sampling box / in situ pump *
18S V4							
Sample processing	0.6	0.60	Water > Sieved in situ pump ~ Not sieved Sampling box > Not sieved Not sieved ~ Sieved Sampling box > in situ pump	Sample processing	0.55	**< 0.001**	Water / Not sieved * Water / Sieved * Not sieved / Sieved Sampling box / in situ pump **
16S V4-V5							
Sample processing	35.7	**< 0.001**	Water < Not sieved *** Water < Sieved *** Not sieved ~ Sieved Sampling box < in situ pump	Sample processing	0.73	**< 0.001**	Water / Not sieved * Water / Sieved * Not sieved / Sieved Sampling box / in situ pump *

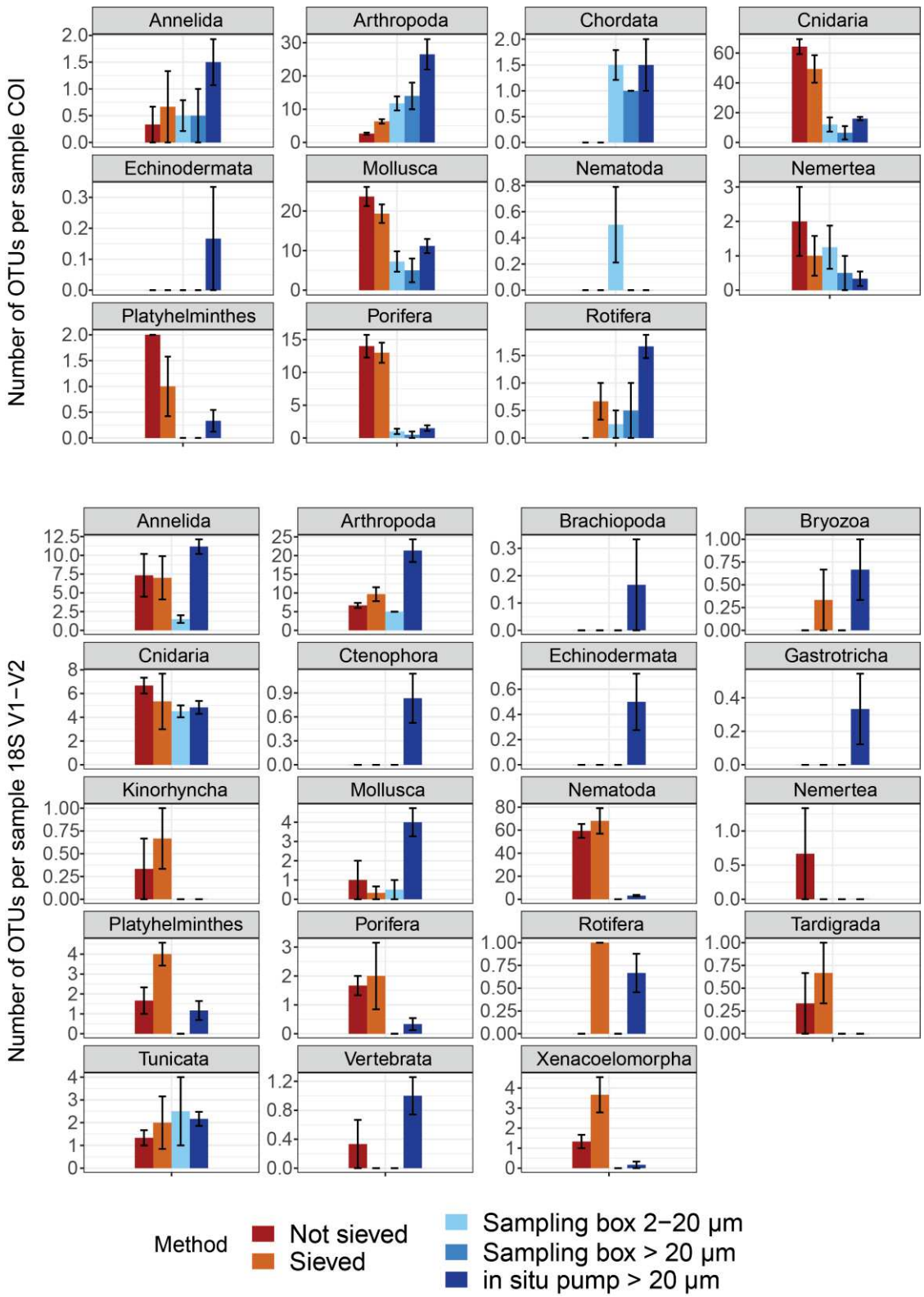


Figure 2. Mean numbers (±SE) of metazoan COI and 18S V1-V2 OTUs detected in target phyla for sediment (brown) and water (blue), using two sampling methods for both sample types. Sediment was either sieved to size-sort organisms prior DNA extraction, or DNA was extracted directly from crude sediment samples. Water was collected with a 7.5 L sampling box, allowing recovery of two size classes, or sampled in large volumes with an *in situ* pump. OTU numbers were calculated on rarefied datasets.

3.3 Effect of sampling method on community structures

Sediment and aboveground water samples detected significantly different communities for all investigated loci (Table 1), and pairwise PERMANOVAs showed that sample type (water or sediment) accounted for 45-54% (COI), 52-60% (18S V1-V2), 37-51% (18S V4), and 58-78% (16S) of variation in data. Relative taxonomic compositions revealed by aboveground water samples differed from sediment samples, with high proportions of arthropods, chordates, annelids, tunicates in the water samples, while nematodes, poriferans, platyhelminths, and xenacoelomorphs were predominant in the sediment samples (Fig. S3). Similarly, Dinophyceae, Haptophyta, Phaeodarea, Syndiniales, Alphaproteobacteria, and Cyanobacteria represented higher proportions of community structures in water than in sediment samples, while Ciliophora, Labyrinthulea, RAD-B, Acidobacteria, Chloroflexi, and Archeae were more abundant in sediment samples (Fig. S3).

Only 6% (COI), 10% (18S V1-V2), 9% (18S V4), and 5% (16S) of clusters were shared between sediment and water samples, and this resulted in strong segregation in PCoA ordinations (Fig. 3). For metazoans, these shared taxa were mostly hydrozoans (COI, 46%, 18S, 12%), calanoid and harpacticoid copepods (COI, 7%, 18S, 22%), gastropods (COI, 14%), demosponges (COI, 11%), or polychaetes (18S, 17%), and chromadorean nematodes (18S, 17%). For protists, ASVs shared among sediment and water samples primarily belonged to the Syndiniales (39%), but other taxa included dinophyceans (10%), filosans (9%), labyrinthuleans (5%), and bacillariophytes (6%). For prokaryotes, shared ASVs were predominantly belonging to the Proteobacteria (Gamma, 19%, Alpha, 10%, Delta, 8%), Bacteroidetes (15%), or Planctomycetes (16%).

Sediment processing did not significantly affect detected community structures as sieved and unsieved sediment resolved comparable communities (Table 1). However, sieving showed a higher impact on the characterization of microbial (eukaryotic or prokaryotic) communities, as indicated by the stronger segregation of sieved and unsieved sediment samples in PCoA ordinations (Fig. 3). Processing method accounted for 23% (COI and 18S V1-V2), 26% (18S V4), and 42% (16S) of variation among sediment samples, and 25% (COI), 28% (18S V1-V2), 9% (18S V4), and 22% (16S) of OTUs/ASVs were shared among sieved and unsieved sediment samples. Shared metazoan OTUs primarily belonged to Hydrozoa (18S, 2.7%, COI, 51%, Siphonophorae, Anthoathecata, Leptothecata), Demospongiae (COI, 14%), Gastropoda (COI, 20%), Nematoda (18S, 63% Chromadorea, 16% Enoplea), Polychaeta (18S, 4.5%), or Copepoda (18S, 5.4%). Microbial ASVs shared among sieved and unsieved sediment mostly

belonged to Syndiniales (18%), Filosa (16%), Ciliophora (11%), Dinophyceae (9%), Planctomycetes (22%), Acidobacteria (10%), or Proteobacteria (Gamma, 9%, Alpha, 8%, Delta, 11%).

In contrast, sampling method significantly affected detected community structure for water, as samples collected with the sampling box resulted in significantly different communities than those from the *in situ* pump (Table 1). Sampling method accounted for 26% (COI), 36% (18S V1-V2), 47% (18S V4), and 46% (16S) of variation among water samples. Only 9% (COI), 5% (18S V1-V2), 7% (18S V4), and 3% (16S) of ASVs/OTUs were shared between the *in situ* pump and the sampling box (Fig. 3). Taxonomic structures resolved by both sampling methods clearly changed due to targeted size fraction (Fig. S3). For metazoans with COI, the > 20 µm size fraction targeted by both the sampling box and the *in situ* pump displayed similar relative taxonomic compositions, while the sampling box's 2-20 µm size fraction resolved different community structures than the *in situ* pump for both metazoan markers. Similarly, both water-sampling methods never targeted the same size fraction for microbial data, resulting in different community structures. The *in situ* pump, targeting the > 20 µm size class, detected higher relative abundances of Bacillariophyta, Ciliophora, and Phaeodarea for protists, and higher relative abundances of Delta-, Gammaproteobacteria, Lentisphaerae, and Planctomycetes for prokaryotes. Both size fractions of the sampling box were characterised by increased relative abundances of cryptophytes, haptophytes, and telonemians (18S V4), as well as Alphaproteobacteria, Marinimicrobia, and Thaumarchaeota (16S).

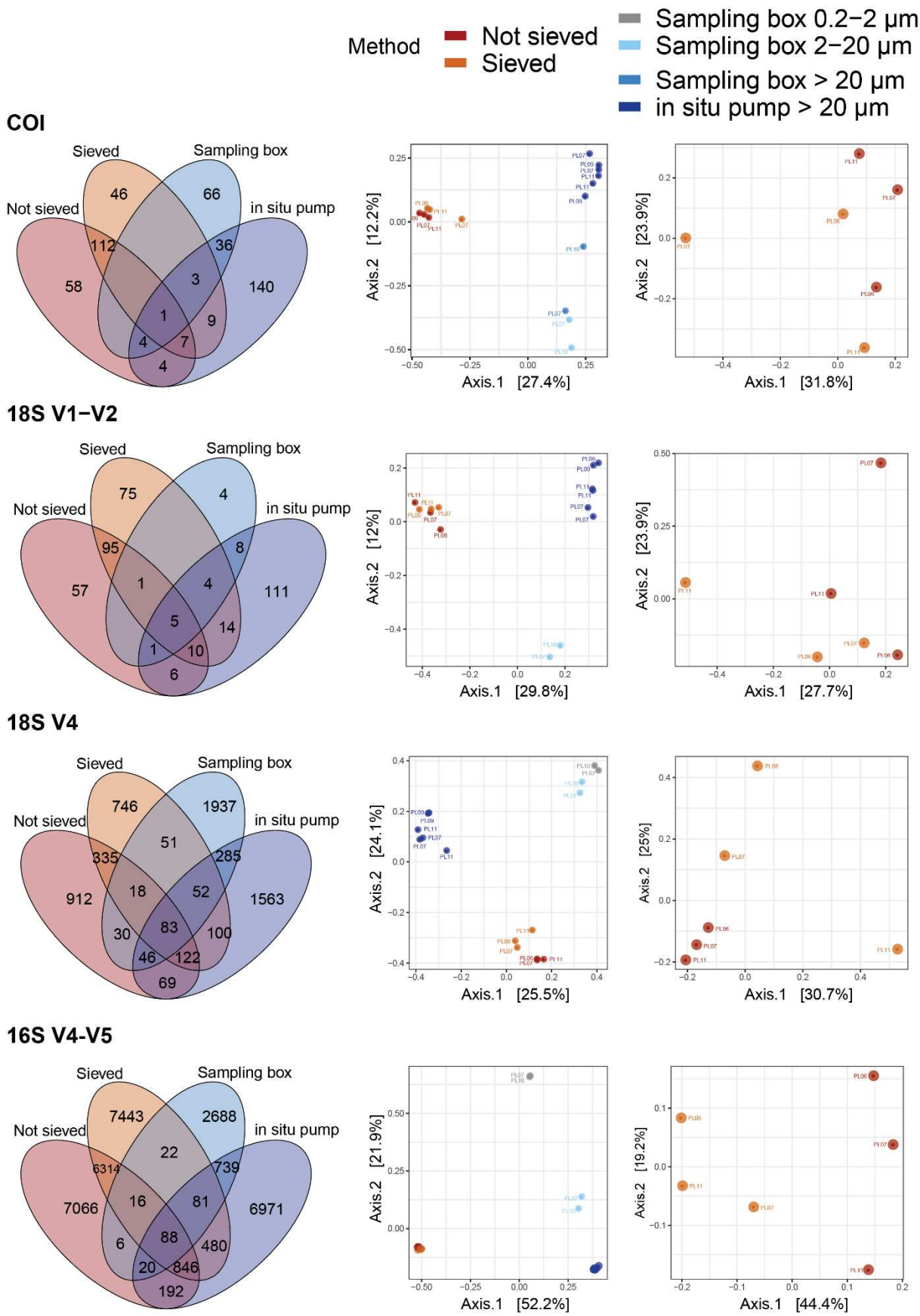


Figure 3. Venn diagrams (left) and Principal Coordinates Analyses (PCoA) ordinations showing differences in community compositions detected by deep-sea sediment (brown) and aboveground water (blue) for metazoans (COI and 18S V1-V2), micro-eukaryotes (18S V4), and prokaryotes (16S V4-V5). Community segregation is strongest between sample types, but also among target size class in the water samples. Sediment was either sieved to size-sort organisms prior DNA extraction, or DNA was extracted directly from crude sediment samples. Water was collected with a 7.5 L sampling box, allowing recovery of two size classes in each taxonomic compartment, or sampled in large volumes with an *in situ* pump.

4 Discussion

4.1 Importance of substrate nature

Sediment samples, whether sieved or unsieved, led to the detection of higher numbers of metazoan OTUs and prokaryote ASVs than water samples (Fig. 1), indicating that more diversity could be found in the benthos compared to the pelagos at this Mediterranean site for those groups. For unicellular eukaryotes, the difference in diversity between sediment and aboveground water was not significant. However, this may primarily be due to the fact that some benthic taxa (filosans, labyrinthuleans, ciliates) were also well detected by water samples (Fig. S2). Indeed, 22% of protist sediment ASVs were also detected in the water samples, while for other loci this percentage was closer to 10%. These findings are congruent with other studies in the marine realm that reported notably higher diversity in sediments compared to seawater (Forster et al. 2016; Probandt et al. 2017; Zinger et al. 2011) for microbial communities, and show that higher diversity can also be expected for metazoans.

Community compositions differed markedly between sediment and aboveground water samples for all life compartments investigated (Fig. 3), and only 5 to 10% of total molecular clusters were shared between substrate types, a range congruent with previous findings (Zinger et al. 2011; Zhao et al. 2020; Antich et al. 2020). Metazoan infauna taxa (e.g., nematodes, platyhelminths, kinorhynchs, tardigrades, and xenacoelomorphs) were specifically detected by sediment samples, while other epibenthic, benthopelagic, and pelagic metazoans were more prevalent in water samples (e.g., echinoderms, chordates, ctenophores). Similarly, with protists and prokaryotes, sediment samples detected lineages typically reported in the deep seafloor, with prokaryotic communities mostly comprised of Proteobacteria, Acidobacteria, Planctomycetes, Thaumarchaeota, Bacteroidetes, and Chloroflexi (Liao et al. 2011; Bienhold et al. 2016; Zhang, J. et al. 2015; Zhang, L. et al. 2016), and protist communities characterized by benthic heterotrophic groups such as ciliates, labyrinthuleans, and filosans (Zhao et al. 2017; Rodríguez-Martínez et al. 2020). Water samples instead recovered taxa commonly reported in pelagic studies, with microbial eukaryotes such as dinoflagellates (Dinophyceae, Syndiniales), radiolarians (Acantharea, Phaeodarea, Spumellarida), or MAST (e.g. diatoms, Chlorophyta, Chrysophyceae) (Pernice et al. 2015a; Massana et al. 2015; Zhao et al. 2020), and bacterial groups such as Proteobacteria, Bacteroidetes and Cyanobacteria (Salazar et al. 2016; Díez-Vives et al. 2019; Lochte and Turley 1988).

Most of the metazoans shared among sediment and water samples displayed benthopelagic life cycles with a benthic adult and a pelagic larvae (hydrozoans, gastropods, demosponges, polychaetes), confirming that the detection of benthic taxa in water samples may predominantly reflect the occurrence of dispersal phases of those organisms. Similarly, Bacteroidetes and Planctomycetes, bacteria that were predominant in this shared fraction are known to occur at the sediment-water interface (Stokke et al. 2015; Probandt et al. 2017). Finally, the fact that diatoms (Bacillariophytes), and dinoflagellates (Dinophyceae and parasitic Syndiniales) were abundant both in sediment and water samples supports the fact that some planktonic protists can sink to deep seafloor (Agusti et al. 2015).

Overall, our results confirm previous findings showing that sample nature strongly affects the type organisms targeted by eDNA metabarcoding (Koziol et al. 2019; Roussel et al. 2011), and underlines that eDNA from water samples cannot be used to comprehensively survey benthic communities (Hajibabaei et al. 2019; Antich et al. 2020; Gleason et al. 2020), even when large volumes of aboveground water are collected.

4.2 Sieving sediment is not essential for comprehensive benthic biodiversity surveys

Studies investigating the effect of size-sorting in macroinvertebrates showed that sorting organisms by size and pooling them proportionately according to their abundance led to a more equal amplification of taxa, the sorted samples recovering 30% more taxa than the unsorted samples at the same sequencing depth (Elbrecht et al. 2017). The size fractions used in this study were specifically aiming to separate the macrofauna (> 1 mm) from the meiofauna (32 µm - 1mm) compartment, which is known to be important in deep-sea sediments, both in terms of abundance and biomass (Thiel 1983; Rex et al. 2006; Zeppilli et al. 2018). Meiofauna taxa, best captured by 18S V1-V2, were more numerous in sieved than unsieved sediment samples, although this difference was only significant for Platyhelminthes (Fig. 2). It could be that the equimolar pooling performed with DNA extracts from each different size fraction maintained biases in abundance, as larger organisms contributed more DNA molecules within each size fraction. This would explain the non-significant differences observed between sieved and unsieved sediment for most metazoan phyla and total OTU numbers. Proportional pooling may be a better approach, but is feasible only if relative abundance of organisms in each size class can be calculated (e.g., using dry sample and specimen weights). A more accurate approach would be to sequence each size fraction separately; this however also increases five-fold

sequencing costs. However, the fact that more diversity was detected when sieving than when not sieving at the same sequencing depth for the 18S marker (Fig. S4), indicates sieving effectively reduces biomass biases, thus allowing detecting more diversity with the same sampling depth. Alternatively, new technologies affording much higher sequencing depths (Singer et al. 2019) might allow circumventing the need for size-class sorting in the future.

The advantage provided by sieving observed in this study for meiofauna may also result from the fact that five DNA extractions were performed for the sieved treatment (one for each size fraction), when only one was performed for non-sieved sediment. As number and type of DNA extraction are known to affect pro- and eukaryote taxon recovery (Webster et al. 2003; Cruaud et al. 2014; Nascimento et al. 2018), it remains to be tested whether several unsieved extractions would allow achieving similar detection levels.

Elutriation (i.e. resuspension of organisms and pouring of supernatant on a 32-µm sieve) or density extraction techniques are other methods traditionally used to study meiofauna (Brannock and Halanych 2015; Burgess 2001). These allow to process whole sediment layers more rapidly than sieving, and effectively concentrate metazoan organisms(Brannock and Halanych 2015). However, if the retention of organisms is achieved using only a single mesh size marking the lower size boundary of meiofauna, this also maintains size-abundance biases. Thus, whether sieving, elutriating, or density extracting, mesh sizes for size-class sorting have to be carefully chosen in order to reach the best compromise between processing time and biomass biases. As underlined by (Elbrecht et al. 2017), sorting is most useful when samples contain specimens with biomasses spanning several orders of magnitude. Given that deep-sea sediments contain large numbers of small organisms, and given the high detection capacity of metabarcoding, implementing five mesh sizes for sorting metazoans may be excessive. Instead, separating organisms into small, medium, and large size categories, as performed by (Elbrecht et al. 2017) for freshwater macroinvertebrates and by (Leray and Knowlton 2015) for coastal benthic communities may be sufficient to maximize metazoan species detection.

However, the rationale behind size sorting should be carefully considered when implementing an eDNA metabarcoding study on the deep seafloor. Indeed, for most integrative ecological studies, the proportion of abundant taxa is most relevant to reach accurate conclusions, and it may not be necessary to detect all small and rare taxa in such studies, at least not for metazoans. Moreover, effects of size sorting on other taxonomic compartments have to be taken into consideration. For microbial organisms, sieving down to a 20-µm mesh size is very likely to result in the loss of most small and/or free-living taxa. This idea is supported by

the fact that metazoan OTUs shared between sieved and unsieved sediment were mainly assigned to large taxonomic groups, indicating that small taxa predominantly explain the differences obtained between both methods. For protists and prokaryotes, although sieved and unsieved sediment uncovered comparable alpha diversity levels (Fig. 1), and resolved similar taxonomic compositions at phylum level (Fig. S3), ordinations indicated that communities segregated considerably with processing method (Fig. 3). Many sediment microorganisms are living within biofilms (e.g., Bacteroidetes, Archeae), attached to sediment particles (e.g, Planctomycetes) or as symbionts of larger taxa (e.g., Syndiniales, some Dinophyceae and Proteobacteria), making their retention on a 20-µm sieve possible. Our results support this idea, as microbial ASVs shared among sieved and unsieved sediment were mostly belonging to those groups or to taxa larger than 20 µm (e.g. ciliates), possibly explaining the non-significant difference we obtained in PERMANOVA (Table 1).

Finally, sieving is associated to higher contamination risks, as sieves need to be carefully washed between samples and water used for sieving (or elutriation) needs to be ultra-filtered (which can be problematic for the large volumes needed). Considering the limited improvement gained by sieving on metazoan communities, the logistic inconvenience, and the risk of bias for other taxonomic compartments, DNA extractions performed directly on 10 g of sediment appear as a satisfactory approach for large-scale biodiversity surveys targeting multiple life compartments.

4.3 Adjusting water sample volume and filter mesh size to target organisms

Numerous aquatic metabarcoding studies have highlighted that sampled water volume is a key factor affecting species detection rates with eDNA, and has to be adapted to the target ecosystem (Goldberg et al. 2016). Positive relationships between increased water volume and increased detection rate have been reported for macroinvertebrates and amphibians (Mächler et al. 2016; Lopes et al. 2017), and studies in freshwater ecosystems have shown that 20 litres to 30-68 litres of water are necessary to detect entire metazoan communities (Hänfling et al. 2016; Cantera et al. 2019; Evans, Nathan T. et al. 2017). While 1 L may be appropriate for macroinvertebrate detection in rivers (Mächler et al. 2016) or marine surface waters (Grey et al. 2018), the results presented here clearly show that 7.5 L of deep-sea water are not sufficient to accurately detect metazoan fauna. The sampling boxes detected less metazoan diversity than the *in situ* pump (Fig. 1), and failed to detect many phyla with 18S V1-V2 (Fig. 2). This reflects

the low abundance and biomass of large organisms in deep waters, combined with the very limited lifetime of extracellular DNA in seawater (Andruszkiewicz et al. 2017; Dejean et al. 2011; Collins et al. 2018; Sassoubre et al. 2016).

Water sampling methods for eDNA metabarcoding relying on on-board filtration or precipitation are intrinsically limited by the amount of water that can be processed. Although purpose-built sampling equipment has been developed for increased efficiency and standardization, filtration flow rates rarely exceed 1 L/min (Thomas et al. 2018). New developments allowing processing thousands of litres of water, such as the SALSA *in situ* pump presented here, or tow net methods developed for lentic ecosystems (Schabacker et al. 2020), improve the detection sensitivity for metazoan taxa in low biomass environments and will allow for more comprehensive and reliable surveys.

With protists and bacteria, taxonomic structures recovered by each sampling method clearly changed with targeted size class (Fig. S2, Fig. S3). Most protistan micro- to mesoplankton were better detected by the *in situ* pump (e.g., diatoms, phaeodareans, Acantharea, Ciliophora), while pico- to nanoplankton were preferentially targeted by the sampling box (e.g., Haptophyta, Telonemia), with many groups mostly by the smallest size fraction (0.2-2 µm, Chlorophyta, Labyrinthulea, Chrysophyceae, MAST). For bacteria, groups known to occur in aggregates, on larger particles, or in association with larger organisms were better recovered by the *in situ* pump (e.g., Actinobacteria, Bacteroidetes, Delta-, Gammaproteobacteria, Lentisphaerae), while other, likely free-living, bacterioplankton were predominant in the sampling box samples (Cyanobacteria, Marinimicrobia). This differential taxon recovery of water collection methods has already been reported in shallower studies (Massana et al. 2015), and highlights the importance of targeting the 0.2-2 µm for accurately surveying microbial diversity.

Although the SALSA prototype presented here has since been improved to pump through a 5-µm nylon mesh, i*n situ* filtration techniques are inherently limited by mesh size in order to filter large volumes of water. Thus, although targeting large volumes such as the ones allowed by SALSA represents the most suitable strategy for assessing metazoan diversity in deep-sea waters, its limitation in terms of mesh size leads to the detection of only a fraction of microbial diversity, i.e. mostly larger planktonic groups or taxa fixed on larger faunal specimens or mineral particles. On board filtration of smaller volumes of water remains necessary to access the pico- and nanoplankton, highlighting that both sampling methods are complementary and should be deployed in parallel for integrative biodiversity surveys across the tree of life.

Overall, this comparative study helps advancing towards more comprehensive and more reliable assessments of metazoan and microbial deep-sea communities based on eDNA metabarcoding. First, only sediment samples can allow the characterization of benthic taxa and aboveground water samples do not provide a good alternative. Second, sieving sediment leads to an improvement of taxa detection for metazoans, but as expected, also modifies the retrieved community composition for protists and prokaryotes. Thus, for studies targeting only metazoans, it is advisable to first separate the organisms from the sediment particles using sieving, elutriation, or density extraction techniques as recommended by Brannock & Halanych (Brannock and Halanych 2015). If both metazoan and microbial communities are targeted, and provided sample volume is large enough, an ideal sampling design would be to use multiple sub-samples for microbial taxa and size-sort the remaining sediment for detecting metazoans, as suggested by Nascimento et al. (Nascimento et al. 2018). Alternatively, as shown here, using sufficient volumes of unsorted sediment seems to be satisfactory for integrative studies across taxonomic compartments. Finally, water sample volume and mesh size need to be carefully chosen depending on taxa of interest, and while volumes collected by sampling boxes (or Niskin bottles) allow surveying microbial diversity, much larger volumes are needed to detect deep-sea metazoans.

Acknowledgements

This work is part of the "*Pourquoi Pas les Abysses?*" project funded by Ifremer, and the project eDNAbyss (AP2016 -228) funded by France Génomique (ANR-10-INBS-09) and Genoscope-CEA. This work also received funding from the 'Investissements d'Avenir' program OCEANOMICS (ANR-11-BTBR-0008; NH). We wish to thank Laure Quintric and Patrick Durand for bioinformatic support and Stéphane Pesant for help in data management. We also wish to express our gratitude to the crew of the R/V Atalante, to the Nautile crew and operators, and to the participants of the EssNaut16 cruise.

Author contributions

MB, DZ, and SA-H designed the study, FP, MAC-B, VC-G carried out the sampling, MB, JP, and CL-H carried out the laboratory work. MB, BT, and CB performed the bioinformatic analyses. MB, BT, and NH performed the statistical analyses. MB, DZ, and SAH wrote the manuscript. All authors contributed to the final manuscript.

Additional Information

Data accessibility: The raw data for this work can be accessed in the European Nucleotide Archive database (Study accession numbers: PRJEB37673 for water, PRJEB33873 for sediment). Please refer to the metadata excel file for ENA file names. The dataset, including raw sequences, databases, as well as raw and refined ASV/OTU tables are available on *https://doi.org/10.12770/2deb785a-74c5-4b9d-84d6-82a81e0dda6d*. Bioinformatic scripts can be accessed following the Gitlab link.

Conflicts of interest: The authors declare no competing financial interests.

Chapter V.
Multi-marker eDNA metabarcoding reveals large and small-scale metazoan biodiversity patterns in the deep Atlantic-Mediterranean transition zone

Manuscript in preparation:

Brandt MI, Zeppilli D, Guenther B, Trouche B, Henry N, Simier M, Orejas C, Pradillon F, Fabri MC, Cambon-Bonavita MA, Guieu C, Tamburini C, Quintric L, Belser C, Poulain J, Wincker P, Maignien L, de Vargas C, and Arnaud-Haond S. Multi-marker eDNA metabarcoding reveals large and small-scale metazoan biodiversity patterns in the deep Atlantic-Mediterranean transition zone.

Morphology-based studies, although essential for species descriptions, are limited in terms of large-scale ecological applications. Environmental DNA metabarcdoing approaches represent useful tools for increasing the spatial scale of deep-sea studies, while allowing to target biodiversity of various biological compartments in parallel, including the commonly overlooked meio- and nanofauna. This chapter demonstrates the application of the optimized eDNA metabarcoding protocols developped in previous chapters, on deep seafloor of the Atlantic-Mediterranean transition zone. The influence of local abiotic factors on deep-sea benthic metazoan OTU richness and community structure are evaluated at the local, habitat, and regional scales, along this west east transect ranging from the Western North Atlantic to the Ionian Sea.

Questions addressed:

a. How do abiotic factors such as sediment layer, sediment grain size, and organic matter content affect metazoan biodiversity patterns across regional scales?

b. How do regional and habitat differences explain deep-sea benthic biodiversity patterns along the Mediterranean-Atlantic transition zone?

c. Does the Gibraltar straight constitute a connectivity barrier between the two ocean basins?

Résumé en français

Le habitats sédimentaires abyssaux couvrent plus de 50% de la planète Terre et sont un grand réservoir de biodiversité encore largement non décrite, bien que de plus en plus sous pression d'activités anthropiques. Dans de tels écosystèmes vastes et difficiles d'accès, le pouvoir de détection élevé du métabarcoding d' ADN environnemental (ADNe), sur des échantillons plus faciles à recueillir que les collections de specimens morphologiques, offre de nouvelles perspectives pour l'investigation standardisée de biodiversité et biogéographie à grande échelle.

En combinant le marqueur génetique mitochondrial COI et la région V1-V2 de l'ARN ribosomal 18S, nous avons étudié la biodiversité métazoaire à petite et à grande échelle dans la zone de transition Atlantique-Méditerranée, à l'aide d'ADNe extrait de sédiments profonds provenant de 13 sites allant de la Méditerranée centrale à la dorsale médio-atlantique. Nous avons évalué l'influence de la couche de sédiments, de la taille des grains de sédiments, de la teneur en matière organique ainsi que des communautés microbiennes (18S V9, 16S V4-V5), sur l'étendue et la structure de la biodiversité métazoaire dans cette région.

Nos résultats soulignent que les facteurs à petite échelle (centimètres) affectent fortement la richesse métazoaire des grands fonds marins et la composition des communautés. Une diminution significative de la richesse en unités taxonomiques moléculaires (OTU) fut observée avec chaque couche de sédiments, de 1 cm à 15 cm de profondeur, et une ségrégation verticale importante dans la structure des communautés a été révélée dans toutes les régions pour la méiofaune et la macrofaune. Les premiers cinq centimètres de sédiment abritaient la plupart des métazoaires (94% pour 18S, 98% pour COI), avec des nombres d'OTU allant de 2 à 168 par échantillon pour 18S et de 81 à 1259 pour COI.

Les facteurs à grande échelle (> 100 km) ont davantage affecté la diversité bêta que la diversité alpha. Le contenu de matière organique et la taille des grains de sédiments ont montré une forte variation à l'échelle régionale, avec une teneur en matière organique plus élevée dans les sédiments méditerranéens et des particules de plus grande taille dans l'Atlantique. Ces deux variables environnementales contribuèrent de manière significative à expliquer les différences de composition des communautés entre sites. La méio et la macrofaune ont révélés un fort niveau de correlation (RV = 0.87), confirmant des interactions trophiques fortes entre ces deux compartiments taxonomiques. De même, les compartiments protistes et procaryotes étaient corrélés à un niveau similaire (RV = 0.84), suggérant que les interactions trophiques sont plus marquées entre organismes de classes de taille comparables. Enfin, le détroit de Gibraltar fut

un facteur supplémentaire expliquant les très fortes différences régionales dans la composition des communautés, soutenant une influence combinée de facteurs historiques et de mouvements actuels des masses d'eau sur la distribution de la diversité benthique.

Abstract

The abyssal sedimentary seafloor covers more than 50% of planet Earth and is a large reservoir of still mostly undescribed biodiversity, although being increasingly under target of resource-extraction industries. In such remote and vast ecosystems, the high detection power of environmental DNA (eDNA) metabarcoding on samples easier to gather than specimen collections, offers new perspectives for the standardized investigation of large-scale biodiversity and biogeography patterns.

Using both mitochondrial COI and the V1-V2 region of 18S ribosomal RNA (rRNA), we investigated small-scale and large-scale metazoan biodiversity patterns in the Atlantic-Mediterranean transition zone, using eDNA extracted from deep-sea sediments of 13 sites spanning from the Central Mediterranean to the Mid Atlantic Ridge. We evaluated the influence of sediment layer, sediment grain size, organic matter content, as well as microbial communities (18S V9 for protists, 16S V4-V5 for prokaryotes), on the extent and structure of metazoan biodiversity in this region.

Our results highlight that small-scale (centimetres) factors strongly affect deep-sea metazoan richness and community composition. A significant decrease in OTU richness was observed with sediment layer, from 1 cm down to 15 cm within the sediment, and significant vertical segregation in community structure was revealed in all regions for both meiofauna and macrofauna. The upper five centimetres harboured most metazoan OTUs (94% for 18S, 98% for COI), with numbers ranging from 2-168 per sample for 18S and 81-1,259 for COI.

Expectedly, large-scale factors (>100 km) affected beta-diversity more than alpha diversity. Organic matter composition and sediment grain size were found to vary strongly at regional scales, with higher organic matter content in Mediterranean sediments and larger particle sizes in the Atlantic. Both significantly contributed to explain differences in community composition among sites. A strong correlation was observed between the meio- and the macrofauna (RV = 0.87), confirming strong trophic interactions between these taxonomic compartments. A similar level of correlation (RV = 0.84) was also observed between protists and prokaryotes, suggesting that trophic interctions are strongest among organisms of similar size classes. Finally, the Gibraltar Strait was an additional factor explaining the very strong regional differences in community compositions, supporting a combined influence of past biogeography and present day movements of water masses on the distribution of benthic diversity.

1 Introduction

While ocean exploration is relatively recent, studies have started shedding light on biodiversity and biogeography patterns in the deep-sea realm during the last decades (Rex 1981; Grassle 1989; Grassle and Maciolek 1992; Vanreusel et al. 2010a; Ramirez-Llodra et al. 2010; Danovaro et al. 2010; Woolley et al. 2016). However, these studies were confronted with the extraordinary vastness of deep-sea ecosystems, the difficulty of sampling in these remote and high-pressure locations, as well as the high costs and time involved in collecting and analysing samples (Danovaro et al. 2014). Analytical methods based on extrapolation from known samples have clearly indicated that deep-sea life is much more diverse than prefiously thought, although estimates remain highly uncertain, primarily due to under-sampling and to the difficulty of identifying specimens. The large marine databases assembled in recent years include too little information about deep-sea species to allow reasonable extrapolation for the estimation of deep-sea biodiversity (Ramirez-Llodra et al. 2010). Studies also highlighted the strong link between surface and deep-ocean regions, showing that benthic deep-sea communities are affected by climate-driven variations in carbon cycles and can therefore directly influence carbon remineralisation and sequestration processes (Smith, K. L. et al. 2009; 2013). However, monitoring these surface-driven changes in deep-sea benthic communities is costly and difficult to sustain over long-term periods.

Deep-sea sedimentary habitats cover more than 50% of the Earth's surface, can host high numbers of organisms (50,000-5 million individuals per square meter), which perform key ecosystem roles such as nutrient cycling, sediment stabilisation and transport, or secondary production (Bik et al. 2012b; Fonseca, V. G. et al. 2010; Snelgrove 1999). The deep seafloor is also characterised by high local and regional diversity (Grassle and Maciolek 1992; Smith, C. R. and Snelgrove 2002; Rosli et al. 2017). Yet, whether this holds true on a global scale is still under debate (Costello and Chaudhary 2017), partly due to the difficulty to integrate local or regional studies made by different taxonomic experts and teams based on distinct protocols (Vanreusel et al. 2010a). Despite this, they are under increased threat from a variety of ongoing or forecasted human activities, ranging from climate change-induced indirect threats due to modifications in ocean biogeochemistry to direct threats from activities such as waste disposal, pollution, or resource exploitation (Ramirez-Llodra et al. 2010; Smith, C. R. et al. 2008; Ramirez-Llodra et al. 2011). Better knowledge of deep-sea biodiversity patterns and the development of large-scale deep-sea biomonitoring protocols are therefore becoming necessary in order to preserve this vast and elusive backyard.

Environmental DNA metabarcoding approaches represent a new perspective for obtaining large-scale inventories of biodiversity and infer ecological and biogeographic drivers of life in deep-sea sediments, to bridge this knowledge gap. They have revolutionized biodiversity research in the past decade and have already been successfully applied in marine sedimentary habitats (Pawlowski et al. 2011; Bik et al. 2012b; Fonseca, V. G. et al. 2010; 2014; Cowart et al. 2015; Sinniger et al. 2016; Forster et al. 2016; Cordier et al. 2019a). Using high-throughput sequencing (HTS) and bioinformatics, these methods allow the detection or the inventory of target organisms using their DNA directly extracted from soil, water, or air samples (Taberlet et al. 2012a). As they do not require specimen isolation, they are practical and efficient tools in large and hard-to-access ecosystems, such as the deep-sea realm. Besides allowing the study of various biological compartments simultaneously, eDNA metabarcoding is also very effective for detecting the diversity of small organisms (micro-organisms, meiofauna), very abundant in deep-sea sediments, but largely disregarded in visual biodiversity inventories due to the difficulty of their identification based on morphological features (Carugati et al. 2015). Finally, given the increased time-efficiency and above all standardization offered by this technique, eDNA metabarcoding also allows increasing the spatial scale of deep-sea studies.

Here, we apply eDNA metabarcoding on deep-sea sediments to investigate small-scale and large-scale metazoan biodiversity patterns in the Atlantic-Mediterranean transition zone. Using both mitochondrial COI and the V1-V2 region of 18S ribosomal RNA (rRNA), our aims were to 1) assess the extent of metazoan biodiversity and its distribution in the Atlantic-Mediterranean transition region, 2) evaluate the influence of current environmental conditions *vs* spatial, i.e. historical effects on metazoan community structure, and 3) evaluate the level of correlation between metazoan and microbial communities, resulting from direct or indirect biotic interactions.

2 Materials and methods

2.1 Preparation of samples

Environmental DNA

Sediment cores were collected from thirteen deep-sea sites located along a west-east gradient in the Mediterranean-Atlantic transition zone (Fig. 1, Table S1). Triplicate tube cores were collected with a multicorer or with a remotely operated vehicle at each sampling site, except for ESN-300m where only a blade corer was available. Each sediment core was sliced into five depth layers down to 15 cm (0-1 cm, 1-3 cm, 3-5 cm, 5-10 cm, 10-15 cm). The latter were transferred into zip-lock bags, homogenised, and frozen at −80°C on board before being shipped on dry ice to the laboratory. In each sampling series, an empty bag was kept as a field control processed through DNA extraction and sequencing.

Processing areas were cleaned with bleach, rinsed with MilliQ water, and dried with 70% ethanol. During all procedures, filter pipet tips and clean gloves were used, by wearing two pairs of gloves, allowing to easily and regularly changing the upper pair. DNA extractions were performed using ~10 g of sediment with the PowerMax Soil DNA Isolation Kit (Qiagen, Hilden, Germany). To increase the DNA yield, the elution buffer was left on the spin filter membrane for 5-10 min at room temperature before centrifugation. For field controls, the first solution of the kit was poured into the control zip lock, before following the usual extraction steps. DNA extracts were stored at -80°C.

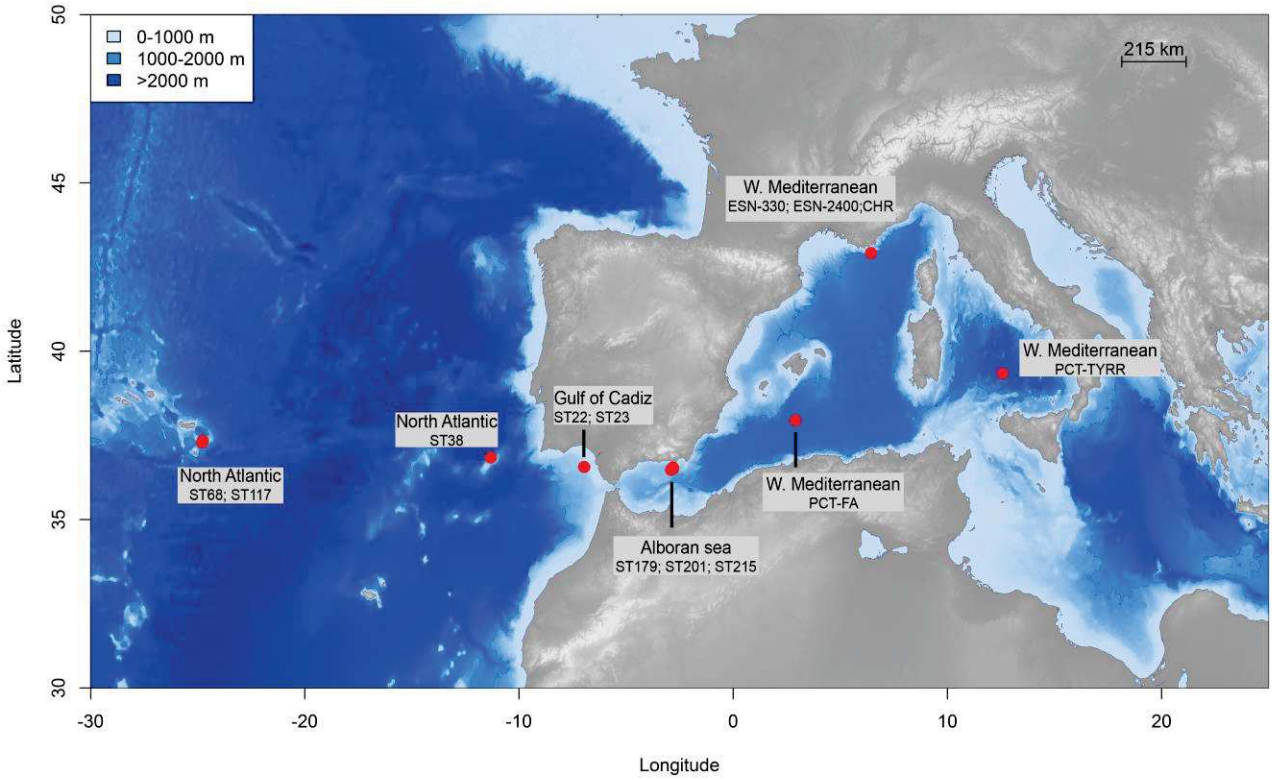


Figure 1. Map of sampling locations. Thirteen sites were sampled in four regions (North Atlantic, Gulf of Cadiz, Alboran Sea, and Western Mediterranean) along a west-east transect covering the Atlantic-Mediterranean transition zone.

Mock samples

Two metazoan mock communities (5 ng/µL) were used as positive controls throughout the PCR and sequencing processes. They were prepared using standardized 10 ng/µL DNA extracts of ten deep-sea specimens belonging to five taxonomic groups (Polychaeta, Crustacea, Anthozoa, Bivalvia, Gastropoda; Table S2). Specimen DNA was extracted using a CTAB extraction protocol, from muscle tissue or from whole polyps for cnidarians. The mock communities differed in terms of ratios of total genomic DNA from each species, with increased dominance of three species and secondary species DNA input decreasing from 3% to 0.7% (Table S2). We individually barcoded the species present in the mock communities: PCRs of the COI and 18S V1-V2 target genes were performed using the same primers as the ones used in metabarcoding (see below). The PCR reactions (25 µL final volume) contained 2 µL DNA template with 0.5 µM concentration of each primer, 1X *Phusion* Master Mix, and an additional 1 mM $MgCl_2$ for COI. PCR products (98°C for 30 s; 40 cycles of 10 s at 98°C, 45 s at 48°C (COI) or 57°C (18S), 30 s at 72°C; and 72°C for 5 min) were cleaned up with ExoSAP (Thermo

Fisher Scientific, Waltham, MA, USA) and sent to Eurofins (Eurofins Scientific, Luxembourg) for Sanger sequencing. The barcode sequences obtained for all mock specimens were added to the databases used for taxonomic assignments of metazoan datasets, and were submitted on Genbank under accession numbers MN826120-MN826130 and MN844176-MN844185.

Organic matter content and sediment grain size

Organic matter (OM) content and the distribution of particle size distributions were measured for each sample (FILAB, Dijon, France). For OM content, ~2 g of sediment were dried by heating them at 100°C overnight. Their percent content of OM was determined by their loss on ignition, the dried samples being decarbonised by heating at 550°C for four hours. Liquid dispersion laser diffraction was performed on each sample for particle size analysis, taking a minimum of four measures per sample.

2.2 PCR amplification and sequencing

Four primer pairs were used to amplify one mitochondrial and three rRNA barcode loci targeting metazoans (COI, 18S V1-V2), micro-eukaryotes (18S V9) and prokaryotes (16S V4-V5). PCR amplifications, library preparation, and sequencing were carried out at Genoscope (Evry, France) as part of the eDNAbyss project.

Eukaryotic 18S V1-V2 rRNA gene amplicon generation

Amplifications were performed with the *Phusion* High Fidelity PCR Master Mix with GC buffer (Thermo Fisher Scientific, Waltham, MA, USA) and the SSUF04 (5'-GCTTGTCTCAAAGATTAAGCC-3') and SSUR22*mod* (5'-CCTGCTGCCTTCCTTRGA-3') primers (Sinniger et al. 2016), preferentially targeting metazoans, the primary focus of this study. The PCR reactions (25 µL final volume) contained 2.5 ng or less of DNA template with 0.4 µM concentration of each primer, 3% of DMSO, and 1X *Phusion* Master Mix. Triplicate PCR amplifications (98 °C for 30 s; 25 cycles of 10 s at 98 °C, 30 s at 45 °C, 30 s at 72 °C; and 72 °C for 10 min) were carried out in order to smooth the intra-sample variance while obtaining sufficient amounts of amplicons for Illumina sequencing.

Eukaryotic COI gene amplicon generation

Metazoan COI barcodes were generated using the mlCOIintF (5'-GGWACWGGWTGAACWGTWTAYCCYCC-3') and jgHCO2198 (5'-TAIACYTCIGGRTGICCRAARAAYCA-3') primers (Leray et al. 2013). Triplicate PCR reactions (20 µl final volume) contained 2.5 ng or less of total DNA template with 0.5 µM final concentration of each primer, 3% of DMSO, 0.175 mM final concentration of dNTPs, and 1X Advantage 2 Polymerase Mix (Takara Bio, Kusatsu, Japan). Cycling conditions included a 10 min denaturation step followed by 16 cycles of 95 °C for 10 s, 30s at 62°C (−1°C per cycle), 68 °C for 60 s, followed by 15 cycles of 95 °C for 10 s, 30s at 46°C, 68 °C for 60 s and a final extension of 68 °C for 7 min.

Eukaryotic 18S V9 rRNA gene amplicon generation

Unicellular eukaryote barcodes were generated using the 1389F 5′-TTGTACACACCGCCC-3′ and 1510R 5′- CCTTCYGCAGGTTCACCTAC-3′ (Amaral-Zettler et al. 2009). Triplicate PCR mixtures were prepared as described above for 18S V1-V2 amplification, but cycling conditions included a 30 s denaturation step followed by 25 cycles of 98 °C for 10 s, 57 °C for 30 s, 72 °C for 30 s, and a final extension of 72 °C for 10 min.

Prokaryotic 16S V4-V5 rRNA gene amplicon generation

Prokaryotic barcodes were generated using 515F-Y (5'-GTGYCAGCMGCCGCGGTAA-3') and 926R (5'-CCGYCAATTYMTTTRAGTTT-3') 16S-V4V5 primers (Parada et al. 2016). Triplicate PCR reactions were prepared as described above for 18S V1-V2, but cycling conditions included a 30 s denaturation step followed by 25 cycles of 98 °C for 10 s, 53 °C for 30 s, 72 °C for 30 s, and a final extension of 72 °C for 10 min.

Amplicon library preparation

PCR triplicates were pooled and PCR products purified using 1X (1.8X for 18S V9) AMPure XP beads (Beckman Coulter, Brea, CA, USA) clean up. Aliquots of purified amplicons were run on an Agilent Bioanalyzer using the DNA High Sensitivity LabChip kit (Agilent Technologies, Santa Clara, CA, USA) to check their lengths, and quantified with a Qubit fluorimeter (Invitrogen, Carlsbad, CA, USA). One hundred nanograms of pooled amplicon triplicates were directly end-repaired, A-tailed and ligated to Illumina adapters on a Biomek FX Laboratory Automation Workstation (Beckman Coulter, Brea, CA, USA). Library

amplification was performed using a Kapa Hifi HotStart NGS Library Amplification kit (Kapa Biosystems, Wilmington, MA, USA) with the same cycling conditions applied for all metagenomic libraries and purified using 1X AMPure XP beads.

Sequencing library quality control

Amplicon libraries were quantified by Quant-iT dsDNA HS assay kits using a Fluoroskan Ascent microplate fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and then by qPCR with the KAPA Library Quantification Kit for Illumina Libraries (Kapa Biosystems, Wilmington, MA, USA) on an MxPro instrument (Agilent Technologies, Santa Clara, CA, USA). Library profiles were assessed using a high-throughput microfluidic capillary electrophoresis system (LabChip GX, Perkin Elmer, Waltham, MA, USA).

Sequencing procedure

Library concentrations were normalized to 10 nM by addition of 10 mM Tris-Cl (pH 8.5) and applied to cluster generation according to the Illumina Cbot User Guide (Part # 15006165). Amplicon libraries are characterized by low diversity sequences at the beginning of the reads due to the presence of the primer sequence. Low-diversity libraries can interfere in correct cluster identification, resulting in a drastic loss of data output. Therefore, loading concentrations of libraries were decreased (8–9 pM instead of 12–14 pM for standard libraries) and PhiX DNA spike-in was increased (20% instead of 1%) in order to minimize the impacts on the run quality. Libraries were sequenced on HiSeq4000 (System User Guide Part #15011190) instruments in a 150 bp paired-end mode for 18S V9, and on HiSeq2500 (System User Guide Part #15035786) instruments (Illumina, San Diego, CA, USA) in a 250 bp paired-end mode for all other amplicons.

2.3 Bioinformatic analyses

All bioinformatic analyses were performed using a Unix shell script available on Gitlab (https://gitlab.ifremer.fr/abyss-project/), on a home-based cluster (DATARMOR, Ifremer). The details of the pipeline, along with specific parameters used for all markers, are given in Table S3 and in Brandt et al. (2020). Pairs of Illumina reads were corrected with DADA2 v.1.10 (Callahan et al. 2016) following the online tutorial for paired-end HiSeq data (https://benjjneb.github.io/dada2/bigdata_paired.html).

Prokaryote and unicellular eukaryote diversity was evaluated with Amplicon Sequence Variants (ASVs), while metazoan data was further clustered into OTUs with FROGS (Escudié et al. 2018) using swarm v2 at *d=1* for 18S V1-V2 and *d=6* for COI (Mahé et al. 2015). ASVs and OTUs were taxonomically assigned via BLAST+ (v2.6.0) based on minimum similarity and minimum coverage (-perc_identity 70 and –qcov_hsp 80). The Silva132 reference database was used for taxonomic assignment of the 18S V1-V2 and 16S rRNA marker genes (Quast et al. 2012), MIDORI-UNIQUE (Machida et al. 2017) subsampled to marine taxa only was used for COI, while Silva132 and PR2 (Guillou et al. 2013) were used for 18S V9.

Molecular inventories were refined in R v.3.5.1 (R Core Team 2018). A blank correction was made using the *decontam* package v.1.2.1 (Davis et al. 2018), removing all clusters that were more prevalent in negative control samples than in true or mock samples. Clusters unassigned at phylum-level and with non-target assignments were removed. For 18S V9, clusters assigned to prokaryotes with Silva132 were removed. For metazoan loci, all clusters with a terrestrial assignment (groups known to be terrestrial-only) were removed. Samples with less than 10,000 target reads were discarded. We then performed an abundance renormalization to remove spurious ASVs/OTUs due to random tag switching (Wangensteen and Turon 2016). The metazoan OTU tables were further curated with LULU v.0.1 (Frøslev et al. 2017) to filter out spurious OTUs originating from intraspecific variation and/or pseudogenes, using a minimum co-occurrence of 0.90, and a minimum match threshold of 84% for COI and 90% for 18S. Finally, we taxonomically filtered the data to ensure taxonomic reliability at phylum-level: only clusters with minimum hit identity of 86% for rDNA loci and 80% for COI were retained. These values were chosen as they represent approximate minimum identity for reliable phylum assignment (Stefanni et al. 2018; Yarza et al. 2014).

2.4 Statistical analyses

Metazoan OTU tables were analysed using R with the packages phyloseq v1.22.3 (McMurdie and Holmes 2013), following guidelines in online tutorials (http://joey711.github.io/phyloseq/tutorials-index.html), and vegan v2.5.2 (Oksanen et al. 2018). To avoid redundancy in the taxonomic resolution of 18S and COI, we subsampled each dataset to its target taxa according to the number of OTUs detected for each phylum. (Fig. 2): data for each phyla was thus extracted from either 18S or COI data depending on the richness each marker allowed revealing. With 18S, we kept OTUs assigned to the phyla Ctenophora,

Gastrotricha, Gnathostomulida, Hemichordata, Kinorhyncha, Loricifera, Nematoda, Nematomorpha, Orthonectida, Platyhelminthes, Tardigrada, Tunicata (in Chordata), and Xenacoelomorpha. Similarly, for COI, we subsampled the data to the following phyla: Annelida, Arthropoda, Brachiopoda, Bryozoa, Chaetognatha, Chordata (mostly Vertebrates), Cnidaria, Echinodermata, Entoprocta, Mollusca, Nemertea, Placozoa, Porifera, Priapulida, and Rotifera.

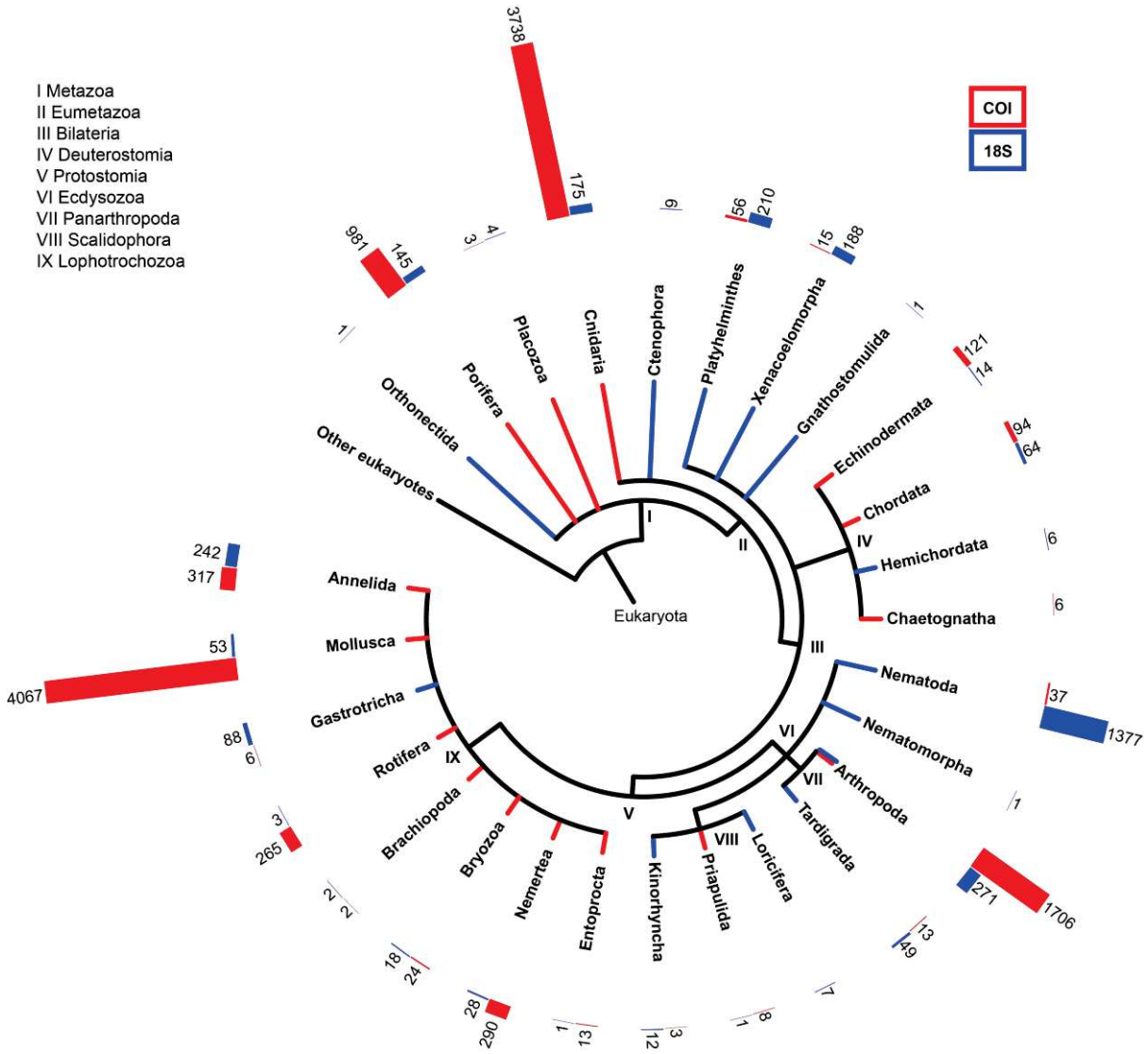


Figure 2. Numbers of Operational Taxonomic Units (OTUs) detected per metazoan phylum by each marker used in this study, after all filtering steps.

Read and rarefied cluster abundances among sediment horizons, regions, and sites were compared via analyses of variance (ANOVA) on mixed models with site as a random factor, using normal or poisson distributions for reads and clusters respectively. Significance was

evaluated with Wald Chi-square and likelihood ratio tests. Pairwise post-hoc comparisons were performed via Tukey HSD tests using the *emmeans* package. Numbers of shared OTUs among sediment horizons and regions in rarefied datasets were visualised with upset plots using the UpSetR package (Conway et al. 2017). Correlation between environmental variables (organic matter content and sediment grain size) and depth in the sediment or OTU richness was measured with the *cor.test* function (Pearson's product–moment correlation).

Homogeneity of multivariate dispersions were evaluated with the *betadisper* function of the *betapart* package v.1.5.1 (Baselga and Orme 2012), and region and site effects were evaluated on balanced datasets, as dispersions were not homogenous among regions. Permutational multivariate analysis of variance (PERMANOVA) was performed on incidence data of rarefied datasets to evaluate the effect of sediment horizon, region, and site on community compositions, using the function *adonis2* (vegan) with Jaccard dissimilarities. The rationale behind this choice is that metazoans are multicellular organisms of extremely varying numbers of cells, organelles, or ribosomal repeats in their genomes, and can also be detected through a diversity of remains. The number of reads can thus not be expected to reflect the abundance of detected OTUs. Significance was evaluated using 1,000 permutations with region as a blocking factor, and site as a plot factor (for evaluating region effect). Pairwise post-hoc comparisons were performed via the *pairwiseAdonis* package, with region as a blocking factor.

Differences among samples for meio- and macrofauna phyla as well as for all metazoan phyla combined were visualized via Principal Coordinates Analyses (PcoA) and Canonical Analysis of Principal Coordinates (CAP) based on Jaccard dissimilarities (Anderson and Willis 2003). Finally, combined analysis of macro-, meiofauna, unicellular eukaryotes (18S V9), and prokaryotes (16S V4-V5) was performed via STATIS analysis (Lavit et al. 1994) in ade4 (Dray and Dufour 2007). Correlation among taxonomic compartments was evaluated through RV coefficients obtained. For these combined analyses, data were reduced to contain only molecular clusters occurring at 0.05% in at least one sample.

3 Results

3.1 High-throughput sequencing results

A number of 163 million 18S V1-V2 reads and 89 million COI reads were obtained from triplicate PCR replicates of 133 (18S) and 82 (COI) sediment samples, 2 mock communities,

12 extraction blanks, and 18-19 PCR negative controls (Table S4). Numbers of sediment samples were lower in the COI dataset as more amplification failures occurred, especially in the deeper horizon samples (5-10 cm and 10-15 cm). Numbers of raw reads varied significantly with sediment horizon, with a decrease in read yield in the first three layers for COI (Chisq = 168.5, $p < 0.001$). For 18S, raw read abundance also varied with sediment horizon, but this not consistently among regions (Chisq = 37.2, $p < 0.001$), and significant differences among sites were also observed (Chisq = 9.2, $p = 0.002$). Quality-filtering and chimera removal reduced read numbers to 83 million for 18S V1-V2 and 61 million for COI (Table S4). Individual sediment samples contained between 0.6% and 0.8% of total processed reads, compared to 0.02% - 0.03% for field and extraction blanks and 0.004% - 0.007% for PCR blanks.

After taxonomic refining, decontamination (Davis et al. 2018), abundance renormalisation (Wangensteen and Turon 2016), and LULU curation (Frøslev et al. 2017), metazoan datasets comprised 29 (18S) and 21.4 (COI) million reads. Rarefaction curves were comparable among sites, but a plateau was not fully reached in some 0-1 cm horizon samples for 18S, suggesting that some diversity was not captured in these samples with this marker (Fig. S1). With 18S, 9 out of 10 species were detected in both metazoan mock samples, although assignment accuracy ranged from genus to class-level, and three bivalve species were not correctly resolved and together only produced 1-2 OTUs. In contrast, COI detected all species in the mock samples, with assignments accurate down to genus-level for six species. The remaining species were correctly assigned to the class level, but two of them (scleractinian and gastropod) produced more than one OTU. Loci were subsampled to target phyla based on detection rate (Fig. 2), and final datasets comprised 6.7 (18S) and 21.3 (COI) million target reads, delivering 1,780 and 11,808 metazoan OTUs for 18S and COI respectively (Table S4).

3.2 OTU richness decreases with depth in the sediment

Metazoan OTU richness significantly decreased with increasing depth in the sediment, although the magnitude of this decrease varied significantly among regions (18S: Chisq = 219.8, $p < 0.001$; COI: Chisq = 1662.5, $p < 0.001$), and although there was significant site variability (18S: Chisq = 238.4, $p < 0.001$; COI: Chisq = 1608.2, $p < 0.001$). This pattern was observed in both studied marker genes (Fig. 3), and in major target phyla, except Tunicates, whose OTU numbers increased below 5 cm with 18S in three regions (Fig. S2).

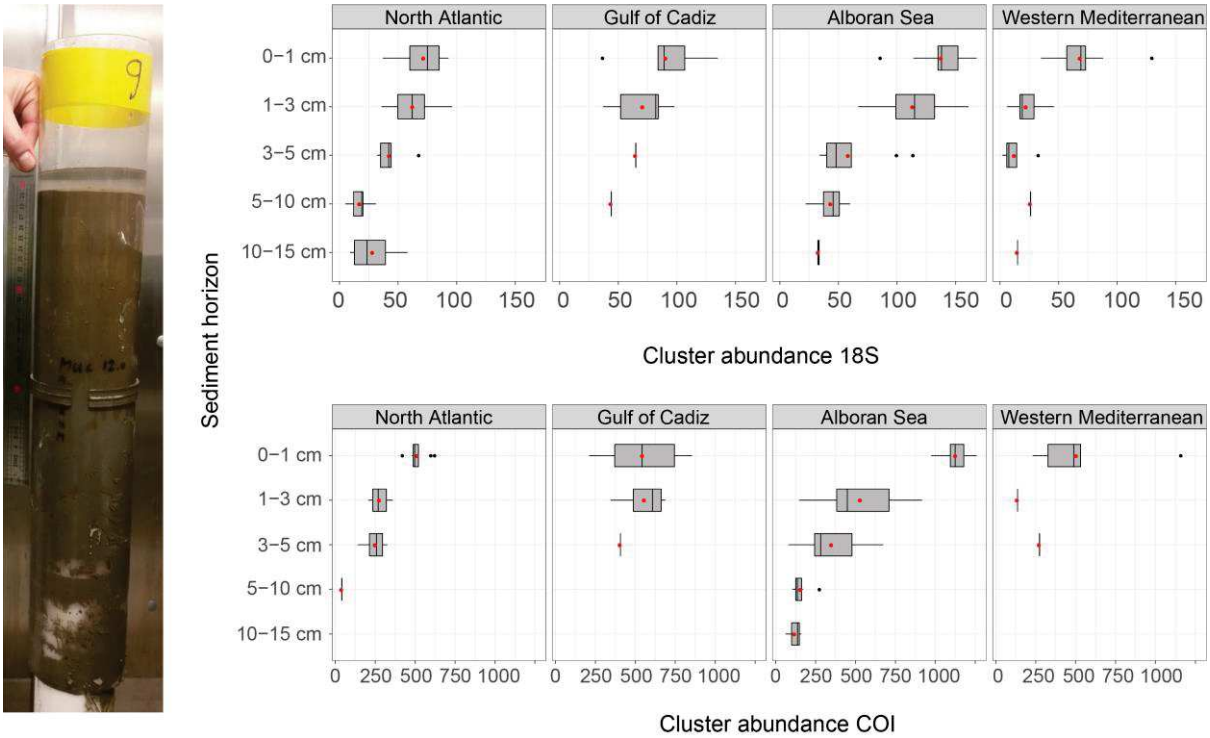


Figure 3. Metazoan Operational Taxonomic Unit (OTU) richness among sediment horizons and regions. Boxplots show detected numbers of metazoan OTUs (18S V1-V2, COI) recovered in each sediment layer, for the four regions evaluated in this study. Cluster abundances were calculated on rarefied datasets. Boxplots show medians with interquartile ranges. Red dots indicate mean values.

The upper 5 cm of sediment comprised 94% (18S) to 98% (COI) of all OTUs. The first horizon (0-1 cm) was the richest and contained the highest amount of unique OTUs, i.e. 46% for COI, 49% for 18S (Fig. S3). Following sediment layers shared more OTUs with their adjacent upper layer, then with their adjacent lower layer. However, communities in deeper sediment layers were not only a subsample of upper layers, as horizon contained from 14% to 31% of unique OTUs. Few OTUs were shared across sediment horizons (Fig. S3). Indeed, only 0.6% (COI) to 1.4% (18S) of OTUs were shared among all horizons (i.e. from 0-15 cm). OTUs shared across the first 10 cm accounted for only ~3% (COI) to 3.5% (18S) of all clusters, and these numbers were at 15% (COI) and 9% (18S) for the first 5 cm. The top two sediment layers shared more OTUs, as 34% (COI) and 30% (18S) of all clusters co-occurred in these horizons (0-1 cm and 1-3 cm).

Organic matter content and particle grain size were negatively correlated with depth in the sediment (Pearson's product-moment ρ = -0.15, p-value = 0.07 and ρ = -0.2, p-value = 0.02 respectively). However, OTU richness was neither correlated to OM content (ρ_{18S} = 0.04, ρ_{COI}

= 0.05, p-value > 0.1) nor to grain size (ρ_{18S} = 0.1, ρ_{COI} = -0.01, p-value > 0.1), and high richness was observed in samples with highly distinct values for these environmental variables (Fig. S4).

3.3 Numbers and nature of OTUs change across the Gibraltar Strait

As regional and habitat (i.e. site) scales significantly affected metazoan OTU richness (Fig. 3), we investigated unique and shared OTUs in each region. Only 12.5% (18S) and 13.3% (COI) of OTUs were found on both sides of the Gibraltar Strait (Fig. 4). The richest sites were located in the regions around the strait , i.e. the Gulf of Cadiz and the Alboran Sea. Both had mean OTU richness values at ~75 (18S) and 500 (COI) per site, compared to ≤ 50 (18S) and ~375 (COI) for other regions (Fig. 4). Both the Gulf of Cadiz (west of the strait) and the Alboran Sea (east of the strait) contained 59-67% of unique OTUs, and shared only 8% (18S, 111 OTUs) and 14% (COI, 947 OTUs) of their OTUs (Fig. 4). The North Atlantic sites harboured the highest levels of unique OTUs (68% for 18S and 83% for COI), compared to 27% (COI) to 40% (18S) of unique OTUs within the Western Mediterranean region. This region shared most of its OTUs with the Alboran Sea, but shared more OTUs with the North Atlantic sites than with the more closely located Gulf of Cadiz (Fig. 4).

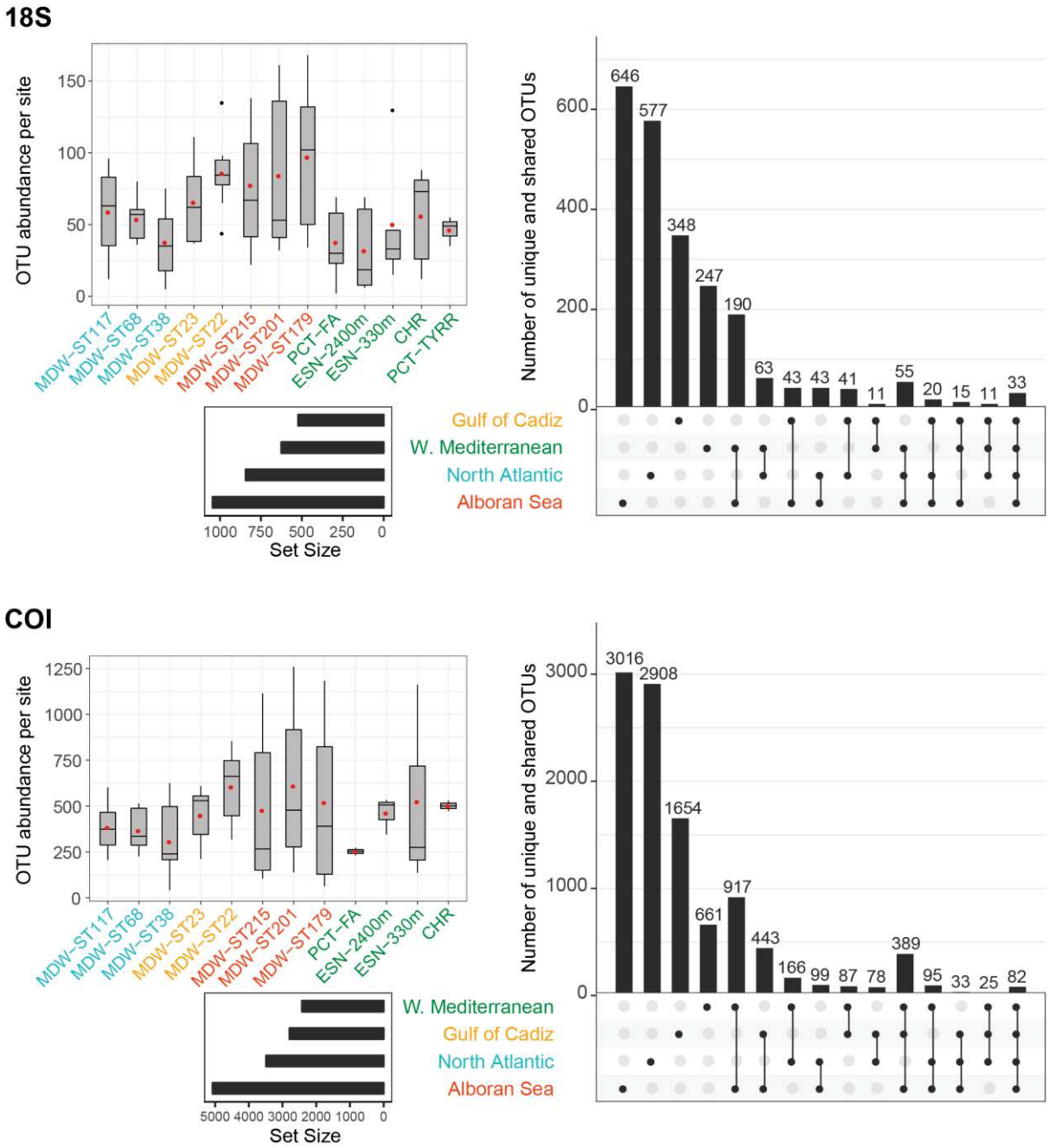


Figure 4. Richness and connectivity among regions of the Mediterranean-Atlantic transition zone. Venn diagrams showing numbers of shared OTUs among regions.

3.4 Influence of small and large-scale factors on community structures

PERMANOVA showed that community structures varied significantly among regions and sites within regions (Table 1). Community structure was also significantly affected by sediment horizon, although the way assemblages segregated with sediment layers varied in magnitude across sites and regions. Large-scale geographic patterns accounted for most variation in data (16-24% for Region, 14-15% for Site), and sediment horizon accounted for approximately 5% (18S) to 8% (COI) of variation among communities.

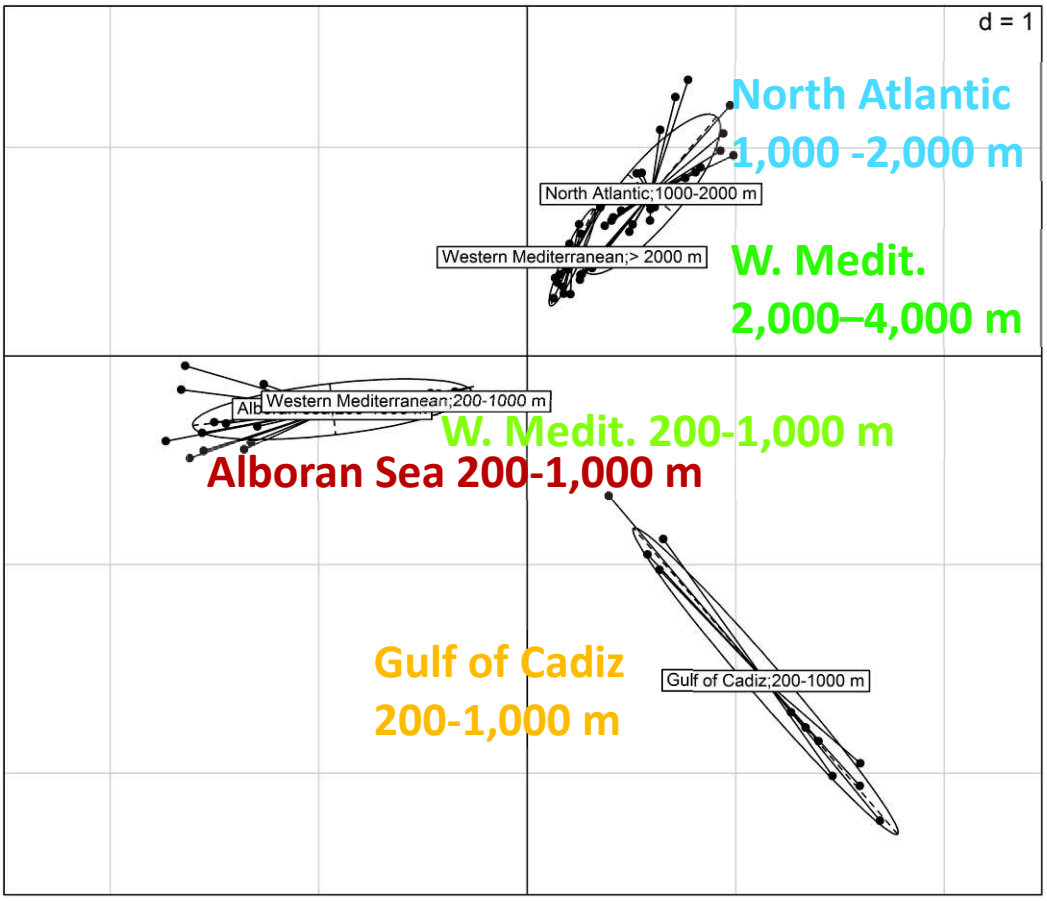
Pairwise comparisons indicated that, for 18S, strongest community segregation among sediment horizons occurred within the first 5 cm, while communities located between 3 cm and 15 cm were similar. In contrast, for COI, the two uppermost sediment layers were similar in terms of community structure, and community segregation was strongest from 3 cm to 10 cm in the sediment.

Table 2. Changes in community structure with region, site, and sediment horizon. PERMANOVAs were calculated on normalised datasets by permuting 1,000 times with Region as a blocking factor and site as a plot factor, using Jaccard distances. Significant p values are in bold. For pairwise comparisons, significance codes are $p < 0.001$: '***'; $p < 0.01$: '**'; $p < 0.05$: '*'.

LOCUS	**F-value**	**R^2**	***p-value***	**pairwise comparisons among horizons**
18S V1-V2				
Region	3.6	0.16	**0.001**	0-1 cm/1-3 cm**; 1-3 cm/ 3-5 cm**; 3-5 cm/5-10 cm; 5-10 cm/10-15 cm
Site(Region)	1.8	0.14	**0.001**	
Sediment horizon	1.6	0.05	**0.001**	
Region:Sediment horizon	1.4	0.11	**0.001**	
Site(Region):Sediment horizon	1.2	0.17	**0.008**	
COI				
Region	3.9	0.24	**0.01**	0-1 cm/1-3 cm; 1-3 cm/ 3-5 cm*; 3-5 cm/5-10 cm*; 5-10 cm/10-15 cm
Site(Region)	1.7	0.15	**0.01**	
Sediment horizon	2.1	0.08	**0.001**	
Region:Sediment horizon	1.5	0.09	**0.001**	
Site(Region):Sediment horizon	1.3	0.15	**0.001**	

Combined analysis of macro-, meiofauna, protists, and prokaryotes confirmed that three bioregions exist across the transition zone, with deep sites hosting similar communities across regions, while mesopelagic sites harboured different communities across the Gibraltar Strait, with the Gulf of Cadiz harbouring different communities than sites in the Mediterranean.

Segregation among sediment layers was observed in all taxonomic compartments, but differed in magnitude depending on size-class. K-tables analyses also showed that similarly sized taxonomic compartments were more strongly correlated, as RV coefficients were highest between the macro- and meiofauna (RV = 0.84) and between protists and prokaryotes (RV = 0.87), while they were at ~0.60 for other taxonomic pairs.



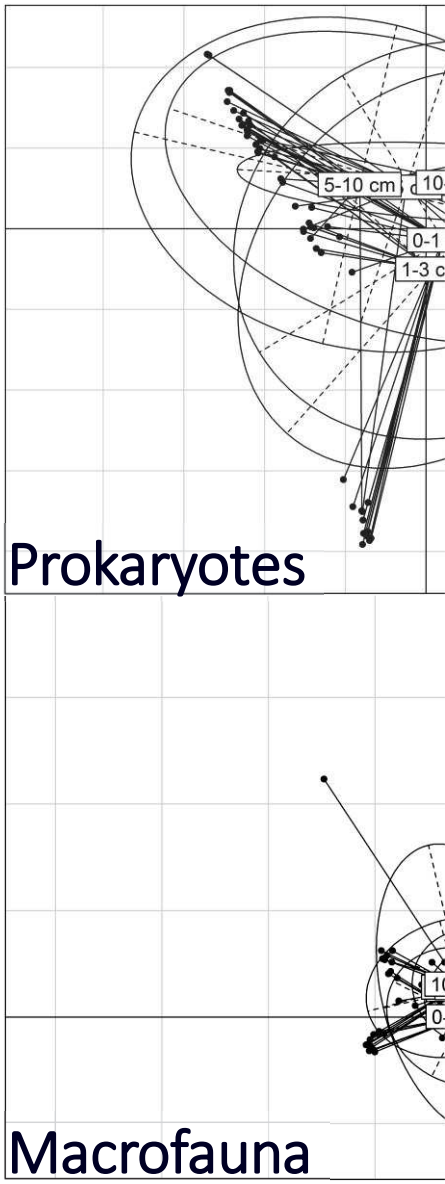


Figure 5. Correspondence Analysis (CA) ordinations of combined macro-, meiofauna, protist, and prokaryote datasets as performed by STATIS in ade4. Sample differences displayed by region (left) and by sediment horizon (right) show that deep (> 1,000 m) sites cluster across the transition zone, while mesopelagic community composition differs on either side of the Gibraltar Strait. Communities also differ among sediment layers, but the segregation differs n magnitude depening on size class.

PCoA ordinations showed different ecological patterns for meio and macrofauna (Fig 6). For meiofauna, communities strongly segregated by depth for Mediterranean sites, while Atlantic sites hosted similar communities across sites and depth zones. While community differentiation among sediment horizons was observed, it was much more pronounced in the

mesopelagic Mediterranean sites. Finally, and in contrast to mesopelagic sites, deep (> 1,000 m) meiofauna communities were similar in Mediterranean and Atlantic sites. For macrofauna, communities showed a strong segregation by ocean depth for both Mediterranean and Atlantic sites, with mesopelagic sites more similar across regions. They also differed among sediment horizons, with deeper horizons hosting more similar communities across regions. Finally, Canonical Analysis of Principal Coordinates (CAP) showed that sediment grain size, ocean depth, depth in the sediment, and longitude were significanlty explaining community structures, while organic matter (OM) was non-significant as it was redundant with longitude (Fig. 6). Local environmental variables mostly explained community differences among sites Indeed, community structures at bathyal and mesopelagic Mediterranean sites were characterized by higher OM content and small particle sizes, while assemblages in the North Atlantic (bathyal) and the Gulf of Cadiz (mesopelagic) were associated to larger particle sizes and lower OM content. However, CAPs also revealed a strong influence of depth on community composition, as mesopelagic sites (200-1,000 m) segregated from deep sites (>1,000 m), regardless the region.

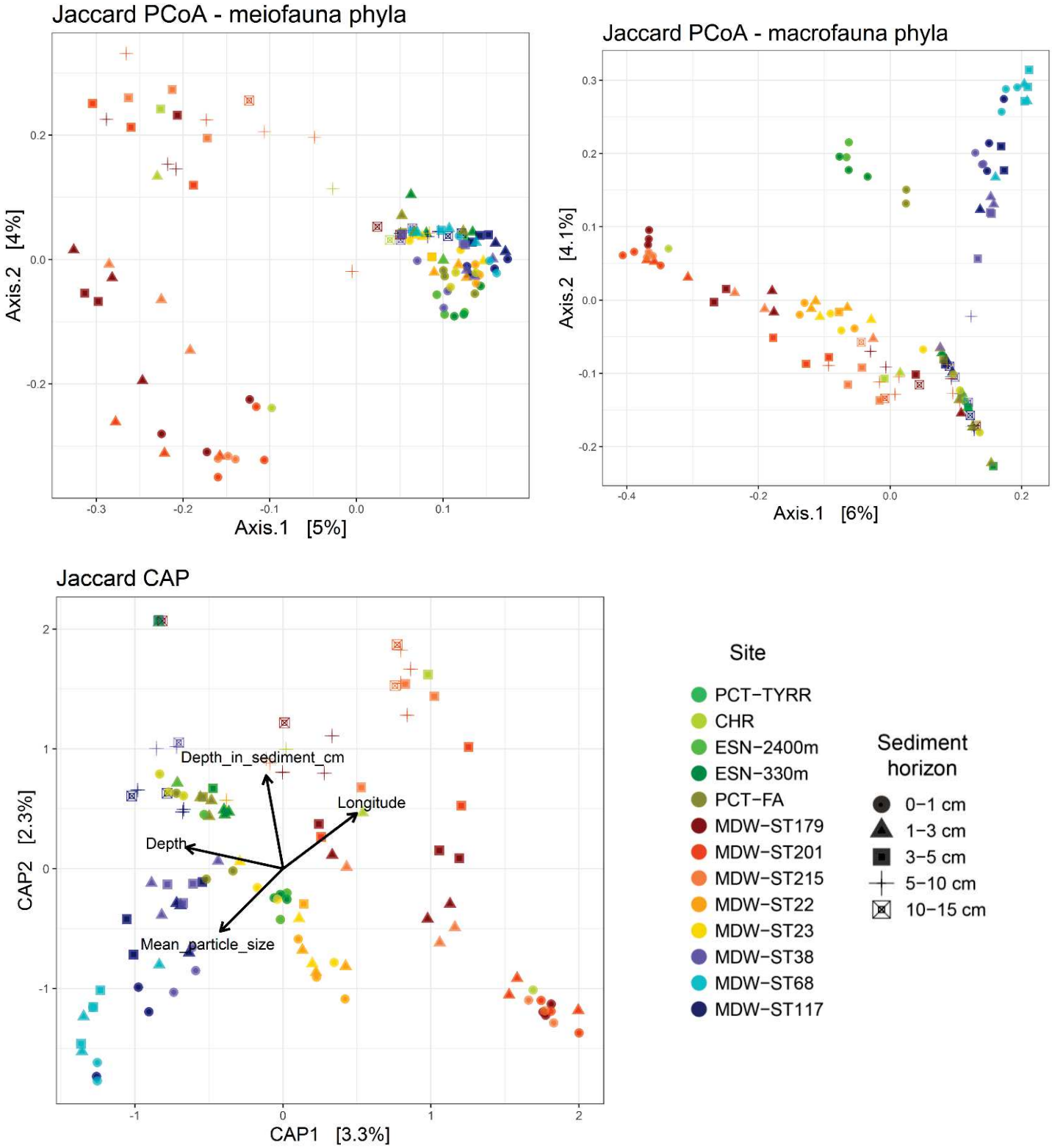


Figure 6. Principarl Coordinates Analysis (PCoA) and Canonical Analysis of Principal Coordinates (CAP) ordinations showing metazoan community differences between 13 deep-sea sites located in four regions covering the Mediterranean-Atlantic transition zone. The sites are coloured according to the region they belong to: green-scale for Western Mediterranean sites, red-scale for Alboran Sea, yellow-scale for Gulf of Cadiz, and blue-scale for North Atlantic. CAPs were calculated on rarefied datasets, using Jaccard

4 Discussion

The aim of this study was to investigate the extent and distribution of metazoan biodiversity at nested spatial scales across the Atlantic-Mediterranean transition zone, based on metabarcoding of environmental DNA using the 18S and COI barcode markers.

The taxonomic resolution of datasets differed markedly between both barcode genes used. This motivated our choice to subsample each marker to its best-detected phyla, thus avoiding taxonomic redundancy in the ecological analysis. Such approach has already been adopted for studying zooplankton patterns at ocean-basin scale (Stefanni et al. 2018) and is an effective way to take advantage of the complementarity of the 18S rRNA and COI barcode regions. Fig. 2 highlights this complementarity, showing that the greatest coverage of metazoan phyla can be achieved by combining both markers, and underlining that there is no clade (e.g., Metazoa, Protostomia, Lophotrochozoa) for which the use of either COI or 18S allows the detection of all phyla. Combining 18S and COI in a taxonomically non-redundant way therefore seems to be the most effective way to achieve more comprehensive biodiversity inventories. Finally, even though we subsampled each dataset based on numbers of OTUs detected in each phylum, it is noteworthy to highlight that 18S seemed to be better at detecting meiofauna (< 1 mm), while COI mostly detected macro- and megafauna (> 1 mm). The metazoan community structures were highly correlated (RV at 0.8), illustrating the strong interactions between metazoan size classes, confirming patterns reported in numerous studies of the deep-sea benthos based on morphological inventories (Gallucci et al. 2008; Buhl-Mortensen et al. 2010; Hasemann and Soltwedel 2011). Meiofauna communities, primarily detected by 18S, were found to be less correlated to prokaryote communities (RV=0.5) than the macro- and megafauna detected by COI (RV=0.65). This is in line with morphology-based studies that did not find an influence of prokaryote abundance, biomass, or activity on meiofauna organisms in the deep-sea (Danovaro et al. 1995; Górska et al. 2014).

This study confirmed that deep-sea metazoan richness and community structure can vary at very small scales, i.e. centimetres (Rosli et al. 2017; Leduc et al. 2015; Rosli et al. 2016; Leduc et al. 2012b; Ingels et al. 2011; Gallucci et al. 2009; Danovaro et al. 1995). Indeed, significant vertical segregation in community structure was revealed by the multivariate analyses, with sediment layer accounting for 5-8% of variation among communities, regardless the sampling region or barcode marker. Diversity as measured by OTU richness also significantly decreased with increasing depth in the sediment, although the magnitude of this decrease varied among regions and sites. Although most morphology-based studies only

investigated the upper 5 cm of sediment, decreases in species abundance and diversity with sediment layer have repeatedly been reported in deep-sea sediment assemblages, for e.g. in the Arctic (Fonseca, G. et al. 2010; Górska et al. 2014; Pfannkuche and Thiel 1987) or Pacific oceans (Rosli et al. 2016; Leduc et al. 2010; Snider et al. 1984). First investigations using eDNA metabarcoding in the deep-sea also reported these patterns (Guardiola et al. 2016). For taxa revealed by 18S, the decrease was most apparent below 3 cm, while for taxa revealed by COI there was a marked drop in richness below the uppermost centimetre of sediment in three out of four sampling regions (Fig. 3). This reflects different segregation patterns between the two types of taxa detected by both markers. Indeed, benthic megafauna, mostly revealed by COI, are epifaunal organisms living and feeding on the sediment surface, while taxa revealed by 18S are predominantly interstitial meiofauna, which can penetrate deeper into the sediment (Rex 1981). Consistently, the top sediment layer (0-1 cm) was dominated by mega and macrofauna OTUs well detected by COI, such as arthropods, cnidarians, molluscs, poriferans, or echinoderms, whose OTU numbers decreased strongly below 1 cm (Fig. S2). In contrast, 18S detected considerable OTUs numbers for meiofauna taxa such as nematodes, gastrotrichs, kinorhynchs, or tardigrades, and revealed a surprising diversity of platyhelminths and xenacoelomorphs in all sampling regions. This was also highlighted by other studies having applied eDNA metabarcoding in deep-sea sediments (Guardiola et al. 2016). Meiofauna have been shown to be primarily located in the upper 3 cm (Snider et al. 1984; Thiel 1983) to 5 cm (Giere 2009) of the sediment, although capable of penetrating as deep as 30 cm (Shirayama 1984). A constant peak in abundance is generally observed in the first centimetre, with the exception of some peculiar ecosystems or where strong currents are present (Zeppilli et al. 2014; 2012). These patterns are in congruence with the results observed here for 18S.

It is important to add that detected OTU numbers were substantially lower for 18S, but this is mostly due to the different taxonomic levels reached by each marker: 18S is less resolutive, revealing family to genus diversity, while COI more accurately reveals species diversity. Although OTU richness of most taxa decreased with sediment depth, with some phyla visibly more diverse in specific regions (e.g. Tardigrada in the Gulf of Cadiz), a notable exception was observed for the tunicates revealed by 18S, whose diversity increased below 5 cm in three out of the four studied regions (Fig. S2), highlighting and confirming that some taxa thrive deeper within the sediment (Steyaert et al. 2003).

Grain size and organic matter slightly decreased with increasing depth in the sediment, yet, OTU richness was not correlated to these variables, and regions with higher OM content

(Alboran Sea and Western Mediterranean, Fig. S3) did not show smoother gradients in OTU decrease with sediment depth, indicating that vertical richness patterns in the sediment cannot be explained by food availability alone. Similarly, while sediment granulometry is important in controlling faunal horizontal patterns (Zeppilli et al. 2016), it has been shown not to be the dominant factor explaining vertical diversity patterns, especially in deeper layers (Steyaert et al. 2003). Rather, it has been suggested that vertical patterns in deep-sea sediments arise from interrelating abiotic and biotic factors, such as oxygen and nitrogen content (Soetaert et al. 1997), organic matter composition and availability (Pfannkuche and Thiel 1987; Danovaro et al. 1995; Pusceddu et al. 2009), as well as interactions with larger fauna, for e.g. predator avoidance or facilitation due to bioturbation resulting in increased sediment oxygenation (Lambshead et al. 1995; Hasemann and Soltwedel 2011; Gallucci et al. 2008). The measurement of *in situ* nutrient levels combined with better *proxies* for food availability such as protein and lipid concentrations (Danovaro et al. 1995) was unfortunately not possible in this study, highlighting the need for integrative research programs combining biological and geochemical measurements to better elucidate deep-sea benthic biodiversity drivers at small scales. However, our results also highlight the importance of standardized sediment sampling schemes to ensure comparability among samples and studies, as different sediment layers do not reveal the same communities (Figs. 5-6).

Large-scale effects were predominant in explaining beta-diversity patterns at habitat (>100 m) and regional (>100 km) scales, accounting for 30-40% of variation among communities detected. Similarly, significant differences in OTU richness were observed among sites within regions and among regions, with lowest richness in the Western Mediterranean Sea compared to regions located westward (Alboran Sea to North Atlantic), a result comparable to observations made by Bianchelli et al. (2010) on the basis of morphological data. Although some studies have found small-scale effects to be more strongly affecting alpha and beta diversity than habitat or region effects, these studies usually focused on a single geographical location and/or habitat type (Górska et al. 2014; Ingels et al. 2011) or a single genus (Fonseca, G. et al. 2007). Consistently to our study, investigations including local to regional scales have found that variability in abundance of organisms, richness, or community structures is higher at larger scales (Pusceddu et al 2009). This could be explained by the very distinct ecosystems occurring over large scales, highlighted here by the diversity of seascapes sampled (Table S1).

Large-scale factors appeared to affect beta-diversity more than alpha diversity, as communities differed much more in terms of structure (Fig. 5) than in terms of richness (Fig 4)

across regions and sites. This in line with a previous study based on morphological inventories and comparing 3 deep-sea regions in the Mediterranean and North East Atlantic found high regional differences in beta diversity, but similar values of alpha diversity (Danovaro et al. 2009a). Similarly, habitat-scale effects have been shown to strongly affect community structure but showed little effect, if any, on taxonomic or functional diversity (Danovaro et al. 2013). This is congruent with our results showing that differences in ommunity composition were significantly linked to changes in organic matter content and sediment particle size (Fig 6), while these variables were not correlated to OTU richness (Fig. S3). Different availability, composition, and size spectra of food particles in sediments at habitat (> 100 m) to local (~1-100 m) scales are known to strongly affect the composition of deep-sea assemblages (Danovaro et al. 2013). Consequently, differences in biochemical composition of sediment organic matter were found to explain high beta-diversity between regions, as they increase diversification of benthic food webs (Pusceddu et al. 2009). Similarly, differences in sediment characteristics are known to affect community diversity and composition (Etter and Grassle 1992; Pape et al. 2013). Coarser sediments are associated with higher diversity on a broad horizontal scale, likely due to an increased range of microhabitats (Snider et al. 1984; Steyaert et al. 2003), and sediment particle size affects organism size, feeding mode, and locomotion mode (Vanreusel et al. 2010a; Leduc et al. 2012b). Our results highlight that organic matter composition and sediment grain size vary more strongly at the habitat (> 100 m) and regional (>100 km) scales, and thus mostly contribute explaining beta diversity patterns at larg spatial scales, rather than local alpha diversity.

Finally, depth also was shown to have a much stronger effect on community structure than on richness, as deep (> 1,000 m) and mesopelagic sites sites harboured significantly different communities (Figs. 5-6), although displaying comparable richness levels (Fig. 4), even though the depth effect was stronger in the Western Mediterranean for the meiofauna. This is in line with previous morphological studies investigating benthic diversity at intra-basin scales (Danovaro et al. 2009a; Bianchelli et al. 2010), highlighting that benthic assemblages strongly differ between mesopelagic and deep-sea environments (> 1,000 m), and supporting the hypothesis that depth is an important factor to consider for defining marine biogeographic realms (Levin, L. A. et al. 2001; Watling et al. 2013). Our results overall support the depth zonation proposed by Costello & Breyer (2017), who, based on a global marine diversity dataset identified 200 m and 1000 m to be critical depths shaping marine diversity.

Howeeever, this work did not provide strong evidence that deep water are less species rich. Instead, it highlighted that deep waters offer very distinct habitat conditions leading to specific niches and thus different species.

The deep sea was long described as homogeneous, lacking the obvious barriers to dispersal that characterize shallow waters, such as emerged lands or strong wind-induced water movements (McClain and Hardy 2010). Few studies have tackled the distribution of deep-sea diversity across distinct ocean basins thus far (Vanreusel et al. 2010a). In the present work, only approx. 13% of total OTUs were shared across the Gibraltar Strait, indicating a very limited exchange between Atlantic and Mediterranean basins. This highlights that the transition between Mediterranean and Atlantic basins is both a biogeographic barrier involved in vicariant events during environmental changes over geological timescales (Patarnello et al. 2007; Duranton et al. 2018) and a present day barrier to connectivity between the Mediterranean and Atlantic ocean basins (Catarino et al. 2017; Duranton et al. 2019). It remains unclear whether this barrier is a barrier to dispersal or to recruitement. K-tables analyses showed that community structures were more similar among deep regions across the strait than among shallow regions within the strait (Fig. 5). The Gulf of Cadiz being an inactive mud volcano habitat, is known to harbour exclusive species (Zeppilli et al. 2011), underlining that habitat-specific conditions are predominant in determining community structures. Our results thus support the hypothesis that the Gibraltar Strait is a barrier to recruitment rather than dispersal, however, our sampling effort was very fragmented, so this remains to be confirmed.

Acknowledgements

This work is part of the "*Pourquoi Pas les Abysses?*" project funded by Ifremer, and the project eDNAbyss (AP2016 -228) funded by France Génomique (ANR-10-INBS-09) and Genoscope-CEA. This work also received funding from the European Union's Horizon 2020 research and innovation programme under grant the agreement No 678760 (ATLAS). This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein. We wish to thank Bastien Mérigot for help on statistical analyses as well as Jan Pawlowski and Eva Ramirez Llodra for useful advice on this work. We also wish to express our gratitude to the crew of the Sarmiento de Gamboa and the participants of the MEDWAVES cruise (supported by the ATLAS project and the Spanish Ministry of Economy, Industry, and Competitivity), especially to all the people who helped collecting samples (Perregrino Cambeiro, Juancho Movilla, Maria Rakka, Joana

Boavida, Anna Addamo). We wish to thank the mission chiefs, crew, and participants of the ESSNAUT, PEACETIME and CANHROV cruises.

Author contributions

MIB and SAH designed the study, MIB, JP carried out the laboratory work. MIB and BT performed the bioinformatic analyses. MIB, BT, and NH performed the statistical analyses. MIB and SAH wrote the manuscript. All authors contributed to the final manuscript.

Data accessibility

The data for this work can be accessed in the European Nucleotide Archive (ENA) database (PRJEB38767 and PRJEB33873, see metadata excel sheet for ENA sample names). The dataset, including sequences, databases, as well as raw and refined ASV/OTU tables are available on https://doi.org/10.12770/cf00aa7b-67e7-49c4-8939-038c4a9d887f. Bioinformatic scripts can be accessed following the Gitlab link.

Chapter VI.
General discussion

The deep seafloor (>200 m depth) covers >60% of planet Earth (Snelgrove 1999; Smith, C. R. et al. 2008; Costello et al. 2010a). It can host high numbers of mostly small organisms (50,000-5 million individuals per square meter) that perform key ecosystem roles such as nutrient cycling, sediment stabilisation and transport, or secondary production (Rex et al. 2006; Zeppilli et al. 2018; Smith, K. L. et al. 2009). Although technological developments in the past 30 years have allowed remarkable advances, most deep-sea studies have been limited to local and regional scales due to the sheer vastness and remoteness of this biome, together with the long time required for morphological inventories and the lack of objective standards needed to merge together works performed by distinct experts. Consequently, we have so far explored less than 1% of the deep seafloor, and this contrasts with the fact that deep-sea ecosystems are under increased threat from a variety of direct and indirect anthropogenic pressures (Ramirez-Llodra et al. 2010).

Reducing deep-sea biodiversity gaps is therefore essential to better understand and predict how biodiversity will change in the context of global climate modifications, and how this will affect Earth's life-support systems. A recent review of scientific advances needed to reduce biodiversity gaps identified seven priorities (Saeedi et al. 2019), two of which are the core elements of this thesis, namely: 1) the "*Improvement and standardization of genetic, genomic, and other "omics" tools to aid in discovery, assessment, description, and cataloguing of biodiversity"* and 2) the need for "*Identifying biodiversity and biogeographic knowledge gaps and promoting efforts to reduce such gaps"*.

Indeed, while eDNA metabarcoding was identified as one of the most promising tools for achieving faster, cost-effective, and more comprehensive marine biodiversity assessments (Danovaro et al. 2016), many challenges remain to be resolved in order to apply eDNA methods on a broad scale (Cristescu and Hebert 2018).

This thesis addresses and evaluates several crucial methodological aspects for applying eDNA metabarcoding in deep-sea ecosystems, and provides an example of how this new tool can accelerate deep-sea exploration, supporting the idea that eDNA metabarcoding offers new perspectives to increase our understanding of deep sea biodiversity and biogeography (Danovaro et al. 2017; 2014). This work presented here explored paths allowing optimizing the eDNA metabarcoding workflow at bioinformatic (**chapter 2**), molecular (**chapter 3**), and sampling (**chapter 4**) levels. Major points for the successful and large application of eDNA metabarcoding in the deep-sea are discussed below.

Bioinformatic pipelines need to combine new tools in a flexible and user-friendly way

Abyss-pipeline (https://gitlab.ifremer.fr/abyss-project), the bioinformatic pipeline developed during this thesis (**chapter 2**) incorporates the newest advances for processing Illumina-sequenced metabarcoding data (Fig. 1). It addresses major sources of error by implementing the following tools. First, raw reads are corrected with DADA2 for effectively removing sequencing errors, the process also producing a read track table for obtaining a valuable and informative overview of read numbers throughout the process. Second, chimeras are removed after ASV inference, and again after swarm-clustering if this process is activated. Third, an abundance-renormalization filter is available to remove spurious clusters due to tag switching, and LULU-curation is available to remove additional spurious clusters. ASVs can optionally be clustered into OTUs using swarm v2, an iterative single-linkage algorithm that allows more fine-scale and data-dependent clustering than previous algorithm based on arbitrary thresholds. Finally, taxonomic assignment can be performed via BLAST and the RDP Bayesian Classifier for both ASVs and OTUs. All these processes are implemented independently to allow maximum user-control, and application on metazoan mock samples showed that the combination of these tools allowed achieving a near 1:1 species-OTU relationship.

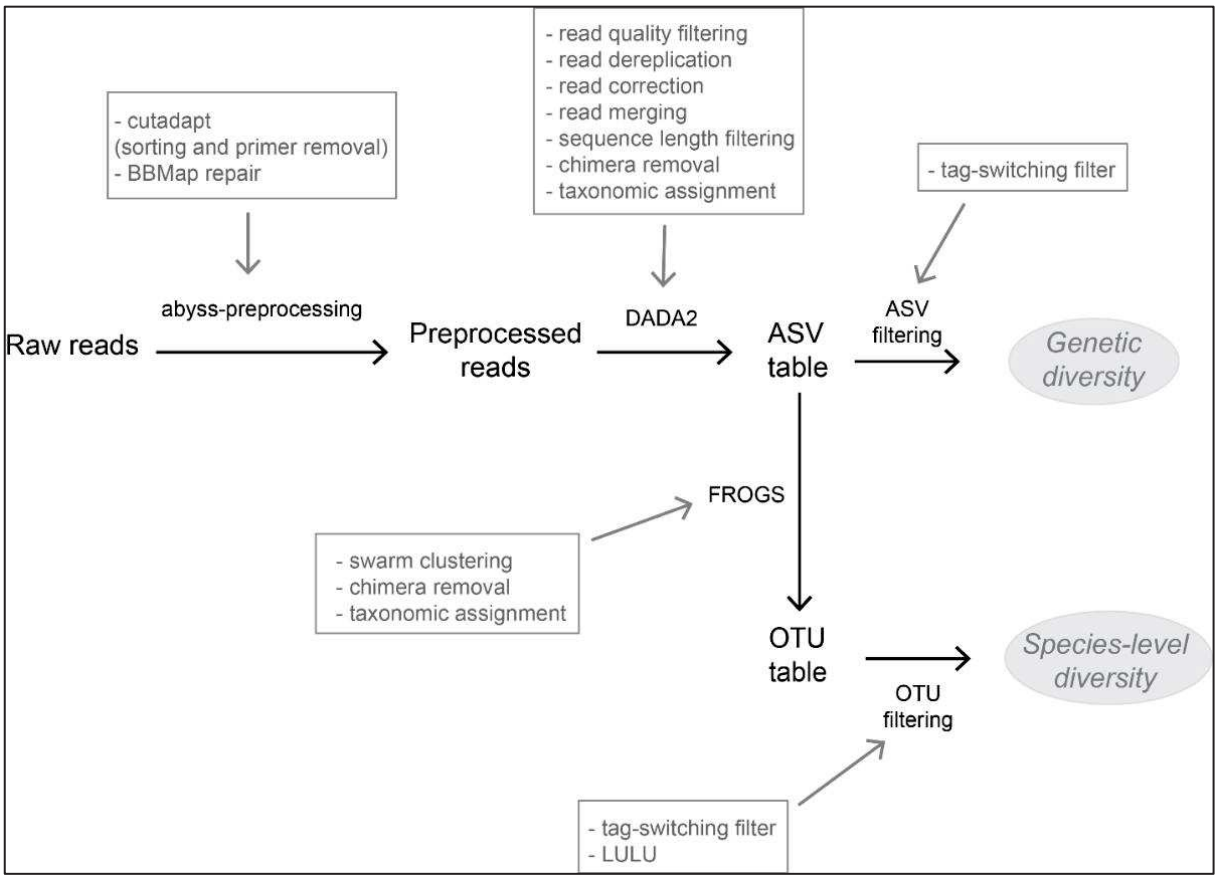


Figure 1. Schematic of the bioinformatic pipeline developed during this thesis for the improved analysis of metabarcoding data.

Chapter 2 highlighted that the choice of the molecular entity used as a *proxy* for taxa is crucial to obtain reliable inventories, and this choice depends on the taxonomic compartment of interest. While ASVs accurately describe high-resolution genetic diversity and may be appropriate for the study of unicellular organisms exhibiting lower intraspecific variation rates, or to infer the distribution of genetic rather than species diversity, they lead to an overestimation of the number of clusters for metazoans (Brandt, M. I. et al. 2020). For these taxa, sequence error correction thus needs to be combined with clustering and LULU-curation in order to obtain more realistic species inventories.

While Graphical User Interfaces in web applications such as SLIM (Dufresne et al. 2019) may facilitate bioinformatic analyses, especially for less-experienced users, web applications remain limited in the quantity of data they can process, and the limited ability for parameter adjustment, especially during initial data preparation, quality filtering, or error correction. The

analysis of large or multi-marker datasets therefore still requires customizable scripts, such as the shell scripts provided by abyss-pipeline.

As E.F. Schumacher highlights in his book "Small Is Beautiful", technology's primary purpose is to lighten the burden of work, and we therefore need methods and equipment which are "*cheap enough so that they are accessible to everyone, suitable for small-scale applications* [i.e. low-cost], *compatible with man's need for creativity*". Most metabarcoding-related bioinformatic tools are freely available online, however implementing and learning to use them remains complex and difficult for most members of the research community, as most are new to informatics. A significant way forward towards the simplification of bioinformatic processing would be to adapt the tools to the users (and not the other way around!) by making them directly usable in R, a software familiar to most biologists. This idea is supported by the fact that recent algorithmic advances used in this thesis such as DADA2 and LULU are already coded in R. Furthermore, these tools could then be combined into an R-based pipeline, as exemplified by the "Just Another Metabarcoding Pipeline" (https://github.com/VascoElbrecht/JAMP) R package (Elbrecht et al. 2018).

eRNA is associated to more bias than eDNA potentially containing traces of ancient DNA

Chapter 3 largely confirmed doubts about the capacity of eRNA to better describe live communities (Cristescu 2019). This study was the first to compare co-extracted eRNA and eDNA biodiversity inventories using ribosomal and mitochondrial markers targeting prokaryote, protistan, and metazoan life compartments.

With ribosomal loci, RNA, while resolving similar spatial patterns than co-extracted DNA, resulted in significantly higher richness estimates. This supports hypotheses of increased persistence of rRNA in the environment, and of increased amounts of spurious clusters with eRNA due to high but unmeasured artefacts produced during reverse transcription of RNA to cDNA (Cristescu 2019; Laroche et al. 2017), highlighted here by the greater amounts of chimeras observed with RNA in ribosomal loci. Contrastingly, with the mitochondrial COI marker, RNA detected significantly lower metazoan richness, resolved less spatial patterns than co-extracted DNA, and was associated to increased sample failures. This reflects high messenger RNA lability, making it unsuitable for large-scale ecological surveys.

Moreover, eRNA may lead to potentially significant taxonomic bias using any marker gene, due to abundant but taxon-specific RNA release, either passively after death (exacerbated by the fact that RNA is way more abundant than DNA in living organisms), or actively and this varying according to metabolic levels and/or life stages (Torti et al. 2015; Blazewicz et al. 2013).

In contrast, our approach aiming to remove ancient DNA by removing DNA fragments in the aDNA size range (< 1,000 bp, Lennon et al. 2018; Boere et al. 2011; Lejzerowicz et al. 2013a; Coolen et al. 2013), did not show any effect on alpha or beta diversity patterns. Of course, aDNA may also be archived in vesicles or other organelles, although there is increasing evidence that DNA from non-living cells is mostly contemporary (Lennon et al. 2018). This suggests that, even if aDNA may be present in deep-sea sediments, the eDNA metabarcoding workflow will primarily target contemporary DNA most likely due to 1) its overabundance in the environment and 2) DNA extraction protocols unsuited for aDNA preservation.

Standardized and replicated sampling is needed to ensure comprehensive, reproducible, and comparable results

It is generally known that ~10% of PCRs fail (Andreson et al. 2008), and this for a number of technical reasons (see **Chapter 1**, Fig. 4). Consequently, strong research focus has gone into evaluating the effect of technical PCR replicates, and numerous studies have stressed that replication of PCR reactions, as well as their independent sequencing is essential to increase detection probability and reliability of results (Ficetola et al. 2015; Alberdi et al. 2017; Leray and Knowlton 2017; Dopheide et al. 2019).

However, PCR replicates are technical replicates, they are not independent as they originate from the same sample, and they can therefore not be used as replicates in statistical tests commonly used in ecology. Similar problems arise with DNA extraction replicates, which, although improving diversity estimates (Lanzén et al. 2017), are not adequate replicates for statistical analyses in a strict sense. Only biological sample replicates allow statistically evaluating within-site variability, and the research focus on PCR and extraction replicates has unfortunately been associated with a decrease in attention on biological replication. This is highlighted by the general absence of stated rationales explaining the number and types of

replicates, and the frequent absence or inadequacy of replication in many metabarcoding studies (Dickie et al. 2018).

Of course, replication should ideally be performed on both technical and biological levels. However, given the cost of sampling, DNA extraction, PCR, and sequencing, it is essential to determine an appropriate trade-off between logistic feasibility and adequate replication, and this depends on the study objectives. For example, it appears important to process PCR replicates independently throughout the metabarcoding workflow if the detection of rare or ancient species is the primary research interest (Ficetola et al. 2015). However, PCR replicates can be pooled for sequencing in order to smooth intra-sample variance and address PCR bias, effectively reducing sequencing costs while still allowing characterizing large-scale patterns (Dickie et al. 2018).

Small-scale (centimetres to metres) patchiness is common in the deep-sea, with patterns of patchiness varying with taxonomic compartment (Grassle and Maciolek 1992; Rex 1981; Lejzerowicz et al. 2014). This leads to considerable within-site variability, which can only be mitigated by collecting adequate numbers of biological replicates per sampling site, but also been addressed by adapting the sampling gear to each benthic size compartment. Indeed, multicorers are generally used for nano- and meiofauna as they are the only tools preserving vertical sediment stratification, box corers are often used for macrofauna as they cover a larger sampling area, and epibenthic sledges are commonly used for megafauna (Montagna et al. 2017). Given that a full deployment series of these gears takes approx. 40 hours, ecological research in the deep-sea is inherently confronted with the trade-off between number of gear deployments per site and number of sites to sample (Daniela Zeppilli, *pers. comm.*). It can be argued that, eDNA metabarcoding effectively detecting small traces of organisms (mucus, shed cells), multicorers may also be adequate for the study of larger fauna, as those will be detected even if not present in the sediment. Moreover, box corers and sledges are known to create strong bow waves that wash off the upper layer of sediment, thus leading to the loss of many organisms. This explains why a recent long-term study found that box corers underestimated total macrofauna density by a factor of 2.9 times compared to multicorers, and reported that they underestimate richness relative to area sampled (Montagna et al. 2017).

Given the extent of the unknown in the deep sea, there is unsurprisingly a lack of consensus on the type and number of replicates appropriate to collect for a typical spatial study on the deep-sea benthos: some authors argue that individual cores from the same multicorer deployment are statistically not ideal as these cores are not collected independently, and are

thus pseudoreplicates rather than replicates (Colegrave and Ruxton 2018). However, Montagna et al. (2017) found more variability between cores of the same multicorer deployments (i.e., pseudoreplicates) than between deployments (true replicates), consistent with the strong small-scale patchiness in benthic fauna reported in the deep-sea and the limited overall variability reported at local scales (see **chapter 1**). Moreover, the difficulty of controlling the exact location of multicorer deployments and the significant patchiness in species distribution, primarily resulting from local scale seafloor heterogeneity (Zeppilli et al. 2016) makes it unclear to what extent multicorer deployments are representative replicates of single sampling location. Results obtained in **chapter 5** highlight this issue, as the multicorer deployments (sites) sampled in each region showed significant differences in richness and community structure, even when they were targeting the same location (e.g., Gulf of Cadiz, Alboran Sea).

It therefore appears that spatial studies should consider sampling more stations/locations, each with few multicore deployments. Whether cores within deployments can be considered true replicates remains to be confirmed on a global scale as results from Montagna et al. (2017) were geographically restricted although covering a 14 year time period. If their findings can e generalised, subsamples of cores within a deployment could allow increased precision per replicate. Measuring this within-plot variability could help better understanding the spatial heterogeneity of deep-sea benthic organisms (Dickie et al. 2018).

Finally, **chapter 3** highlighted that sediment quantities ≥10 g should be used to accurately detect eukaryote (incl. non-metazoan) diversity and that 2 g of sediment were insufficient to account for small-scale spatial heterogeneity, a finding already reported in numerous other studies in terrestrial soils and marine sediments (Creer et al. 2016; Nascimento et al. 2018). While organism size sorting through sediment sieving allowed, as expected (Elbrecht et al. 2017), detecting higher meiofauna diversity (**chapter 4**), similar spatial patterns and taxon compositions were obtained in sieved and non-sieved samples, indicating that the considerable time-costs associated with sieving are not essential for inferring robust ecological patterns. This is supported by the fact that sample washing and sieving may lead to substantial loss of organisms (Montagna et al. 2017), added to the increased risk of contamination by foreign DNA, particularly if the protist and prokaryote size compartments are of interest. Moreover, **chapter 4** highlighted that aboveground water cannot be used for assessing benthic diversity with eDNA, even when large water volumes are sampled, supporting recent studies performed at shallower depths (Antich et al. 2020). Finally, **chapter 5** highlighted that reporting sediment depth layers used is crucial to allow comparability among studies, and that sampling design

should include the 0-1 cm, 1-3 cm, and 3-5 cm sediment horizons as these were associated with low processing failures and consistent differences across all sites for both the COI and 18S markers. Moreover, sampling should combine measurements of physical and chemical parameters with biological species detection to allow better estimation of community structure and function (Costello et al. 2018). While deep water samples specifically targeted pelagic organisms (**chapter 4**), sampling tools significantly affected the type of taxa detected. I*n situ* pumps were shown to have great potential in low biomass deep-sea waters, but being limited by mesh size, they are therefore more appropriate for assessing metazoan diversity, but only capture a fraction of microorganisms. Sampling tools allowing the recovery of small size classes remain necessary to comprehensively detect microbial diversity.

Taxonomic assignments of deep-sea metabarcoding datasets are unreliable beyond phylum-level when using public reference databases

Obtaining species names is useful for inferring ecological traits, as behind each name, there is a phenotype (with all its variability and life forms), an ecological role, and a geographic distribution. However, taxonomic assignments of sequences are only as good as the databases they are based upon. Deep-sea benthic metabarcoding datasets face the double challenge of focusing on taxonomic groups that are both highly diverse and poorly represented in public sequence reference databases. **Chapter 2** thus highlighted that it is difficult to obtain accurate taxonomic assignments even for megafauna, as we failed to obtain high-resolution taxonomic assignments for several species in the mock samples, for both the COI and 18S markers. With 18S, a high number of OTUs were left unassigned at the phylum-level, and percentage identities to reference sequences of OTUs in the sediment samples ranged from 80% to 100%. They were ≤ 86% for most OTUs with COI.

It has been suggested that critical cut-off values to ensure correct phylum-level assignment are 80% identity for COI reads, and 86% identity for 18S reads (Stefanni et al. 2018). This suggests that, while most of the 18S OTUs analysed in this thesis were most likely correctly identified at phylum-level, this is not the case for COI OTUs. Similarly, accurate taxonomic assignments down to genus or species level are unlikely using currently available databases, even for rRNA, where confidence in taxonomic assignments at the genus-level can only be ensured above 95% identity (Edgar 2018b). This explains why only ~2% of COI and ~39% of 18S OTUs were found to have acceptable genus RDP bootstraps in **chapter 2**. As taxonomic

assignments of OTUs in our deep-sea metabarcoding datasets were not satisfactory using publicly available databases, especially not for COI, we thus chose to focus **chapter 5** solely on numerical ecology.

Sequence-based techniques require the availability of comprehensive but also high-quality reference databases. Concerns about misannotation errors in large public repositories (i.e. GenBank, ENA, and DNA Data Bank of Japan) have been emitted based on analyses of particular groups (Leray et al. 2019), and these have resulted in a general mistrust in these repositories (Bidartondo 2008). This has prompted the development of curated sequence databases for particular taxonomic groups and genes, including BOLD, PR2, SILVA, GREENGENES, and MIDORI. Surprisingly, latest studies evaluate at ~17% the annotation errors in some of these curated databases (Edgar 2018c) and found remarkably accurate metazoan identifications in GenBank, even at low taxonomic levels (likely < 1% error rate at the genus level, Leray et al. 2019). This suggests that the limiting factor towards accurate taxonomic assignments is not the quality of database submissions, but rather their quantity.

Accurate taxonomic assignments in large-scale deep-sea biodiversity studies will thus only be obtained if a concerted effort is made to fill database gaps, which requires the integration of barcoding (and associated morphological identification) and metabarcoding approaches. Moreover, arbitrarily large databases containing a great diversity of taxa, and sequences that have not been truncated to the target sequence length can decrease the number of accurate taxonomic assignments (Macheriotou et al. 2019). Thus, concerted effort should also be directed towards developing user-friendly methods to build taxon-specific databases from large repositories. These methods could easily be part of R-based bioinformatic pipelines, as R scripts that automatically retrieve target sequences from databases are already available at https://github.com/metabarpark/R_scripts_querying_databases.

Improving and filling database gaps is going to take time. However, in the context of global change, there is urgent need to develop and apply biomonitoring programmes in the marine biome. This is highlighted by the significant number of studies evaluating the performance of metabarcoding-based environmental impact assessment (Pawlowski et al. 2016b; a; Cordier et al. 2019a; Chariton et al. 2015; Gibson, J. F. et al. 2015; Aylagas et al. 2018; Stat et al. 2017; Vivien et al. 2015), and developing genetic biotic indices (Pawlowski et al. 2015; Visco et al. 2015; Aylagas et al. 2014; Cordier and Pawlowski 2018; Keeley et al. 2018; Pawlowski et al. 2018). As the calculation of most indices relies on taxonomic identities, applications of eDNA metabarcoding in biomonitoring and impact assessment are especially dependent on good

taxonomy. To circumvent this, and allow efficient application in a time of database gaps, new approaches using supervised machine learning (SML) have demonstrated that SML can be used to predict accurate biotic indices from metabarcoding data (Cordier et al. 2017). SML approaches even outperform assessments relying solely on taxonomically assigned sequences, and this for a variety of eukaryotic and prokaryotic markers (Cordier et al. 2018), highlighting that accurate eDNA bioassessment is possible even in poorly referenced ecosystems such as the deep-sea.

Large-scale ecological studies are now possible in the deep-sea, even within a typical research-project timeframe.

Morphology-based studies, although essential for species descriptions, are limited in terms of large-scale ecological applications. They usually focus on a few, well-described taxa due to the limited amount of specialised taxonomists. The time-consuming (~ 1 month for one sediment core with five sediment horizons) identification of organisms explains why these studies rarely go beyond local to regional scales. Moreover, taxon identification is highly dependent on investigators, making it difficult to combine inventories from several studies.

Chapter 5, investigating benthic biodiversity patterns in the deep Atlantic-Mediterranean transition zone, found concordant results with morphological studies from the last decades, indicating that eDNA metabarcoding is an appropriate tool for ecological research on the deep seafloor. It provides major advantages including allowing studying various biological compartments simultaneously, effectively detecting diversity of small organisms (micro-organisms, meiofauna) and even traces of organisms, as well as being cost and time efficient.

As significantly different assemblages were found at mesopelagic (<1,000 m) vs bathyal and abyssal depths (> 1,000 m), **Chapter 5** also highlighted the need to consider depth for defining biogeographical realms, something often disregarded in previous studies, even those covering “all accessible data for all taxa in all oceans” (Costello et al. 2017). Recent work has shown that the global latitudinal marine species richness gradient follows a bi-modal pattern related to temperature and habitat availability, temperature being primarily an indicator of food availability (Chaudhary et al. 2016; 2017). However, temperature is not a good proxy for food availability in the deep-sea, where organic matter input, as measured by, for e.g., POC flux, is considered more important in defining species distribution patterns (Woolley et al. 2016).

Organic matter input varies among regions but also with depth, highlighting that it is unlikely that depth does not play a role in marine species distributions.

Conclusions and future directions

In 2010, the Conference of the Parties of the Convention on Biological Diversity agreed on the Strategic Plan for Biodiversity 2011-2020, and established five "Strategic Goals" that were divided into 20 targets. Each so-called 'Aichi Target' was designed to better understand and predict biodiversity, in particular, how biological diversity underpins ecosystem function, and how ecosystem services are essential for human well-being. Meeting these Aichi Targets ultimately secures livelihoods and economic development, and is essential for biodiversity maintenance and poverty reduction (Tittensor et al. 2010; Shepherd et al. 2016).

The Strategic Plan for Biodiversity 2011-2020 made bridging biodiversity knowledge gaps, as well as improving marine environmental status assessment and biodiversity monitoring a requirement in many countries worldwide, and eDNA metabarcoding was been identified as a tool that will allow achieving these requirements (Danovaro et al. 2016).

Biodiversity knowledge gaps are especially severe in the deep-sea, where >90% of expanses remain unexplored. This thesis evaluated and optimized the eDNA metabarcoding workflow for deep-sea sediments on a bioinformatic, molecular, and sample processing level, across multiple life compartments, opening the door to large-scale biodiversity surveys in the deep-sea realm, thus contributing to a better understanding of biodiversity, biogeography, and ecosystem functioning in this vast and elusive backyard. This also paves the way for establishing efficient biomonitoring protocols that are increasingly needed in areas of resource extraction.

Elucidating biodiversity-ecosystem functioning relationships with eDNA metabarcoding still remains a challenge, especially in poorly referenced ecosystems such as the deep-sea where taxonomic assignments are of poor accuracy. Methods for detecting functional traits from sequence data are only starting to be investigated, and various approaches, including attribute prediction software (Edgar 2017a), ancestral state reconstruction via reference phylogenetic trees (Keck et al. 2018) and machine-learning approaches (Vacher et al. 2016; Cordier et al. 2018) have been evaluated. Currently, supervised machine learning represents the most promising way towards ascribing ecological roles to OTUs. However, these SML approaches still require reference datasets upon which they can be trained to make accurate predictions of unknown samples, and this will require controlled mesocosm experiments. Ecological networks are useful for representing and analysing all the interactions between species, thereby offering

an understanding of ecosystem functioning, and machine-learning approaches can be used for reconstructing such networks based on HTS co-occurrence data (Vacher et al. 2016). Future developments of these methods may allow thus us to discover and monitor species interactions, even in unknown environments.

PCR-based approaches ultimately limit the ability of metabarcoding to accurately depict species abundance in complex communities, especially for metazoans with strongly varying numbers of marker gene copies and cells. Shotgun sequencing of mitochondrial genomes, i.e. genome skimming, has been shown to more reliably describe read-biomass relationships, while also allowing more accurate species identification as taxonomic assignment is based on whole mitogenomes (Bista et al. 2018; Gómez-Rodríguez et al. 2015; Fernández et al. 2015). This approach has been successfully applied on bulk samples, where DNA concentrations are typically high compared to extracts based on environmental samples. The application of these methods on deep-sea sediments will thus require mtDNA enrichment method, such as sequence-capture by hybridization techniques (Liu et al. 2016; Jones and Good 2016; Gasc et al. 2016; Wilcox et al. 2018; Cruz-Dávalos et al. 2017).

Finally, while new sequencing technologies are increasingly reliable, cost effective, and accessible, they remain inadequate for low-resource field-based applications. New portable, low-cost HTS devices (e.g. MinION sequencer from Oxford Nanopore Technologies) combined with portable lab systems such as miniPCR (Marx 2015), are opening the door to real-time biodiversity assessment, and successful applications have already been reported in space (Castro-Wallace et al. 2017), in the arctic (Goordial et al. 2017), or in the rainforest (Pomerantz et al. 2018).

Bibliography

Acinas, S.G., L.A. Marcelino, V. Klepac-Ceraj, and M.F. Polz. 2004. "Divergence and Redundancy of 16S RRNA Sequences in Genomes with Multiple Rrn Operons." *Journal of Bacteriology* 186 (9): 2629–35. https://doi.org/10.1128/JB.186.9.2629-2635.2004.

Agusti, S., J.I. Gonzalez-Gordillo, D. Vaque, M. Estrada, M.I. Cerezo, G. Salazar, J.M. Gasol, and C.M. Duarte. 2015. "Ubiquitous Healthy Diatoms in the Deep Sea Confirm Deep Carbon Injection by the Biological Pump." *Nature Communications* 6. https://doi.org/Artn 760810.1038/Ncomms8608.

Alberdi, A., O. Aizpurua, M.T.P. Gilbert, and K. Bohmann. 2017. "Scrutinizing Key Steps for Reliable Metabarcoding of Environmental Samples." Edited by Mahon, A. *Methods in Ecology and Evolution*, 2017. https://doi.org/10.1111/2041-210X.12849.

Alberti, A., J. Poulain, S. Engelen, K. Labadie, S. Romac, I. Ferrera, G. Albini, et al. 2017. "Viral to Metazoan Marine Plankton Nucleotide Sequences from the Tara Oceans Expedition." *Scientific Data* 4: 1–20. https://doi.org/10.1038/sdata.2017.93.

Alroy, J. 2002. "How Many Named Species Are Valid?" *Proceedings of the National Academy of Sciences of the United States of America* 99 (6): 3706–11. https://doi.org/10.1073/pnas.062691099.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215 (3): 403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.

Amaral-Zettler, L.A., E.A. McCliment, H.W. Ducklow, and S.M. Huse. 2009. "A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes." *PLoS ONE* 4 (7): 1–9. https://doi.org/10.1371/journal.pone.0006372.

Amir, A., D. McDonald, J.A. Navas-Molina, E. Kopylova, J.T. Morton, Z. Zech Xu, E.P. Kightley, et al. 2017. "Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns." *MSystems* 2 (2): e00191-16. https://doi.org/10.1128/msystems.00191-16.

Anderson, M.J., T.O. Crist, J.M. Chase, M. Vellend, B.D. Inouye, A.L. Freestone, N.J. Sanders, et al. 2011. "Navigating the Multiple Meanings of Beta Diversity: A Roadmap for the Practicing Ecologist." *Ecology Letters* 14 (1): 19–28. https://doi.org/10.1111/j.1461-0248.2010.01552.x.

Anderson, M.J., and T.J. Willis. 2003. "Canonical Analysis of Principal Coordinates: A Useful Method of Constrained Ordination for Ecology." *Ecology* 84 (2): 511–25. https://doi.org/10.1890/0012-9658(2003)084[0511:CAOPCA]2.0.CO;2.

Andreson, R., T. Mos, and M. Remm. 2008. "Predicting Failure Rate of PCR in Large Genomes." *Nucleic Acids Research* 36 (11): e66. https://doi.org/10.1093/nar/gkn290.

Andruszkiewicz, E.A., L.M. Sassoubre, and A.B. Boehm. 2017. "Persistence of Marine Fish Environmental DNA and the Influence of Sunlight." Edited by Doi, H. *PLoS ONE* 12 (9): e0185043. https://doi.org/10.1371/journal.pone.0185043.

Andújar, C., P. Arribas, C. Gray, C. Bruce, G. Woodward, D.W. Yu, and A.P. Vogler. 2018a. "Metabarcoding of Freshwater Invertebrates to Detect the Effects of a Pesticide Spill." *Molecular Ecology* 27 (1): 146–66. https://doi.org/10.1111/mec.14410.

Andújar, C., P. Arribas, D.W. Yu, A.P. Vogler, and B.C. Emerson. 2018b. "Why the COI Barcode Should Be the Community DNA Metabarcode for the Metazoa." *Molecular Ecology* 27 (20): 3968–75. https://doi.org/10.1111/mec.14844.

Antich, A., C. Palacín, E. Cebrian, R. Golo, O.S. Wangensteen, and X. Turon. 2020. "Marine Biomonitoring with EDNA: Can Metabarcoding of Water Samples Cut It as a Tool for Surveying Benthic Communities?" *Molecular Ecology*, September, mec.15641. https://doi.org/10.1111/mec.15641.

Appeltans, W., S.T. Ahyong, G. Anderson, M. V. Angel, T. Artois, N. Bailly, R. Bamber, et al. 2012. "The Magnitude of Global Marine Species Diversity." *Current Biology* 22 (23): 2189–2202. https://doi.org/10.1016/j.cub.2012.09.036.

Atienza, S., M. Guardiola, K. Præbel, A. Antich, X. Turon, and O.S. Wangensteen. 2020. "DNA Metabarcoding of Deep-Sea Sediment Communities Using COI: Community Assessment, Spatio-Temporal Patterns and Comparison with 18S RDNA." *Diversity* 12 (4). https://doi.org/10.3390/D12040123.

Aylagas, E., Á. Borja, X. Irigoien, and N. Rodríguez-Ezpeleta. 2016. "Benchmarking DNA Metabarcoding for Biodiversity-Based Monitoring and Assessment." *Frontiers in Marine Science* 3. https://doi.org/10.3389/fmars.2016.00096.

Aylagas, E., Á. Borja, I. Muxika, and N. Rodríguez-Ezpeleta. 2018. "Adapting Metabarcoding-Based Benthic Biomonitoring into Routine Marine Ecological Status Assessment Networks." *Ecological Indicators* 95 (December): 194–202. https://doi.org/10.1016/j.ecolind.2018.07.044.

Aylagas, E., Á. Borja, and N. Rodríguez-Ezpeleta. 2014. "Environmental Status Assessment Using DNA Metabarcoding: Towards a Genetics Based Marine Biotic Index (GAMBI)." Edited by Consuegra, S. *PLoS ONE* 9 (3): e90529. https://doi.org/10.1371/journal.pone.0090529.

Azovsky, A.I. 2002. "Size-Dependent Species-Area Relationships in Benthos: Is the World More Diverse for Microbes?" *Ecography* 25 (3): 273–82. https://doi.org/10.1034/j.1600-0587.2002.250303.x.

Bakker, J., O.S. Wangensteen, D.D. Chapman, G. Boussarie, D. Buddo, T.L. Guttridge, H. Hertler, D. Mouillot, L. Vigliola, and S. Mariani. 2017. "Environmental DNA Reveals Tropical Shark Diversity and Abundance in Contrasting Levels of Anthropogenic Impact." *Scientific Reports* 7. https://doi.org/10.1038/s41598-017-17150-2.

Baldrian, P., M. Kolaiřík, M. Štursová, J. Kopecký, V. Valášková, T. Větrovský, L. Žifčáková, et al. 2012. "Active and Total Microbial Communities in Forest Soil Are Largely Different and Highly Stratified during Decomposition." *ISME Journal* 6 (2): 248–58. https://doi.org/10.1038/ismej.2011.95.

Baselga, A., and C.D.L. Orme. 2012. "Betapart : An R Package for the Study of Beta Diversity." *Methods in Ecology and Evolution* 3 (5): 808–12. https://doi.org/10.1111/j.2041-210X.2012.00224.x.

Bazin, E., S. Glémin, and N. Galtier. 2006. "Population Size Does Not Influence Mitochondrial Genetic Diversity in Animals." *Science* 312 (5773): 570–72. https://doi.org/10.1126/science.1122033.

Beaumont, M.A., and R.A. Nichols. 1996. "Evaluating Loci for Use in the Genetic Analysis of Population Structure." *Proceedings of the Royal Society of London. Series B: Biological Sciences* 263 (1377): 1619–26. https://doi.org/10.1098/rspb.1996.0237.

Bellec, L., M.-A.A. Cambon-Bonavita, V. Cueff-Gauchard, L. Durand, N. Gayet, and D. Zeppilli. 2018. "A Nematode of the Mid-Atlantic Ridge Hydrothermal Vents Harbors a Possible Symbiotic Relationship." *Frontiers in Microbiology* 9 (September): 2246. https://doi.org/10.3389/fmicb.2018.02246.

Bellemain, E., T. Carlsen, C. Brochmann, E. Coissac, P. Taberlet, and H. Kauserud. 2010. "ITS as an Environmental DNA Barcode for Fungi: An in Silico Approach Reveals Potential PCR Biases." *BMC Microbiology* 10 (July): 189. https://doi.org/10.1186/1471-2180-10-189.

Bensasson, D., D.X. Zhang, D.L. Hartl, and G.M. Hewitt. 2001. "Mitochondrial Pseudogenes: Evolution's Misplaced Witnesses." *Trends in Ecology and Evolution*. https://doi.org/10.1016/S0169-5347(01)02151-6.

Bernardino, A.F., L.A. Levin, A.R. Thurber, and C.R. Smith. 2012. "Comparative Composition, Diversity and Trophic Ecology of Sediment Macrofauna at Vents, Seeps and Organic Falls." *PLoS One* 7 (4): e33515. https://doi.org/10.1371/journal.pone.0033515.

Bhadury, P., M.C. Austen, D.T. Bilton, P.J.D. Lambshead, A.D. Rogers, and G.R. Smerdon. 2006. "Molecular Detection of Marine Nematodes from Environmental Samples: Overcoming Eukaryotic Interference." *Aquatic Microbial Ecology* 44 (1): 97–103. https://doi.org/Doi 10.3354/Ame044097.

Bianchelli, S., C. Gambi, M. Mea, A. Pusceddu, and R. Danovaro. 2013. "Nematode Diversity Patterns at Different Spatial Scales in Bathyal Sediments of the Mediterranean Sea." *Biogeosciences* 10 (8): 5465–79. https://doi.org/10.5194/bg-10-5465-2013.

Bianchelli, S., C. Gambi, D. Zeppilli, and R. Danovaro. 2010. "Metazoan Meiofauna in Deep-Sea Canyons and Adjacent Open Slopes: A Large-Scale Comparison with Focus on the Rare Taxa." *Deep-Sea Research Part I: Oceanographic Research Papers* 57 (3): 420–33. https://doi.org/10.1016/j.dsr.2009.12.001.

Bidartondo, M.I. 2008. "Preserving Accuracy in GenBank." *Science* 319 (5870): 1616a-1616a. https://doi.org/10.1126/science.319.5870.1616a.

Bienhold, C., P. Pop Ristova, F. Wenzhöfer, T. Dittmar, and A. Boetius. 2013. "How Deep-Sea Wood Falls Sustain Chemosynthetic Life." Edited by Kirchman, D.L. *PLoS ONE* 8 (1): e53590. https://doi.org/10.1371/journal.pone.0053590.

Bienhold, C., L. Zinger, A. Boetius, and A. Ramette. 2016. "Diversity and Biogeography of Bathyal and Abyssal Seafloor Bacteria." Edited by Kellogg, C.A. *PLoS ONE* 11 (1): e0148016. https://doi.org/10.1371/journal.pone.0148016.

Bik, H.M., D. Fournier, W. Sung, R.D. Bergeron, and W.K. Thomas. 2013. "Intra-Genomic Variation in the Ribosomal Repeats of Nematodes." Edited by Robinson-Rechavi, M. *PLoS ONE* 8 (10): e78230. https://doi.org/10.1371/journal.pone.0078230.

Bik, H.M., D.L. Porazinska, S. Creer, J.G. Caporaso, R. Knight, and W.K. Thomas. 2012a. "Sequencing Our Way towards Understanding Global Eukaryotic Biodiversity." *Trends in Ecology & Evolution* 27 (4): 233–43. https://doi.org/10.1016/j.tree.2011.11.010.

Bik, H.M., W. Sung, P. De Ley, J.G. Baldwin, J. Sharma, A. Rocha-Olivares, and W.K. Thomas. 2012b. "Metagenetic Community Analysis of Microbial Eukaryotes Illuminates Biogeographic Patterns in Deep-Sea and Shallow Water Sediments." *Molecular Ecology* 21 (5): 1048–59. https://doi.org/10.1111/j.1365-294X.2011.05297.x.

Bista, I., G. Carvalho, K. Walsh, M. Christmas, M. Hajibabaei, P. Kille, D. Lallias, and S. Creer. 2015. "Monitoring Lake Ecosystem Health Using Metabarcoding of Environmental DNA: Temporal Persistence and Ecological Relevance." *Genome* 58 (5): 197.

Bista, I., G.R. Carvalho, M. Tang, K. Walsh, X. Zhou, M. Hajibabaei, S. Shokralla, et al. 2018. "Performance of Amplicon and Shotgun Sequencing for Accurate Biomass Estimation in Invertebrate Community Samples." *Molecular Ecology Resources* 18 (5): 1020–34. https://doi.org/10.1111/1755-0998.12888.

Blankenship-Williams, L.E., and L.A. Levin. 2009. "Living Deep: A Synopsis of Hadal Trench Ecology." *Marine Technology Society Journal* 43 (5): 137–43. https://doi.org/10.4031/MTSJ.43.5.23.

Blaxter, M. 2003. “Molecular Systematics: Counting Angels with DNA.” *Nature* 421 (6919): 122–24. https://doi.org/10.1038/421122a.

———. 2016. “Imagining Sisyphus Happy: DNA Barcoding and the Unnamed Majority.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 371 (1702). https://doi.org/10.1098/rstb.2015.0329.

Blazewicz, S.J., R.L. Barnard, R.A. Daly, and M.K. Firestone. 2013. “Evaluating RRNA as an Indicator of Microbial Activity in Environmental Communities: Limitations and Uses.” *ISME Journal* 7 (11): 2061–68. https://doi.org/10.1038/ismej.2013.102.

Bodil, B.A., W.G. Ambrose, M. Bergmann, L.M. Clough, A. V. Gebruk, C. Hasemann, K. Iken, et al. 2011. “Diversity of the Arctic Deep-Sea Benthos.” *Marine Biodiversity* 41 (1): 87–107. https://doi.org/10.1007/s12526-010-0078-4.

Boeckner, M.J., J. Sharma, and H.C. Proctor. 2009. “Revisiting the Meiofauna Paradox: Dispersal and Colonization of Nematodes and Other Meiofaunal Organisms in Low- and High-Energy Environments.” *Hydrobiologia* 624 (1): 91–106. https://doi.org/10.1007/s10750-008-9669-5.

Boere, A.C., W.I.C. Rijpstra, G.J. De Lange, J.S. Sinninghe Damsté, and M.J.L. Coolen. 2011. “Preservation Potential of Ancient Plankton DNA in Pleistocene Marine Sediments.” *Geobiology* 9 (5): 377–93. https://doi.org/10.1111/j.1472-4669.2011.00290.x.

Bokulich, N.A., S. Subramanian, J.J. Faith, D. Gevers, J.I. Gordon, R. Knight, D.A. Mills, and J.G. Caporaso. 2013. “Quality-Filtering Vastly Improves Diversity Estimates from Illumina Amplicon Sequencing.” *Nature Methods* 10 (1): 57–59. https://doi.org/10.1038/nmeth.2276.

Boussarie, G., J. Bakker, O.S. Wangensteen, S. Mariani, L. Bonnin, J.B. Juhel, J.J. Kiszka, et al. 2018. “Environmental DNA Illuminates the Dark Diversity of Sharks.” *Science Advances* 4 (5): eaap9661. https://doi.org/10.1126/sciadv.aap9661.

Boyer, F., C. Mercier, A. Bonin, Y. Le Bras, P. Taberlet, and E. Coissac. 2016. “OBITOOLS: A UNIX-Inspired Software Package for DNA Metabarcoding.” *Molecular Ecology Resources* 16 (1): 176–82. https://doi.org/10.1111/1755-0998.12428.

Brandt, A., C. De Broyer, I. De Mesel, K.E. Ellingsen, A.J. Gooday, B. Hilbig, K. Linse, M.R.A. Thomson, and P.A. Tyler. 2007a. “The Biodiversity of the Deep Southern Ocean Benthos.” *Philosophical Transactions of the Royal Society B-Biological Sciences* 362 (1477): 39–66. https://doi.org/DOI 10.1098/rstb.2006.1952.

Brandt, A., and B. Ebbe. 2009. “Southern Ocean Deep-Sea Biodiversity-From Patterns to Processes.” *Deep-Sea Research Part II: Topical Studies in Oceanography* 56 (19–20): 1732–38. https://doi.org/10.1016/j.dsr2.2009.05.017.

Brandt, A., A.J. Gooday, S.N. Brandão, S. Brix, W. Brökeland, T. Cedhagen, M. Choudhury, et al. 2007b. “First Insights into the Biodiversity and Biogeography of the Southern Ocean Deep Sea.” *Nature* 447 (7142): 307–11. https://doi.org/10.1038/nature05827.

Brandt, M.I., B. Trouche, L. Quintric, P. Wincker, J. Poulain, and S. Arnaud-Haond. 2020. “A Flexible Pipeline Combining Bioinformatic Correction Tools for Prokaryotic and Eukaryotic Metabarcoding.” *BioRxiv 717355, Ver. 3 Peer-Reviewed and Recommended by PCI Ecology*. https://doi.org/10.1101/717355.

Brannock, P.M., and K.M. Halanych. 2015. “Meiofaunal Community Analysis by High-Throughput Sequencing: Comparison of Extraction, Quality Filtering, and Clustering Methods.” *Marine Genomics* 23: 67–75. https://doi.org/10.1016/j.margen.2015.05.007.

Brown, E.A., F.J.J. Chain, T.J. Crease, H.J. Macisaac, and M.E. Cristescu. 2015. “Divergence Thresholds and Divergent Biodiversity Estimates: Can Metabarcoding Reliably Describe Zooplankton Communities?” *Ecology and Evolution* 5 (11): 2234–51. https://doi.org/10.1002/ece3.1485.

Bucklin, A., D. Steinke, and L. Blanco-Bercial. 2011. "DNA Barcoding of Marine Metazoa." *Annual Review of Marine Science* 3 (1): 471–508. https://doi.org/10.1146/annurev-marine-120308-080950.

Buhl-Mortensen, L., A. Vanreusel, A.J. Gooday, L.A. Levin, I.G. Priede, P. Buhl-Mortensen, H. Gheerardyn, N.J. King, and M. Raes. 2010. "Biological Structures as a Source of Habitat Heterogeneity and Biodiversity on the Deep Ocean Margins." *Marine Ecology* 31 (1): 21–50. https://doi.org/10.1111/j.1439-0485.2010.00359.x.

Burgess, R. 2001. "An Improved Protocol for Separating Meiofauna from Sediments Using Colloidal Silica Sols." *Marine Ecology Progress Series* 214 (Iler 1979): 161–65. https://doi.org/10.3354/meps214161.

Butchart, S.H.M., M. Walpole, B. Collen, A. Van Strien, J.P.W. Scharlemann, R.E.A. Almond, J.E.M. Baillie, et al. 2010. "Global Biodiversity: Indicators of Recent Declines." *Science* 328 (5982): 1164–68. https://doi.org/10.1126/science.1187512.

Caley, M.J., R. Fisher, and K. Mengersen. 2014. "Global Species Richness Estimates Have Not Converged." *Trends in Ecology & Evolution* 29 (4): 187–88. https://doi.org/10.1016/j.tree.2014.02.002.

Callahan, B.J., P.J. McMurdie, and S.P. Holmes. 2017. "Exact Sequence Variants Should Replace Operational Taxonomic Units in Marker-Gene Data Analysis." *ISME Journal* 11 (12): 2639–43. https://doi.org/10.1038/ismej.2017.119.

Callahan, B.J., P.J. McMurdie, M.J. Rosen, A.W. Han, A.J.A. Johnson, and S.P. Holmes. 2016. "DADA2: High-Resolution Sample Inference from Illumina Amplicon Data." *Nature Methods* 13 (7): 581–83. https://doi.org/10.1038/nmeth.3869.

Candek, K., and M. Kuntner. 2015. "DNA Barcoding Gap: Reliable Species Identification over Morphological and Geographical Scales." *Molecular Ecology Resources* 15 (2): 268–77. https://doi.org/10.1111/1755-0998.12304.

Cantera, I., K. Cilleros, A. Valentini, A. Cerdan, T. Dejean, A. Iribar, P. Taberlet, R. Vigouroux, and S. Brosse. 2019. "Optimizing Environmental DNA Sampling Effort for Fish Inventories in Tropical Streams and Rivers." *Scientific Reports* 9 (1): 1–11. https://doi.org/10.1038/s41598-019-39399-5.

Cardinale, B.J., J.E. Duffy, A. Gonzalez, D.U. Hooper, C. Perrings, P. Venail, A. Narwani, et al. 2012. "Biodiversity Loss and Its Impact on Humanity." *Nature* 486 (7401): 59–67. https://doi.org/10.1038/nature11148.

Carranza, S., G. Giribet, C. Ribera, J. Baguñà, and M. Riutort. 1996. "Evidence That Two Types of 18S RDNA Coexist in the Genome of Dugesia (Schmidtea) Mediterranea (Platyhelminthes, Turbellaria, Tricladida)." *Molecular Biology and Evolution* 13 (6): 824–32. https://doi.org/10.1093/oxfordjournals.molbev.a025643.

Carugati, L., C. Corinaldesi, A. Dell'Anno, and R. Danovaro. 2015. "Metagenetic Tools for the Census of Marine Meiofaunal Biodiversity: An Overview." *Marine Genomics* 24: 11–20. https://doi.org/10.1016/j.margen.2015.04.010.

Castro-Wallace, S.L., C.Y. Chiu, K.K. John, S.E. Stahl, K.H. Rubins, A.B.R. McIntyre, J.P. Dworkin, et al. 2017. "Nanopore DNA Sequencing and Genome Assembly on the International Space Station." *Scientific Reports* 7 (1): 1–12. https://doi.org/10.1038/s41598-017-18364-0.

Catarino, D., S. Stefanni, P.E. Jorde, G.M. Menezes, J.B. Company, F. Neat, and H. Knutsen. 2017. "The Role of the Strait of Gibraltar in Shaping the Genetic Structure of the Mediterranean Grenadier, Coryphaenoides Mediterraneus, between the Atlantic and Mediterranean Sea." Edited by Chiang, T.-Y. *PLoS ONE* 12 (5): 1–24. https://doi.org/10.1371/journal.pone.0174988.

Chao, A., and C.-H. Chiu. 2016. "Species Richness: Estimation and Comparison." *Statistics*

Reference Online, 1–11. https://doi.org/10.1002/9780470015902.a0026329.

Chao, A., N.J. Gotelli, T.C. Hsieh, E.L. Sander, K.H. Ma, R.K. Colwell, and A.M. Ellison. 2014. "Rarefaction and Extrapolation with Hill Numbers: A Framework for Sampling and Estimation in Species Diversity Studies." *Ecological Monographs* 84 (1): 45–67. https://doi.org/10.1890/13-0133.1.

Chariton, A.A., S. Stephenson, M.J. Morgan, A.D.L. Steven, M.J. Colloff, L.N. Court, and C.M. Hardy. 2015. "Metabarcoding of Benthic Eukaryote Communities Predicts the Ecological Condition of Estuaries." *Environmental Pollution* 203: 165–74. https://doi.org/10.1016/j.envpol.2015.03.047.

Chaudhary, C., H. Saeedi, and M.J. Costello. 2016. "Bimodality of Latitudinal Gradients in Marine Species Richness." *Trends in Ecology and Evolution*. https://doi.org/10.1016/j.tree.2016.06.001.

———. 2017. "Marine Species Richness Is Bimodal with Latitude: A Reply to Fernandez and Marques." *Trends in Ecology and Evolution* 32 (4): 234–37. https://doi.org/10.1016/j.tree.2017.02.007.

Chisholm, R.A., M.A. Burgman, S.P. Hubbell, and L. Borda-De-??gua. 2004. "The Unified Neutral Theory of Biodiversity and Biogeography: Comment." *Ecology*. Princeton University Press. http://www.citeulike.org/group/894/article/522578.

Clare, E.L., F.J.J. Chain, J.E. Littlefair, and M.E. Cristescu. 2016. "The Effects of Parameter Choice on Defining Molecular Operational Taxonomic Units and Resulting Ecological Analyses of Metabarcoding Data." Edited by Deiner, K. *Genome* 59 (11): 981–90. https://doi.org/10.1139/gen-2015-0184.

Clark, M.R., A.A. Rowden, T. Schlacher, A. Williams, M. Consalvey, K.I. Stocks, A.D. Rogers, et al. 2010. "The Ecology of Seamounts: Structure, Function, and Human Impacts." *Annual Review of Marine Science* 2 (1): 253–78. https://doi.org/10.1146/annurev-marine-120308-081109.

Clarke, L.J., J.M. Beard, K.M. Swadling, and B.E. Deagle. 2017. "Effect of Marker Choice and Thermal Cycling Protocol on Zooplankton DNA Metabarcoding Studies." *Ecology and Evolution* 7 (3): 873–83. https://doi.org/10.1002/ece3.2667.

Cohan, F.M. 2001. "Bacterial Species and Speciation." Edited by Kane, M. *Systematic Biology* 50 (4): 513–24. https://doi.org/10.1080/10635150118398.

———. 2002. "What Are Bacterial Species?" *Annual Review of Microbiology* 56 (1): 457–87. https://doi.org/10.1146/annurev.micro.56.012302.160634.

Coissac, E., T. Riaz, and N. Puillandre. 2012. "Bioinformatic Challenges for DNA Metabarcoding of Plants and Animals." *Molecular Ecology* 21 (8): 1834–47. https://doi.org/10.1111/j.1365-294X.2012.05550.x.

Colegrave, N., and G.D. Ruxton. 2018. "Using Biological Insight and Pragmatism When Thinking about Pseudoreplication." *Trends in Ecology and Evolution* 33 (1): 28–35. https://doi.org/10.1016/j.tree.2017.10.007.

Collins, R.A., O.S. Wangensteen, E.J. O'Gorman, S. Mariani, D.W. Sims, and M.J. Genner. 2018. "Persistence of Environmental DNA in Marine Systems." *Communications Biology* 1 (1): 185. https://doi.org/10.1038/s42003-018-0192-6.

Conway, J.R., A. Lex, and N. Gehlenborg. 2017. "UpSetR: An R Package for the Visualization of Intersecting Sets and Their Properties." Edited by Hancock, J. *Bioinformatics* 33 (18): 2938–40. https://doi.org/10.1093/bioinformatics/btx364.

Coolen, M.J.L., W.D. Orsi, C. Balkema, C. Quince, K. Harris, S.P. Sylva, M. Filipova-Marinova, and L. Giosan. 2013. "Evolution of the Plankton Paleome in the Black Sea from the Deglacial to Anthropocene." *Proceedings of the National Academy of Sciences of the United States of America* 110 (21): 8609–14. https://doi.org/10.1073/pnas.1219283110.

Cordier, T., P. Esling, F. Lejzerowicz, J. Visco, A. Ouadahi, C. Martins, T. Cedhagen, and J.W. Pawlowski. 2017. "Predicting the Ecological Quality Status of Marine Environments from EDNA Metabarcoding Data Using Supervised Machine Learning." *Environmental Science and Technology* 51 (16): 9118–26. https://doi.org/10.1021/acs.est.7b01518.

Cordier, T., D. Forster, Y. Dufresne, C.I.M. Martins, T. Stoeck, and J.W. Pawlowski. 2018. "Supervised Machine Learning Outperforms Taxonomy-Based Environmental DNA Metabarcoding Applied to Biomonitoring." *Molecular Ecology Resources* 18 (6): 1381–91. https://doi.org/10.1111/1755-0998.12926.

Cordier, T., F. Frontalini, K. Cermakova, L. Apothéloz-Perret-Gentil, M. Treglia, E. Scantamburlo, V. Bonamin, and J.W. Pawlowski. 2019a. "Multi-Marker EDNA Metabarcoding Survey to Assess the Environmental Impact of Three Offshore Gas Platforms in the North Adriatic Sea (Italy)." *Marine Environmental Research* 146: 24–34. https://doi.org/10.1016/j.marenvres.2018.12.009.

Cordier, T., A. Lanzén, L. Apothéloz-Perret-Gentil, T. Stoeck, and J.W. Pawlowski. 2019b. "Embracing Environmental Genomics and Machine Learning for Routine Biomonitoring." *Trends in Microbiology*. https://doi.org/10.1016/j.tim.2018.10.012.

Cordier, T., and J.W. Pawlowski. 2018. "BBI: An R Package for the Computation of Benthic Biotic Indices from Composition Data." *Metabarcoding and Metagenomics* 2 (July): e25649. https://doi.org/10.3897/mbmg.2.25649.

Corinaldesi, C., M. Tangherlini, E. Manea, and A. Dell'Anno. 2018. "Extracellular DNA as a Genetic Recorder of Microbial Diversity in Benthic Deep-Sea Ecosystems." *Scientific Reports* 8 (1): 1839. https://doi.org/10.1038/s41598-018-20302-7.

Costello, M.J. 2009. "Distinguishing Marine Habitat Classification Concepts for Ecological Data Management." *Marine Ecology Progress Series* 397: 253–68. https://doi.org/10.3354/meps08317.

Costello, M.J., Z. Basher, R. Sayre, S. Breyer, and D.J. Wright. 2018. "Stratifying Ocean Sampling Globally and with Depth to Account for Environmental Variability." *Scientific Reports* 8 (1): 1–9. https://doi.org/10.1038/s41598-018-29419-1.

Costello, M.J., and S. Breyer. 2017. "Ocean Depths: The Mesopelagic and Implications for Global Warming." *Current Biology*. Cell Press. https://doi.org/10.1016/j.cub.2016.11.042.

Costello, M.J., and C. Chaudhary. 2017. "Marine Biodiversity, Biogeography, Deep-Sea Gradients, and Conservation." *Current Biology* 27 (11): R511–27. https://doi.org/10.1016/j.cub.2017.04.060.

Costello, M.J., A. Cheung, and N. De Hauwere. 2010a. "Surface Area and the Seabed Area, Volume, Depth, Slope, and Topographic Variation for the World's Seas, Oceans, and Countries." *Environmental Science and Technology* 44 (23): 8821–28. https://doi.org/10.1021/es1012752.

Costello, M.J., M. Coll, R. Danovaro, P. Halpin, H. Ojaveer, and P. Miloslavich. 2010b. "A Census of Marine Biodiversity Knowledge, Resources, and Future Challenges." Edited by Humphries, S. *PLoS ONE* 5 (8): e12110. https://doi.org/10.1371/journal.pone.0012110.

Costello, M.J., P. Tsai, P.S. Wong, A.K.L. Cheung, Z. Basher, and C. Chaudhary. 2017. "Marine Biogeographic Realms and Species Endemicity." *Nature Communications* 8 (1): 1057. https://doi.org/10.1038/s41467-017-01121-2.

Costello, M.J., S. Wilson, and B. Houlding. 2012. "Predicting Total Global Species Richness Using Rates of Species Description and Estimates of Taxonomic Effort." *Systematic Biology* 61 (5): 871–83. https://doi.org/10.1093/sysbio/syr080.

Cowart, D.A., M. Pinheiro, O. Mouchel, M. Maguer, J. Grall, J. Miné, and S. Arnaud-Haond. 2015. "Metabarcoding Is Powerful yet Still Blind: A Comparative Analysis of Morphological and Molecular Surveys of Seagrass Communities." *PLoS One* 10 (2):

e0117562. https://doi.org/10.1371/journal.pone.0117562.

Creer, S., K. Deiner, S. Frey, D. Porazinska, P. Taberlet, W.K. Thomas, C. Potter, and H.M. Bik. 2016. "The Ecologist's Field Guide to Sequence-Based Identification of Biodiversity." Edited by Freckleton, R. *Methods in Ecology and Evolution* 7 (9): 1008–18. https://doi.org/10.1111/2041-210X.12574.

Cristescu, M.E. 2014. "From Barcoding Single Individuals to Metabarcoding Biological Communities: Towards an Integrative Approach to the Study of Global Biodiversity." *Trends in Ecology & Evolution* 29 (10): 566–71. https://doi.org/10.1016/j.tree.2014.08.001.

———. 2019. "Can Environmental RNA Revolutionize Biodiversity Science?" *Trends in Ecology and Evolution*. https://doi.org/10.1016/j.tree.2019.05.003.

Cristescu, M.E., and P.D.N. Hebert. 2018. "Uses and Misuses of Environmental DNA in Biodiversity Science and Conservation." *Rev. Ecol. Evol. Syst* 49 (July): 209–39. https://doi.org/10.1146/annurev-ecolsys-110617-062306.

Cruaud, P., A. Vigneron, C. Lucchetti-Miganeh, P.E. Ciron, A. Godfroy, and M.A. Cambon-Bonavita. 2014. "Influence of DNA Extraction Method, 16S RRNA Targeted Hypervariable Regions, and Sample Origin on Microbial Diversity Detected by 454 Pyrosequencing in Marine Chemosynthetic Ecosystems." *Applied and Environmental Microbiology* 80 (15): 4626–39. https://doi.org/10.1128/AEM.00592-14.

Cruz-Dávalos, D.I., B. Llamas, C. Gaunitz, A. Fages, C. Gamba, J. Soubrier, P. Librado, et al. 2017. "Experimental Conditions Improving In-Solution Target Enrichment for Ancient DNA." *Molecular Ecology Resources* 17 (3): 508–22. https://doi.org/10.1111/1755-0998.12595.

Curini-Galletti, M., T. Artois, V. Delogu, W.H. de Smet, D. Fontaneto, U. Jondelius, F. Leasi, et al. 2012. "Patterns of Diversity in Soft-Bodied Meiofauna: Dispersal Ability and Body Size Matter." Edited by Archambault, P. *PLoS ONE* 7 (3): e33801. https://doi.org/10.1371/journal.pone.0033801.

Danchin, É., A. Charmantier, F.A. Champagne, A. Mesoudi, B. Pujol, and S. Blanchet. 2011. "Beyond DNA: Integrating Inclusive Inheritance into an Extended Theory of Evolution." *Nature Reviews Genetics* 12 (7): 475–86. https://doi.org/10.1038/nrg3028.

Danovaro, R., S. Bianchelli, C. Gambi, M. Mea, and D. Zeppilli. 2009a. "α-, β-, γ-, δ-and ε-Diversity of Deep-Sea Nematodes in Canyons and Open Slopes of Northeast Atlantic and Mediterranean Margins." *Marine Ecology Progress Series* 396: 197–209. https://doi.org/10.3354/meps08269.

Danovaro, R., M. Canals, C. Gambi, S. Heussner, N. Lampadariou, and A. Vanreusel. 2009b. "Exploring Benthic Biodiversity Patterns and Hot Spots on European Margin Slopes." *Oceanography* 22 (1): 16–25. https://doi.org/10.5670/oceanog.2009.02.

Danovaro, R., L. Carugati, M. Berzano, A.E. Cahill, S. Carvalho, A. Chenuil, C. Corinaldesi, et al. 2016. "Implementing and Innovating Marine Monitoring Approaches for Assessing Marine Environmental Status." *Frontiers in Marine Science* 3. https://doi.org/10.3389/fmars.2016.00213.

Danovaro, R., L. Carugati, C. Corinaldesi, C. Gambi, K. Guilini, A. Pusceddu, and A. Vanreusel. 2013. "Multiple Spatial Scale Analyses Provide New Clues on Patterns and Drivers of Deep-Sea Nematode Diversity." *Deep-Sea Research Part II: Topical Studies in Oceanography* 92: 97–106. https://doi.org/10.1016/j.dsr2.2013.03.035.

Danovaro, R., J.B. Company, C. Corinaldesi, G. D'Onghia, B. Galil, C. Gambi, A.J. Gooday, et al. 2010. "Deep-Sea Biodiversity in the Mediterranean Sea: The Known, the Unknown, and the Unknowable." Edited by Gratwicke, B. *PLoS ONE* 5 (8): e11832. https://doi.org/10.1371/journal.pone.0011832.

Danovaro, R., C. Corinaldesi, A. Dell'Anno, and P.V.R. Snelgrove. 2017. "The Deep-Sea under Global Change." *Current Biology*. https://doi.org/10.1016/j.cub.2017.02.046.

Danovaro, R., M. Fabiano, G. Albertelli, and N. Della Croce. 1995. "Vertical Distribution of Meiobenthos in Bathyal Sediments of the Eastern Mediterranean Sea: Relationship with Labile Organic Matter and Bacterial Biomasses." *Marine Ecology* 16 (2): 103–16. https://doi.org/10.1111/j.1439-0485.1995.tb00398.x.

Danovaro, R., P.V.R. Snelgrove, and P. Tyler. 2014. "Challenging the Paradigms of Deep-Sea Ecology." *Trends in Ecology and Evolution*. https://doi.org/10.1016/j.tree.2014.06.002.

Davis, N.M., D.M. Proctor, S.P. Holmes, D.A. Relman, and B.J. Callahan. 2018. "Simple Statistical Identification and Removal of Contaminant Sequences in Marker-Gene and Metagenomics Data." *Microbiome* 6 (1): 226. https://doi.org/10.1186/s40168-018-0605-2.

Deagle, B.E., S.N. Jarman, E. Coissac, F. Pompanon, and P. Taberlet. 2014. "DNA Metabarcoding and the Cytochrome c Oxidase Subunit I Marker: Not a Perfect Match." *Biology Letters* 10 (9): 20140562–20140562. https://doi.org/10.1098/rsbl.2014.0562.

Deiner, K., E.A. Fronhofer, E. Mächler, J.C. Walser, and F. Altermatt. 2016. "Environmental DNA Reveals That Rivers Are Conveyer Belts of Biodiversity Information." *Nature Communications* 7 (1): 12544. https://doi.org/10.1038/ncomms12544.

Deiner, K., J.-C.C. Walser, E. Machler, F. Altermatt, E. Mächler, and F. Altermatt. 2015. "Choice of Capture and Extraction Methods Affect Detection of Freshwater Biodiversity from Environmental DNA." *Biological Conservation* 183: 53–63. https://doi.org/10.1016/j.biocon.2014.11.018.

Dejean, T., A. Valentini, A. Duparc, S. Pellier-Cuit, F. Pompanon, P. Taberlet, and C. Miaud. 2011. "Persistence of Environmental DNA in Freshwater Ecosystems." *PLoS One* 6 (8). https://doi.org/10.1371/journal.pone.0023398.

Dell'Anno, A., and R. Danovaro. 2005. "Ecology: Extracellular DNA Plays a Key Role in Deep-Sea Ecosystem Functioning." *Science* 309 (5744): 2179. https://doi.org/10.1126/science.1117475.

Desbruyeres, D., M. Segonzac, M. Bright, and A. Decapoda. 2006. "HANDBOOK OF DEEP-SEA HYDROTHERMAL VENT FAUNA." Edited by Desbruyeres, D., Segonzac, M., and Bright, M. *Marine Ecology* 27 (3): 271–271. https://doi.org/10.1111/j.1439-0485.2006.00107.x.

Dickie, I.A., S. Boyer, H.L. Buckley, R.P. Duncan, P.P. Gardner, I.D. Hogg, R.J. Holdaway, et al. 2018. "Towards Robust and Repeatable Sampling Methods in EDNA-Based Studies." *Molecular Ecology Resources*. Wiley/Blackwell (10.1111). https://doi.org/10.1111/1755-0998.12907.

Díez-Vives, C., S. Nielsen, P. Sánchez, O. Palenzuela, I. Ferrera, M. Sebastián, C. Pedrós-Alió, J.M. Gasol, and S.G. Acinas. 2019. "Delineation of Ecologically Distinct Units of Marine Bacteroidetes in the Northwestern Mediterranean Sea." *Molecular Ecology* 28 (11): 2846–59. https://doi.org/10.1111/mec.15068.

Djurhuus, A., J. Port, C.J. Closek, K.M. Yamahara, O.C. Romero-Maraccini, K.R. Walz, D.B. Goldsmith, et al. 2017. "Evaluation of Filtration and DNA Extraction Methods for Environmental DNA Biodiversity Assessments across Multiple Trophic Levels." *Frontiers in Marine Science* 4 (October): 314. https://doi.org/10.3389/fmars.2017.00314.

Dopheide, A., D. Xie, T.R. Buckley, A.J. Drummond, and R.D. Newcomb. 2019. "Impacts of DNA Extraction and PCR on DNA Metabarcoding Estimates of Soil Biodiversity." Edited by Bunce, M. *Methods in Ecology and Evolution* 10 (1): 120–33. https://doi.org/10.1111/2041-210X.13086.

Dover, C.L. Van, C.R. German, K.G. Speer, L.M. Parson, and R.C. Vrijenhoek. 2002.

"Evolution and Biogeography of Deep-Sea Vent and Seep Invertebrates." *Science* 295 (2002): 1253–57. https://doi.org/10.1126/science.1067361.

Dover, C.L. Van, S.E. Humphris, D. Fornari, C.M. Cavanaugh, R. Collier, S.K. Goffredi, J. Hashimoto, et al. 2001. "Biogeography and Ecological Setting of Indian Ocean Hydrothermal Vents." *Science* 294 (5543): 818–23. https://doi.org/10.1126/science.1064574.

Dray, S., and A.B. Dufour. 2007. "The Ade4 Package: Implementing the Duality Diagram for Ecologists." *Journal of Statistical Software* 22 (4): 1–20. https://doi.org/10.18637/jss.v022.i04.

Drummond, A.J., R.D. Newcomb, T.R. Buckley, D. Xie, A. Dopheide, B.C.M. Potter, J. Heled, et al. 2015. "Evaluating a Multigene Environmental DNA Approach for Biodiversity Assessment." *Gigascience* 4. https://doi.org/ARTN 4610.1186/s13742-015-0086-1.

Dubilier, N., C. Bergin, and C. Lott. 2008. "Symbiotic Diversity in Marine Animals: The Art of Harnessing Chemosynthesis." *Nature Reviews Microbiology* 6 (10): 725–40. https://doi.org/10.1038/nrmicro1992.

Dufresne, Y., F. Lejzerowicz, L.A. Perret-Gentil, J.W. Pawlowski, and T. Cordier. 2019. "SLIM: A Flexible Web Application for the Reproducible Processing of Environmental DNA Metabarcoding Data." *BMC Bioinformatics* 20 (1): 88. https://doi.org/10.1186/s12859-019-2663-2.

Duranton, M., F. Allal, C. Fraïsse, N. Bierne, F. Bonhomme, and P.A. Gagnaire. 2018. "The Origin and Remolding of Genomic Islands of Differentiation in the European Sea Bass." *Nature Communications* 9 (1): 1–11. https://doi.org/10.1038/s41467-018-04963-6.

Duranton, M., F. Bonhomme, and P. Gagnaire. 2019. "The Spatial Scale of Dispersal Revealed by Admixture Tracts." *Evolutionary Applications* 12 (9): 1743–56. https://doi.org/10.1111/eva.12829.

Easton, E.E., and D. Thistle. 2016. "Do Some Deep-Sea, Sediment-Dwelling Species of Harpacticoid Copepods Have 1000-Km-Scale Range Sizes?" *Molecular Ecology* 25 (17): 4301–18. https://doi.org/10.1111/mec.13744.

Edgar, R.C. 2010. "Search and Clustering Orders of Magnitude Faster than BLAST." *Bioinformatics* 26 (19): 2460–61. https://doi.org/10.1093/bioinformatics/btq461.

———. 2013. "UPARSE: Highly Accurate OTU Sequences from Microbial Amplicon Reads." *Nature Methods* 10 (10): 996–98. https://doi.org/10.1038/nmeth.2604.

———. 2016a. "SINTAX: A Simple Non-Bayesian Taxonomy Classifier for 16S and ITS Sequences." *BioRxiv [Preprint]*, 074161. https://doi.org/10.1101/074161.

———. 2016b. "UNCROSS: Filtering of High-Frequency Cross-Talk in 16S Amplicon Reads," 088666. https://doi.org/10.1101/088666.

———. 2016c. "UNOISE2: Improved Error-Correction for Illumina 16S and ITS Amplicon Sequencing." *BioRxiv [Preprint]*, October, 081257. https://doi.org/10.1101/081257.

———. 2017a. "SINAPS: Prediction of Microbial Traits from Marker Gene Sequences." *BioRxiv [Preprint]*, 124156. https://doi.org/10.1101/124156.

———. 2017b. "UNBIAS: An Attempt to Correct Abundance Bias in 16S Sequencing, with Limited Success." *BioRxiv [Preprint]*, April, 124149. https://doi.org/10.1101/124149.

———. 2018a. "UNCROSS2: Identification of Cross-Talk in 16S RRNA OTU Tables." *BioRxiv [Preprint]*, 400762. https://doi.org/10.1101/400762.

———. 2018b. "Accuracy of Taxonomy Prediction for 16S RRNA and Fungal ITS Sequences." *PeerJ* 2018 (4): e4652. https://doi.org/10.7717/peerj.4652.

———. 2018c. "Taxonomy Annotation and Guide Tree Errors in 16S RRNA Databases." *PeerJ* 2018 (6): e5030. https://doi.org/10.7717/peerj.5030.

———. 2018d. "Updating the 97% Identity Threshold for 16S Ribosomal RNA OTUs." Edited

by Valencia, A. *Bioinformatics* 34 (14): 2371–75. https://doi.org/10.1093/bioinformatics/bty113.

Edgar, R.C., and H. Flyvbjerg. 2018. "Octave Plots for Visualizing Diversity of Microbial OTUs." *BioRxiv [Preprint]*, August, 389833. https://doi.org/10.1101/389833.

Edgar, R.C., B.J. Haas, J.C. Clemente, C. Quince, and R. Knight. 2011. "UCHIME Improves Sensitivity and Speed of Chimera Detection." *Bioinformatics* 27 (16): 2194–2200. https://doi.org/10.1093/bioinformatics/btr381.

Elbrecht, V., and F. Leese. 2015. "Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias and Biomass-Sequence Relationships with an Innovative Metabarcoding Protocol." Edited by Hajibabaei, M. *PLoS ONE* 10 (7): e0130324. https://doi.org/10.1371/journal.pone.0130324.

Elbrecht, V., B. Peinert, and F. Leese. 2017. "Sorting Things out: Assessing Effects of Unequal Specimen Biomass on DNA Metabarcoding." *Ecology and Evolution* 7 (17): 6918–26. https://doi.org/10.1002/ece3.3192.

Elbrecht, V., E.E. Vamos, D. Steinke, and F. Leese. 2018. "Estimating Intraspecific Genetic Diversity from Community DNA Metabarcoding Data." *PeerJ*, no. 4 (April). https://doi.org/10.7717/peerj.4644.

Emmerson, M.C., M. Solan, C. Emes, D.M. Paterson, and D. Raffaelli. 2001. "Consistent Patterns and the Idiosyncratic Effects of Biodiversity in Marine Ecosystems." *Nature* 411 (6833): 73–77. https://doi.org/10.1038/35075055.

Eren, A.M., J.H. Vineis, H.G. Morrison, and M.L. Sogin. 2013. "A Filtering Method to Generate High Quality Short Reads Using Illumina Paired-End Technology." *PLoS ONE* 8 (6): e66643. https://doi.org/10.1371/journal.pone.0066643.

Escudié, F., L. Auer, M. Bernard, M. Mariadassou, L. Cauquil, K. Vidal, S. Maman, et al. 2018. "FROGS: Find, Rapidly, OTUs with Galaxy Solution." Edited by Berger, B. *Bioinformatics* 34 (8): 1287–94. https://doi.org/10.1093/bioinformatics/btx791.

Etnoyer, P.J., J. Wood, and T.C. Shirley. 2010. "How Large Is the Seamount Biome?" *Oceanography* 23 (1): 206–9. https://doi.org/10.5670/oceanog.2010.96.

Etter, R.J., and J.F. Grassle. 1992. "Patterns of Species Diversity in the Deep Sea as a Function of Sediment Particle Size Diversity." *Nature* 360 (6404): 576–78. https://doi.org/10.1038/360576a0.

Evans, N T, B.P. Olds, M.A. Renshaw, C.R. Turner, Y.Y. Li, C.L. Jerde, A.R. Mahon, M.E. Pfrender, G.A. Lamberti, and D.M. Lodge. 2016. "Quantification of Mesocosm Fish and Amphibian Species Diversity via Environmental DNA Metabarcoding." *Molecular Ecology Resources* 16 (1): 29–41. https://doi.org/10.1111/1755-0998.12433.

Evans, Nathan T., Y. Li, M.A. Renshaw, B.P. Olds, K. Deiner, C.R. Turner, C.L. Jerde, D.M. Lodge, G.A. Lamberti, and M.E. Pfrender. 2017. "Fish Community Assessment with EDNA Metabarcoding: Effects of Sampling Design and Bioinformatic Filtering." *Canadian Journal of Fisheries and Aquatic Sciences* 74 (9): 1362–74. https://doi.org/10.1139/cjfas-2016-0306.

Fenchel, T. 2005. "Where Are All the Species?" *Environmental Microbiology* 7 (4): 473–75. https://doi.org/10.1111/j.1462-2920.2005.803_3.x.

Fernández, C.A., P. Arribas, F. Ruzicka, A. Crampton-Platt, M.J.T.N. Timmermans, and A.P. Vogler. 2015. "Phylogenetic Community Ecology of Soil Biodiversity Using Mitochondrial Metagenomics." *Molecular Ecology* 24 (14): 3603–17. https://doi.org/10.1111/mec.13195.

Ficetola, G.F., J. Pansu, A. Bonin, E. Coissac, C. Giguet-Covex, M. De Barba, L. Gielly, et al. 2015. "Replication Levels, False Presences and the Estimation of the Presence/Absence from EDNA Metabarcoding Data." *Molecular Ecology Resources* 15 (3): 543–56.

https://doi.org/10.1111/1755-0998.12338.

Finlay, B.J. 2002. "Global Dispersal of Free-Living Microbial Eukaryote Species." *Science*. https://doi.org/10.1126/science.1070710.

Finlay, B.J., G.F. Esteban, and T. o. m. Fenchel. 2004. "Protist Diversity Is Different?" *Protist* 155 (1): 15–22. https://doi.org/10.1078/1434461000160.

Fisher, C.R., P.A. Montagna, and T.T. Sutton. 2016. "How Did the Deepwater Horizon Oil Spill Impact Deep-Sea Ecosystems?" *Oceanography* 29 (3): 182–95. https://doi.org/10.5670/oceanog.2016.82.

Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. "DNA Primers for Amplification of Mitochondrial Cytochrome c Oxidase Subunit I from Diverse Metazoan Invertebrates." *Molecular Marine Biology and Biotechnology* 3 (5): 294–99. https://doi.org/10.1071/ZO9660275.

Fonseca, G., A.W. Muthumbi, and A. Vanreusel. 2007. "Species Richness of the Genus Molgolaimus (Nematoda) from Local to Ocean Scale along Continental Slopes." *Marine Ecology* 28 (4): 446–59. https://doi.org/10.1111/j.1439-0485.2007.00202.x.

Fonseca, G., T. Soltwedel, A. Vanreusel, and M. Lindegarth. 2010. "Variation in Nematode Assemblages over Multiple Spatial Scales and Environmental Conditions in Arctic Deep Seas." *Progress in Oceanography* 84 (3–4): 174–84. https://doi.org/10.1016/j.pocean.2009.11.001.

Fonseca, V.G. 2018. "Pitfalls in Relative Abundance Estimation Using Edna Metabarcoding." *Molecular Ecology Resources* 18 (5): 923–26. https://doi.org/10.1111/1755-0998.12902.

Fonseca, V.G., G.R. Carvalho, B. Nichols, C. Quince, H.F. Johnson, S.P. Neill, P.J.D. Lambshead, W.K. Thomas, D.M. Power, and S. Creer. 2014. "Metagenetic Analysis of Patterns of Distribution and Diversity of Marine Meiobenthic Eukaryotes." *Global Ecology and Biogeography* 23 (11): 1293–1302. https://doi.org/10.1111/geb.12223.

Fonseca, V.G., G.R. Carvalho, W. Sung, H.F. Johnson, D.M. Power, S.P. Neill, M. Packer, et al. 2010. "Second-Generation Environmental Sequencing Unmasks Marine Metazoan Biodiversity." *Nature Communications* 1. https://doi.org/9810.1038/ncomms1095.

Forbes, E. 1844. "Report on the Mollusca and Radiata of the Aegean Sea, and on Their Distribution, Considered as Bearing on Geology." *Report of the 13th British Association for the Advancement of Science. London.* 13: 130–193.

Forster, D., M. Dunthorn, F. Mahé, J.R. Dolan, S. Audic, D. Bass, L. Bittner, et al. 2016. "Benthic Protists: The under-Charted Majority." *FEMS Microbiology Ecology*. https://doi.org/10.1093/femsec/fiw120.

Freudenstein, J. V., M.B. Broe, R.A. Folk, and B.T. Sinn. 2017. "Biodiversity and the Species Concept - Lineages Are Not Enough." *Systematic Biology* 66 (4): 644–56. https://doi.org/10.1093/sysbio/syw098.

Frøslev, T.G., R. Kjøller, H.H. Bruun, R. Ejrnæs, A.K. Brunbjerg, C. Pietroni, and A.J. Hansen. 2017. "Algorithm for Post-Clustering Curation of DNA Amplicon Data Yields Reliable Biodiversity Estimates." *Nature Communications* 8 (1). https://doi.org/10.1038/s41467-017-01312-x.

Fujiwara, Y., M. Kawato, T. Yamamoto, T. Yamanaka, W. Sato-Okoshi, C. Noda, S. Tsuchida, T. Komai, S.S. Cubelio, and T. Sasaki. 2007. "Three-year Investigations into Sperm Whale-fall Ecosystems in Japan." *Marine Ecology* 28 (1): 219–32.

Gaever, S. Van, M. Raes, F. Pasotti, and A. Vanreusel. 2010. "Spatial Scale and Habitat-Dependent Diversity Patterns in Nematode Communities in Three Seepage Related Sites along the Norwegian Sea Margin." *Marine Ecology* 31 (1): 66–77. https://doi.org/10.1111/j.1439-0485.2009.00314.x.

Gallucci, F., G. Fonseca, and T. Soltwedel. 2008. "Effects of Megafauna Exclusion on

Nematode Assemblages at a Deep-Sea Site." *Deep-Sea Research Part I: Oceanographic Research Papers* 55 (3): 332–49. https://doi.org/10.1016/j.dsr.2007.12.001.

Gallucci, F., T. Moens, and G. Fonseca. 2009. "Small-Scale Spatial Patterns of Meiobenthos in the Arctic Deep Sea." *Marine Biodiversity* 39 (1): 9–25. https://doi.org/10.1007/s12526-009-0003-x.

Galtier, N., B. Nabholz, S. GlÉmin, and G.D.D. Hurst. 2009. "Mitochondrial DNA as a Marker of Molecular Diversity: A Reappraisal." *Molecular Ecology* 18 (22): 4541–50. https://doi.org/10.1111/j.1365-294X.2009.04380.x.

Gambi, C., and R. Danovaro. 2006. "A Multiple-Scale Analysis of Metazoan Meiofaunal Distribution in the Deep Mediterranean Sea." *Deep-Sea Research Part I: Oceanographic Research Papers* 53 (7): 1117–34. https://doi.org/10.1016/j.dsr.2006.05.003.

Gasc, C., E. Peyretaillade, and P. Peyret. 2016. "Sequence Capture by Hybridization to Explore Modern and Ancient Genomic Diversity in Model and Nonmodel Organisms." *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkw309.

Gaston, K.J. 2000. "Global Patterns in Biodiversity." *Nature* 405 (6783): 220–27. http://dx.doi.org/10.1038/35012228.

Geller, J., C. Meyer, M. Parker, and H. Hawk. 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 (5). https://doi.org/10.1111/1755-0998.12138.

Gevers, D., F.M. Cohan, J.G. Lawrence, B.G. Spratt, T. Coenye, E.J. Feil, E. Stackebrandt, et al. 2005. "Re-Evaluating Prokaryotic Species." *Nature Reviews Microbiology* 3 (9): 733–39. https://doi.org/10.1038/nrmicro1236.

Gibson, J.F., S. Shokralla, C. Curry, D.J. Baird, W.A. Monk, I. King, and M. Hajibabaei. 2015. "Large-Scale Biomonitoring of Remote and Threatened Ecosystems via High-Throughput Sequencing." Edited by Fontaneto, D. *PLoS ONE* 10 (10): e0138432. https://doi.org/10.1371/journal.pone.0138432.

Giere, O. 2009. *Meiobenthology: The Microscopic Motile Fauna of Aquatic Sediments*. *Meiobenthology: The Microscopic Motile Fauna of Aquatic Sediments*. https://doi.org/10.1007/978-3-540-68661-3.

Gleason, J.E., V. Elbrecht, T.W.A. Braukmann, R.H. Hanner, and K. Cottenie. 2020. "Assessment of Stream Macroinvertebrate Communities with EDNA Is Not Congruent with Tissue-Based Metabarcoding." *Molecular Ecology*. https://doi.org/10.1111/mec.15597.

Glud, R.N., F. Wenzhöfer, M. Middelboe, K. Oguri, R. Turnewitsch, D.E. Canfield, and H. Kitazato. 2013. "High Rates of Microbial Carbon Turnover in Sediments in the Deepest Oceanic Trench on Earth." *Nature Geoscience* 6 (4): 284–88. https://doi.org/10.1038/ngeo1773.

Goffredi, S.K., C.K. Paull, K. Fulton-Bennett, L.A. Hurtado, and R.C. Vrijenhoek. 2004. "Unusual Benthic Fauna Associated with a Whale Fall in Monterey Canyon, California." *Deep Sea Research Part I: Oceanographic Research Papers* 51 (10): 1295–1306.

Goldberg, C.S., C.R. Turner, K. Deiner, K.E. Klymus, P.F. Thomsen, M.A. Murphy, S.F. Spear, et al. 2016. "Critical Considerations for the Application of Environmental DNA Methods to Detect Aquatic Species." *Methods in Ecology and Evolution* 7 (11): 1299–1307. https://doi.org/10.1111/2041-210X.12595.

Gómez-Rodríguez, C., A. Crampton-Platt, M.J.T.N. Timmermans, A. Baselga, and A.P. Vogler. 2015. "Validating the Power of Mitochondrial Metagenomics for Community Ecology and Phylogenetics of Complex Assemblages." Edited by Gilbert, M. *Methods in Ecology and Evolution* 6 (8): 883–94. https://doi.org/10.1111/2041-210X.12376.

Gooday, A.J. 1999. "Biodiversity of Foraminifera and Other Protists in the Deep Sea: Scales and Patterns." *Belg. J. Zool.* 129 (1): 61–80.

Gooday, A.J., S. Hori, Y. Todo, T. Okamoto, H. Kitazato, and A. Sabbatini. 2004. "Soft-Walled, Monothalamous Benthic Foraminiferans in the Pacific, Indian and Atlantic Oceans: Aspects of Biodiversity and Biogeography." *Deep-Sea Research Part I: Oceanographic Research Papers* 51 (1): 33–53. https://doi.org/10.1016/j.dsr.2003.07.002.

Goordial, J., I. Altshuler, K. Hindson, K. Chan-Yam, E. Marcolefas, and L.G. Whyte. 2017. "In Situ Field Sequencing and Life Detection in Remote (79°26′N) Canadian High Arctic Permafrost Ice Wedge Microbial Communities." *Frontiers in Microbiology* 8 (DEC): 2594. https://doi.org/10.3389/fmicb.2017.02594.

Górska, B., K. Grzelak, L. Kotwicki, C. Hasemann, I. Schewe, T. Soltwedel, and M. Włodarska-Kowalczuk. 2014. "Bathymetric Variations in Vertical Distribution Patterns of Meiofauna in the Surface Sediments of the Deep Arctic Ocean (HAUSGARTEN, Fram Strait)." *Deep-Sea Research Part I: Oceanographic Research Papers* 91: 36–49. https://doi.org/10.1016/j.dsr.2014.05.010.

Grabchak, M., E. Marcon, G. Lang, and Z. Zhang. 2017. "The Generalized Simpson's Entropy Is a Measure of Biodiversity." Edited by Green, S.J. *PLoS ONE* 12 (3): e0173305. https://doi.org/10.1371/journal.pone.0173305.

Grassle, J.F. 1989. "Species Diversity in Deep-Sea Communities." *Trends in Ecology and Evolution* 4 (1): 12–15. https://doi.org/10.1016/0169-5347(89)90007-4.

Grassle, J.F., and N.J. Maciolek. 1992. "Deep-Sea Species Richness: Regional and Local Diversity Estimates from Quantitative Bottom Samples." *American Naturalist* 139 (2): 313–41. https://doi.org/10.1086/285329.

Grey, E.K., L. Bernatchez, P. Cassey, K. Deiner, M. Deveney, K.L. Howland, A. Lacoursière-Roussel, et al. 2018. "Effects of Sampling Effort on Biodiversity Patterns Estimated from Environmental DNA Metabarcoding Surveys." *Scientific Reports* 8 (1): 8843. https://doi.org/10.1038/s41598-018-27048-2.

Guardiola, M., O.S. Wangensteen, P. Taberlet, E. Coissac, M.J. Uriz, and X. Turon. 2016. "Spatio-Temporal Monitoring of Deep-Sea Communities Using Metabarcoding of Sediment DNA and RNA." *PeerJ* 4: e2807. https://doi.org/10.7717/peerj.2807.

Guillou, L., D. Bachar, S. Audic, D. Bass, C. Berney, L. Bittner, C. Boutte, et al. 2013. "The Protist Ribosomal Reference Database (PR2): A Catalog of Unicellular Eukaryote Small Sub-Unit RRNA Sequences with Curated Taxonomy." *Nucleic Acids Research* 41 (D1): D597-604. https://doi.org/10.1093/nar/gks1160.

Haase, K.M., S. Petersen, A. Koschinsky, R. Seifert, C.W. Devey, R. Keir, K.S. Lackschewitz, et al. 2009. "Fluid Compositions and Mineralogy of Precipitates from Mid Atlantic Ridge Hydrothermal Vents at 4°48'S." PANGAEA. https://doi.org/10.1594/PANGAEA.727454.

Hajibabaei, M., T.M. Porter, C. V. Robinson, D.J. Baird, S. Shokralla, and M.T.G. Wright. 2019. "Watered-down Biodiversity? A Comparison of Metabarcoding Results from DNA Extracted from Matched Water and Bulk Tissue Biomonitoring Samples." *PLoS ONE* 14 (12): 1–16. https://doi.org/10.1371/journal.pone.0225409.

Halpern, B.S., S. Walbridge, K.A. Selkoe, C. V Kappel, F. Micheli, C. Agrosa, J.F. Bruno, et al. 2008. "A Global Map of Human Impact on Marine Ecosystems." *Science* 319 (5865): 948 LP – 952. http://science.sciencemag.org/content/319/5865/948.abstract.

Hänfling, B., L.L. Handley, D.S. Read, C. Hahn, J. Li, P. Nichols, R.C. Blackman, A. Oliver, and I.J. Winfield. 2016. "Environmental DNA Metabarcoding of Lake Fish Communities Reflects Long-Term Data from Established Survey Methods." *Molecular Ecology* 25 (13): 3101–19. https://doi.org/10.1111/mec.13660.

Hardy, S.M., C.R. Smith, and A.M. Thurnherr. 2015. "Can the Source-Sink Hypothesis Explain Macrofaunal Abundance Patterns in the Abyss? A Modelling Test." *Proc. R. Soc. B* 282 (1808): 20150193-. https://doi.org/10.1098/rspb.2015.0193.

Hasemann, C., and T. Soltwedel. 2011. "Small-Scale Heterogeneity in Deep-Sea Nematode Communities around Biogenic Structures." Edited by Voolstra, C.R. *PLoS ONE* 6 (12): e29152. https://doi.org/10.1371/journal.pone.0029152.

Hashimoto, J.G., B.S. Stevenson, and T.M. Schmidt. 2003. "Rates and Consequences of Recombination between RRNA Operons." *Journal of Bacteriology* 185 (3): 966–72. https://doi.org/10.1128/JB.185.3.966-972.2003.

Hauquier, F., L. Macheriotou, T. Bezerra, G. Egho, P. Martínez Arbizu, and A. Vanreusel. 2019. "Distribution of Free-Living Marine Nematodes in the Clarion-Clipperton Zone: Implications for Future Deep-Sea Mining Scenarios." *Biogeosciences* 16 (18): 3475–89. https://doi.org/10.5194/bg-16-3475-2019.

Havermans, C., G. Sonet, C. d'Udekem d'Acoz, Z.T. Nagy, P. Martin, S. Brix, T. Riehl, S. Agrawal, and C. Held. 2013. "Genetic and Morphological Divergences in the Cosmopolitan Deep-Sea Amphipod Eurythenes Gryllus Reveal a Diverse Abyss and a Bipolar Species." Edited by Fontaneto, D. *PLoS ONE* 8 (9): e74218. https://doi.org/10.1371/journal.pone.0074218.

Hebert, P.D.N., A. Cywinska, S.L. Ball, and J.R. DeWaard. 2003a. "Biological Identifications through DNA Barcodes." *Proceedings of the Royal Society B: Biological Sciences* 270 (1512): 313–21. https://doi.org/10.1098/rspb.2002.2218.

Hebert, P.D.N., P.M. Hollingsworth, and M. Hajibabaei. 2016a. "From Writing to Reading the Encyclopedia of Life." *Philosophical Transactions of the Royal Society B: Biological Sciences* 371 (1702). https://doi.org/10.1098/rstb.2015.0321.

Hebert, P.D.N., S. Ratnasingham, and J.R. de Waard. 2003b. "Barcoding Animal Life: Cytochrome c Oxidase Subunit 1 Divergences among Closely Related Species." *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270 (suppl_1): S96-9. https://doi.org/10.1098/rsbl.2003.0025.

Hebert, P.D.N., S. Ratnasingham, E. V Zakharov, A.C. Telfer, V. Levesque-Beaudin, M.A. Milton, S. Pedersen, P. Jannetta, and J.R. deWaard. 2016b. "Counting Animal Species with DNA Barcodes: Canadian Insects." *Philosophical Transactions of the Royal Society B: Biological Sciences* 371 (1702). https://doi.org/10.1098/rstb.2015.0333.

Hendriks, I.E., and C.M. Duarte. 2008. "Allocation of Effort and Imbalances in Biodiversity Research." *Journal of Experimental Marine Biology and Ecology* 360 (1): 15–20. https://doi.org/10.1016/j.jembe.2008.03.004.

Hendriks, I.E., C.M. Duarte, and C.H.R. Heip. 2006. "Biodiversity Research Still Grounded." *Science* 312 (5781): 1715. https://doi.org/10.1126/science.1128548.

Hengeveld, R. 1988. "Mayr's Ecological Species Criterion." *Systematic Zoology* 37 (1): 47. https://doi.org/10.2307/2413188.

Herrera, S., H. Watanabe, and T.M. Shank. 2015. "Evolutionary and Biogeographical Patterns of Barnacles from Deep-Sea Hydrothermal Vents." *Molecular Ecology* 24 (3): 673–89. https://doi.org/10.1111/mec.13054.

Hsieh, T.C., K.H. Ma, and A. Chao. 2016. "INEXT: An R Package for Rarefaction and Extrapolation of Species Diversity (Hill Numbers)." Edited by McInerny, G. *Methods in Ecology and Evolution* 7 (12): 1451–56. https://doi.org/10.1111/2041-210X.12613.

Hu, S., J. Sprintall, C. Guan, M.J. McPhaden, F. Wang, D. Hu, and W. Cai. 2020. "Deep-Reaching Acceleration of Global Mean Ocean Circulation over the Past Two Decades." *Science Advances* 6 (6): eaax7727. https://doi.org/10.1126/sciadv.aax7727.

Huang, D., R. Meier, P.A. Todd, and L.M. Chou. 2008. "Slow Mitochondrial COI Sequence

Evolution at the Base of the Metazoan Tree and Its Implications for DNA Barcoding." *Journal of Molecular Evolution* 66 (2): 167–74. https://doi.org/10.1007/s00239-008-9069-5.

Hurst, G.D.., and F.M. Jiggins. 2005. "Problems with Mitochondrial DNA as a Marker in Population, Phylogeographic and Phylogenetic Studies: The Effects of Inherited Symbionts." *Proceedings of the Royal Society B: Biological Sciences* 272 (1572): 1525–34. https://doi.org/10.1098/rspb.2005.3056.

Ibarbalz, F.M., N. Henry, M.C. Brandão, S. Martini, G. Busseni, H. Byrne, L.P. Coelho, et al. 2019. "Global Trends in Marine Plankton Diversity across Kingdoms of Life." *Cell* 179 (5): 1084-1097.e21. https://doi.org/10.1016/j.cell.2019.10.008.

Ingels, J., D.S.M. Billett, K. Kiriakoulakis, G.A. Wolff, and A. Vanreusel. 2011. "Structural and Functional Diversity of Nematoda in Relation with Environmental Variables in the Setúbal and Cascais Canyons, Western Iberian Margin." *Deep-Sea Research Part II: Topical Studies in Oceanography* 58 (23–24): 2354–68. https://doi.org/10.1016/j.dsr2.2011.04.002.

Itoh, M., K. Kawamura, T. Kitahashi, S. Kojima, H. Katagiri, and M. Shimanaga. 2011. "Bathymetric Patterns of Meiofaunal Abundance and Biomass Associated with the Kuril and Ryukyu Trenches, Western North Pacific Ocean." *Deep-Sea Research Part I: Oceanographic Research Papers* 58 (1): 86–97. https://doi.org/10.1016/j.dsr.2010.12.004.

Jamieson, A.J. 2011. "Ecology of Deep Oceans: Hadal Trenches." In *ELS*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470015902.a0023606.

———. 2015. *The Hadal Zone: Life in the Deepest Oceans*. *The Hadal Zone: Life in the Deepest Oceans*. Cambridge University Press. https://doi.org/10.1017/CBO9781139061384.

Jamieson, A.J., T. Fujii, D.J. Mayor, M. Solan, and I.G. Priede. 2010. "Hadal Trenches: The Ecology of the Deepest Places on Earth." *Trends in Ecology and Evolution*. https://doi.org/10.1016/j.tree.2009.09.009.

Ji, Y., L. Ashton, S.M. Pedley, D.P. Edwards, Y. Tang, A. Nakamura, R. Kitching, et al. 2013. "Reliable, Verifiable and Efficient Monitoring of Biodiversity via Metabarcoding." *Ecology Letters* 16 (10): 1245–57. https://doi.org/10.1111/ele.12162.

Jo, T., H. Murakami, R. Masuda, M.K. Sakata, S. Yamamoto, and T. Minamoto. 2017. "Rapid Degradation of Longer DNA Fragments Enables the Improved Estimation of Distribution and Biomass Using Environmental DNA." *Molecular Ecology Resources* 17 (6): e25–33. https://doi.org/10.1111/1755-0998.12685.

Jones, M.R., and J.M. Good. 2016. "Targeted Capture in Evolutionary and Ecological Genomics." *Molecular Ecology* 25 (1): 185–202. https://doi.org/10.1111/mec.13304.

Jorda, G., N. Marbà, S. Bennett, J. Santana-Garcon, S. Agusti, and C.M. Duarte. 2020. "Ocean Warming Compresses the Three-Dimensional Habitat of Marine Life." *Nature Ecology and Evolution* 4 (1): 109–14. https://doi.org/10.1038/s41559-019-1058-0.

Jost, L. 2007. "Partitioning Diversity into Independent Alpha and Beta Components." *Ecology* 88 (10): 2427–39. https://doi.org/10.1890/06-1736.1.

———. 2008. "GST and Its Relatives Do Not Measure Differentiation." *Molecular Ecology* 17 (18): 4015–26. https://doi.org/10.1111/j.1365-294X.2008.03887.x.

Keck, F., V. Vasselon, F. Rimet, A. Bouchez, and M. Kahlert. 2018. "Boosting DNA Metabarcoding for Biomonitoring with Phylogenetic Estimation of Operational Taxonomic Units' Ecological Profiles." *Molecular Ecology Resources* 18 (6): 1299–1309. https://doi.org/10.1111/1755-0998.12919.

Keeley, N., S.A. Wood, and X. Pochon. 2018. "Development and Preliminary Validation of a Multi-Trophic Metabarcoding Biotic Index for Monitoring Benthic Organic Enrichment."

Ecological Indicators 85 (February): 1044–57. https://doi.org/10.1016/j.ecolind.2017.11.014.

Kitahashi, T., K. Kawamura, S. Kojima, and M. Shimanaga. 2013. "Assemblages Gradually Change from Bathyal to Hadal Depth: A Case Study on Harpacticoid Copepods around the Kuril Trench (North-West Pacific Ocean)." *Deep-Sea Research Part I: Oceanographic Research Papers* 74: 39–47. https://doi.org/10.1016/j.dsr.2012.12.010.

Klappenbach, J.A., P.R. Saxman, C.J. R., and T.M. Schmidt. 2001. "Rrndb: The Ribosomal RNA Operon Copy Number Database." *Nucleic Acids Research* 29 (1): 181–84. https://doi.org/10.1093/nar/29.1.181.

Klunder, L., H. de Stigter, M.S.S. Lavaleye, J.D.L. van Bleijswijk, H.W. van der Veer, G.J. Reichart, and G.C.A. Duineveld. 2020. "A Molecular Approach to Explore the Background Benthic Fauna Around a Hydrothermal Vent and Their Larvae: Implications for Future Mining of Deep-Sea SMS Deposits." *Frontiers in Marine Science* 7 (March): 1–12. https://doi.org/10.3389/fmars.2020.00134.

Konstantinidis, K.T., A. Ramette, and J.M. Tiedje. 2006. "The Bacterial Species Definition in the Genomic Era." *Philosophical Transactions of the Royal Society B: Biological Sciences* 361 (1475): 1929–40. https://doi.org/10.1098/rstb.2006.1920.

Koziol, A., M. Stat, T. Simpson, S. Jarman, J.D. DiBattista, E.S. Harvey, M. Marnane, J. McDonald, and M. Bunce. 2019. "Environmental DNA Metabarcoding Studies Are Critically Affected by Substrate Selection." *Molecular Ecology Resources* 19 (2): 366–76. https://doi.org/10.1111/1755-0998.12971.

Krehenwinkel, H., M. Wolf, J.Y. Lim, A.J. Rominger, W.B. Simison, and R.G. Gillespie. 2017. "Estimating and Mitigating Amplification Bias in Qualitative and Quantitative Arthropod Metabarcoding." *Scientific Reports* 7 (1): 17668. https://doi.org/10.1038/s41598-017-17333-x.

Lamb, P.D., E. Hunter, J.K. Pinnegar, S. Creer, R.G. Davies, and M.I. Taylor. 2019. "How Quantitative Is Metabarcoding: A Meta-Analytical Approach." *Molecular Ecology* 28 (2): 420–30. https://doi.org/10.1111/mec.14920.

Lambshead, P.J.D., C.J. Brown, T.J. Ferrero, N.J. Mitchell, C.R. Smith, L.E. Hawkins, and J.H. Tietjen. 2002. "Latitudinal Diversity Patterns of Deep-Sea Marine Nematodes and Organic Fluxes: A Test from the Central Equatorial Pacific." *Marine Ecology Progress Series* 236: 129–35. https://doi.org/10.3354/meps236129.

Lambshead, P.J.D., T.J. Ferrero, and G.A. Wollf. 1995. "Comparison of the Vertical Distribution of Nematodes from Two Contrasting Abyssal Sites in the Northeast Atlantic Subject to Different Seasonal Inputs of Phytodetritus." *Internationale Revue Der Gesamten Hydrobiologie Und Hydrographie* 80 (2): 327–31. https://doi.org/10.1002/iroh.19950800219.

Lanzén, A., K. Lekang, I. Jonassen, E.M. Thompson, and C. Troedsson. 2017. "DNA Extraction Replicates Improve Diversity and Compositional Dissimilarity in Metabarcoding of Eukaryotes in Marine Sediments." Edited by Hajibabaei, M. *PLoS ONE* 12 (6): e0179443. https://doi.org/10.1371/journal.pone.0179443.

Laroche, O., S.A. Wood, L.A. Tremblay, J.I. Ellis, G. Lear, and X. Pochon. 2018. "A Cross-Taxa Study Using Environmental DNA/RNA Metabarcoding to Measure Biological Impacts of Offshore Oil and Gas Drilling and Production Operations." *Marine Pollution Bulletin* 127: 97–107. https://doi.org/10.1016/j.marpolbul.2017.11.042.

Laroche, O., S.A. Wood, L.A. Tremblay, J.I. Ellis, F. Lejzerowicz, J.W. Pawlowski, G. Lear, J. Atalah, and X. Pochon. 2016. "First Evaluation of Foraminiferal Metabarcoding for Monitoring Environmental Impact from an Offshore Oil Drilling Site." *Marine Environmental Research* 120: 225–35.

https://doi.org/10.1016/J.MARENVRES.2016.08.009.

Laroche, O., S.A. Wood, L.A. Tremblay, G. Lear, J.I. Ellis, and X. Pochon. 2017. "Metabarcoding Monitoring Analysis: The Pros and Cons of Using Co-Extracted Environmental DNA and RNA Data to Assess Offshore Oil Production Impacts on Benthic Communities." *PeerJ* 2017 (5): e3347. https://doi.org/10.7717/peerj.3347.

Lavit, C., Y. Escoufier, R. Sabatier, and P. Traissac. 1994. "The ACT (STATIS Method)." *Computational Statistics and Data Analysis* 18 (1): 97–119. https://doi.org/10.1016/0167-9473(94)90134-1.

Leduc, D., S.D. Nodder, K. Berkenbusch, and A.A. Rowden. 2015. "Effect of Core Surface Area and Sediment Depth on Estimates of Deep-Sea Nematode Genus Richness and Community Structure." *Marine Biodiversity* 45 (3): 349–56. https://doi.org/10.1007/s12526-014-0271-y.

Leduc, D., P.K. Probert, and S.D. Nodder. 2010. "Influence of Mesh Size and Core Penetration on Estimates of Deep-Sea Nematode Abundance, Biomass, and Diversity." *Deep-Sea Research Part I: Oceanographic Research Papers* 57 (10): 1354–62. https://doi.org/10.1016/j.dsr.2010.06.005.

Leduc, D., A.A. Rowden, D.A. Bowden, S.D.S. Nodder, P.K. Probert, C.C.A. Pilditch, G.G.C.A. Duineveld, and R. Witbaard. 2012a. "Nematode Beta Diversity on the Continental Slope of New Zealand : Spatial Patterns and Environmental Drivers." *Marine Ecology Progress Series* 454 (Legendre 1993): 37–52. https://doi.org/10.3354/meps09690.

Leduc, D., A.A. Rowden, R.N. Glud, F. Wenzhöfer, H. Kitazato, and M.R. Clark. 2016. "Comparison between Infaunal Communities of the Deep Floor and Edge of the Tonga Trench: Possible Effects of Differences in Organic Matter Supply." *Deep-Sea Research Part I: Oceanographic Research Papers* 116 (October): 264–75. https://doi.org/10.1016/j.dsr.2015.11.003.

Leduc, D., A.A. Rowden, P.K. Probert, C.A. Pilditch, S.D. Nodder, A. Vanreusel, G.C.A. Duineveld, and R. Witbaard. 2012b. "Further Evidence for the Effect of Particle-Size Diversity on Deep-Sea Benthic Biodiversity." *Deep-Sea Research Part I: Oceanographic Research Papers* 63: 164–69. https://doi.org/10.1016/j.dsr.2011.10.009.

Legendre, P., and M. De Cáceres. 2013. "Beta Diversity as the Variance of Community Data: Dissimilarity Coefficients and Partitioning." *Ecology Letters* 16 (8): 951–63. https://doi.org/10.1111/ele.12141.

Lejzerowicz, F., P. Esling, W. Majewski, W. Szczuciński, J. Decelle, C. Obadia, P.M. Arbizu, and J.W. Pawlowski. 2013a. "Ancient DNA Complements Microfossil Record in Deep-Sea Subsurface Sediments." *Biology Letters* 9 (4): 20130283. https://doi.org/10.1098/rsbl.2013.0283.

Lejzerowicz, F., P. Esling, and J.W. Pawlowski. 2014. "Patchiness of Deep-Sea Benthic Foraminifera across the Southern Ocean: Insights from High-Throughput DNA Sequencing." *Deep-Sea Research Part II: Topical Studies in Oceanography* 108: 17–26. https://doi.org/10.1016/j.dsr2.2014.07.018.

Lejzerowicz, F., I. Voltsky, and J.W. Pawlowski. 2013b. "Identifying Active Foraminifera in the Sea of Japan Using Metatranscriptomic Approach." *Deep-Sea Research Part II: Topical Studies in Oceanography* 86–87: 214–20. https://doi.org/10.1016/j.dsr2.2012.08.008.

Lennon, J.T., M.E. Muscarella, S.A. Placella, and B.K. Lehmkuhl. 2018. "How, When, and Where Relic DNA Affects Microbial Diversity." *MBio* 9 (3): e00637-18. https://doi.org/10.1128/mBio.00637-18.

Leray, M., and N. Knowlton. 2015. "DNA Barcoding and Metabarcoding of Standardized

Samples Reveal Patterns of Marine Benthic Diversity." *Proceedings of the National Academy of Sciences of the United States of America* 112 (7): 2076–81. https://doi.org/10.1073/pnas.1424997112.

———. 2017. "Random Sampling Causes the Low Reproducibility of Rare Eukaryotic OTUs in Illumina COI Metabarcoding." *PeerJ* 2017 (3): e3006. https://doi.org/10.7717/peerj.3006.

Leray, M., N. Knowlton, S.L. Ho, B.N. Nguyen, and R.J. Machida. 2019. "GenBank Is a Reliable Resource for 21st Century Biodiversity Research." *Proceedings of the National Academy of Sciences of the United States of America* 116 (45): 22651–56. https://doi.org/10.1073/pnas.1911714116.

Leray, M., J.Y. Yang, C.P. Meyer, S.C. Mills, N. Agudelo, V. Ranwez, J.T. Boehm, and R.J. Machida. 2013. "A New Versatile Primer Set Targeting a Short Fragment of the Mitochondrial COI Region for Metabarcoding Metazoan Diversity: Application for Characterizing Coral Reef Fish Gut Contents." *Front Zool* 10: 34. https://doi.org/10.1186/1742-9994-10-34.

Levin, D.A. 2000. *The Origin, Expansion, and Demise of Plant Species*. Oxford: Oxford University Press. https://books.google.fr/books?hl=en&lr=&id=bI83LxrWKAcC&oi=fnd&pg=PR7&dq=Levin+(2000)+ecogenetic+concept&ots=nEZzYsN7Cs&sig=Ho-ZiFIuFuCZgpXJclMSatRw31Y#v=onepage&q=Levin (2000) ecogenetic concept&f=false.

Levin, L.A. 2003. "Oxygen Minimum Zone Benthos: Adaptation and Community Response to Hypoxia." In *Oceanography and Marine Biology: An Annual Review*, edited by Gibson, R.N. and Atkinson, R.J.A., 41:1–45. Taylor & Francis Group.

Levin, L.A., R.J. Etter, M.A. Rex, A.J. Gooday, C.R. Smith, J. Pineda, C.T. Stuart, R.R. Hessler, and D. Pawson. 2001. "Environmental Influences on Regional Deep-Sea Species Diversity." *Annual Review of Ecology and Systematics* 32 (1): 51–93. https://doi.org/10.1146/annurev.ecolsys.32.081501.114002.

Levin, L.A., K. Mengerink, K.M. Gjerde, A.A. Rowden, C.L. Van Dover, M.R. Clark, E. Ramirez-Llodra, et al. 2016. "Defining 'Serious Harm' to the Marine Environment in the Context of Deep-Seabed Mining." *Marine Policy* 74 (December): 245–59. https://doi.org/10.1016/j.marpol.2016.09.032.

Liao, L., X.W. Xu, X.W. Jiang, C.S. Wang, D.S. Zhang, J.Y. Ni, and M. Wu. 2011. "Microbial Diversity in Deep-Sea Sediment from the Cobalt-Rich Crust Deposit Region in the Pacific Ocean." *FEMS Microbiology Ecology* 78 (3): 565–85. https://doi.org/10.1111/j.1574-6941.2011.01186.x.

Linley, T.D., M.E. Gerringer, P.H. Yancey, J.C. Drazen, C.L. Weinstock, and A.J. Jamieson. 2016. "Fishes of the Hadal Zone Including New Species, in Situ Observations and Depth Records of Liparidae." *Deep-Sea Research Part I: Oceanographic Research Papers* 114 (August): 99–110. https://doi.org/10.1016/j.dsr.2016.05.003.

Liu, S., X. Wang, L. Xie, M. Tan, Z. Li, X. Su, H. Zhang, et al. 2016. "Mitochondrial Capture Enriches Mito-DNA 100 Fold, Enabling PCR-Free Mitogenomics Biodiversity Analysis." *Molecular Ecology Resources* 16 (2): 470–79. https://doi.org/10.1111/1755-0998.12472.

Locey, K.J., and J.T. Lennon. 2016. "Scaling Laws Predict Global Microbial Diversity." *Proceedings of the National Academy of Sciences of the United States of America* 113 (21): 5970–75. https://doi.org/10.1073/pnas.1521291113.

Lochte, K., and C.M. Turley. 1988. "Bacteria and Cyanobacteria Associated with Phytodetritus in the Deep Sea." *Nature* 333 (6168): 67–69. https://doi.org/10.1038/333067a0.

Lopes, C.M., T. Sasso, A. Valentini, T. Dejean, M. Martins, K.R. Zamudio, and C.F.B. Haddad.

2017. “EDNA Metabarcoding: A Promising Method for Anuran Surveys in Highly Diverse Tropical Forests.” *Molecular Ecology Resources* 17 (5): 904–14. https://doi.org/10.1111/1755-0998.12643.

Loreau, M., S. Naeem, P. Inchausti, J. Bengtsson, J.P. Grime, A. Hector, D.U. Hooper, et al. 2001. “Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges.” *Science* 294 (5543): 804–8. https://doi.org/10.1126/science.1064088.

Lynch, M., and J.S. Conery. 2003. “The Origins of Genome Complexity.” *Science* 302 (5649): 1401–4. https://doi.org/10.1126/science.1089370.

Macheriotou, L., K. Guilini, T.N. Bezerra, B. Tytgat, D.T. Nguyen, T.X. Phuong Nguyen, F. Noppe, et al. 2019. “Metabarcoding Free-Living Marine Nematodes Using Curated 18S and CO1 Reference Sequence Databases for Species-Level Taxonomic Assignments.” *Ecology and Evolution* 9 (1): 1–16. https://doi.org/10.1002/ece3.4814.

Machida, R.J., and N. Knowlton. 2012. “PCR Primers for Metazoan Nuclear 18S and 28S Ribosomal DNA Sequences.” Edited by Gilbert, J.A. *PLoS ONE* 7 (9): e46180. https://doi.org/10.1371/journal.pone.0046180.

Machida, R.J., M. Kweskin, and N. Knowlton. 2012. “PCR Primers for Metazoan Mitochondrial 12S Ribosomal DNA Sequences.” *PLoS ONE* 7 (4). https://doi.org/10.1371/journal.pone.0035887.

Machida, R.J., M. Leray, S.L. Ho, and N. Knowlton. 2017. “Data Descriptor: Metazoan Mitochondrial Gene Sequence Reference Datasets for Taxonomic Assignment of Environmental Samples.” *Scientific Data* 4. https://doi.org/10.1038/sdata.2017.27.

Mächler, E., K. Deiner, F. Spahn, and F. Altermatt. 2016. “Fishing in the Water: Effect of Sampled Water Volume on Environmental DNA-Based Detection of Macroinvertebrates.” *Environmental Science and Technology* 50 (1): 305–12. https://doi.org/10.1021/acs.est.5b04188.

Mahé, F., T. Rognes, C. Quince, C. de Vargas, and M. Dunthorn. 2015. “Swarm v2: Highly-Scalable and High-Resolution Amplicon Clustering.” *PeerJ* 3 (December): e1420. https://doi.org/10.7717/peerj.1420.

Martin, J.W., and T.A. Haney. 2005. “Decapod Crustaceans from Hydrothermal Vents and Cold Seeps: A Review through 2005.” *Zoological Journal of the Linnean Society*. Blackwell Science Ltd. https://doi.org/10.1111/j.1096-3642.2005.00178.x.

Marx, V. 2015. “PCR Heads into the Field.” *Nature Methods* 12 (5): 393–97. https://doi.org/10.1038/nmeth.3369.

Massana, R.R., A. Gobet, S. Audic, D. Bass, L. Bittner, C. Boutte, A. Chambouvet, et al. 2015. “Marine Protist Diversity in European Coastal Waters and Sediments as Revealed by High-Throughput Sequencing.” *Environmental Microbiology* 17 (10): 4035–49. https://doi.org/10.1111/1462-2920.12955.

May, R.M. 1988. “How Many Species Are There on Earth?” *Science* 241 (4872): 1441–49. https://doi.org/10.1126/science.241.4872.1441.

Mayr, E. 1942. *Systematics and the Origin of Species, from the Viewpoint of a Zoologist*. New York, NY: Columbia University Press. http://www.hup.harvard.edu/catalog.php?isbn=9780674862500.

———. 1982. *The Growth of Biological Thought : Diversity, Evolution, and Inheritance*. Belknap Press. https://www.hup.harvard.edu/catalog.php?isbn=9780674364462.

McCauley, D.J., M.L. Pinsky, S.R. Palumbi, J.A. Estes, F.H. Joyce, and R.R. Warner. 2015. “Marine Defaunation: Animal Loss in the Global Ocean.” *Science* 347 (6219). https://doi.org/10.1126/science.1255641.

McClain, C.R., and S.M. Hardy. 2010. “The Dynamics of Biogeographic Ranges in the Deep Sea.” *Proceedings of the Royal Society B: Biological Sciences* 277 (1700): 3533–46.

https://doi.org/10.1098/rspb.2010.1057.

McIntyre, A.D., R.J. Menzies, R.Y. George, and G.T. Rowe. 1975. *Abyssal Environment and Ecology of the World Oceans*. *The Journal of Animal Ecology*. Vol. 44. New York: John Wiley and Sons. https://doi.org/10.2307/3873.

McMurdie, P.J., and S. Holmes. 2013. "Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data." *PLoS ONE* 8 (4): e61217. https://doi.org/10.1371/journal.pone.0061217.

Menzel, L., K.H. George, and P.M. Arbizu. 2011. "Submarine Ridges Do Not Prevent Large-Scale Dispersal of Abyssal Fauna: A Case Study of Mesocletodes (Crustacea, Copepoda, Harpacticoida)." *Deep-Sea Research Part I: Oceanographic Research Papers* 58 (8): 839–64. https://doi.org/10.1016/j.dsr.2011.05.008.

Merlino, G., A. Barozzi, G. Michoud, D.K. Ngugi, and D. Daffonchio. 2018. "Microbial Ecology of Deep-Sea Hypersaline Anoxic Basins." *FEMS Microbiology Ecology*. Oxford University Press. https://doi.org/10.1093/femsec/fiy085.

Meyer, C.P., and G. Paulay. 2005. "DNA Barcoding: Error Rates Based on Comprehensive Sampling." Edited by Godfray, C. *PLoS Biology* 3 (12): 1–10. https://doi.org/10.1371/journal.pbio.0030422.

Minoche, A.E., J.C. Dohm, and H. Himmelbauer. 2011. "Evaluation of Genomic High-Throughput Sequencing Data Generated on Illumina HiSeq and Genome Analyzer Systems." *Genome Biology* 12 (11): R112. https://doi.org/10.1186/gb-2011-12-11-r112.

Moalic, Y., D. Desbruyeres, C.M. Duarte, A.F. Rozenfeld, C. Bachraty, and S. Arnaud-Haond. 2012. "Biogeography Revisited with Network Theory: Retracing the History of Hydrothermal Vent Communities." *Systematic Biology* 61 (1): 127–37. https://doi.org/10.1093/sysbio/syr088.

Montagna, P.A., J.G. Baguley, C.Y. Hsiang, and M.G. Reuscher. 2017. "Comparison of Sampling Methods for Deep-Sea Infauna." *Limnology and Oceanography: Methods* 15 (2): 166–83. https://doi.org/10.1002/lom3.10150.

Mora, C., D.P. Tittensor, S. Adl, A.G.B. Simpson, and B. Worm. 2011. "How Many Species Are There on Earth and in the Ocean?" Edited by Mace, G.M. *PLoS Biology* 9 (8). https://doi.org/10.1371/journal.pbio.1001127.

Mulligan, C.J., A. Kitchen, and M.M. Miyamoto. 2006. "Comment on 'Population Size Does Not Influence Mitochondrial Genetic Diversity in Animals.'" *Science* 314 (5804): 1390a-1390a. https://doi.org/10.1126/science.1132585.

Murray Roberts, J., A.J. Wheeler, A.A. Freiwald, and J.M. Roberts. 2006. "Reefs of the Deep: The Biology and Geology of Cold-Water Coral Ecosystems." *Science* 312 (5773): 543 LP – 547. https://doi.org/DOI 10.1126/science.1119861.

Nabholz, B., S. Glémin, and N. Galtier. 2009. "The Erratic Mitochondrial Clock: Variations of Mutation Rate, Not Population Size, Affect MtDNA Diversity across Birds and Mammals." *BMC Evolutionary Biology* 9 (1): 54. https://doi.org/10.1186/1471-2148-9-54.

Nabholz, B., J.F. Mauffrey, E. Bazin, N. Galtier, and S. Glemin. 2008. "Determination of Mitochondrial Genetic Diversity in Mammals." *Genetics* 178 (1): 351–61. https://doi.org/10.1534/genetics.107.073346.

Nagler, M., H. Insam, G. Pietramellara, and J. Ascher-Jenull. 2018. "Extracellular DNA in Natural Environments: Features, Relevance and Applications." *Applied Microbiology and Biotechnology* 102 (15): 6343–56. https://doi.org/10.1007/s00253-018-9120-4.

Nascimento, F.J.A., D. Lallias, H.M. Bik, and S. Creer. 2018. "Sample Size Effects on the Assessment of Eukaryotic Diversity and Community Structure in Aquatic Sediments Using High-Throughput Sequencing." *Scientific Reports* 8 (1).

https://doi.org/10.1038/s41598-018-30179-1.

Nearing, J.T., G.M. Douglas, A.M. Comeau, and M.G.I. Langille. 2018. "Denoising the Denoisers: An Independent Evaluation of Microbiome Sequence Error-Correction Approaches." *PeerJ* 6: e5364. https://doi.org/10.7717/peerj.5364.

Nichols, R. V., C. Vollmers, L.A. Newsom, Y. Wang, P.D. Heintzman, M. Leighton, R.E. Green, and B. Shapiro. 2018. "Minimizing Polymerase Biases in Metabarcoding." *Molecular Ecology Resources* 18 (5): 927–39. https://doi.org/10.1111/1755-0998.12895.

Oksanen, J., M. Blanchet, Guillaume F. Friendly, R. Kindt, P. Legendre, D. McGlinn, R.P. Minchin, R.B. O'Hara, et al. 2018. "Vegan: Community Ecology Package." https://cran.r-project.org/package=vegan.

Olu-Le Roy, K., M. Sibuet, A. Fiala-Médioni, S. Gofas, C. Salas, A. Mariotti, J.P. Foucher, and J. Woodside. 2004. "Cold Seep Communities in the Deep Eastern Mediterranean Sea: Composition, Symbiosis and Spatial Distribution on Mud Volcanoes." *Deep-Sea Research Part I: Oceanographic Research Papers* 51 (12): 1915–36. https://doi.org/10.1016/j.dsr.2004.07.004.

Olu, K., E.E. Cordes, C.R. Fisher, J.M. Brooks, M. Sibuet, and D. Desbruyeres. 2010. "Biogeography and Potential Exchanges among the Atlantic Equatorial Belt Cold-Seep Faunas." *PLoS ONE*. https://doi.org/10.1371/journal.pone.0011967.

Olu, K., A. Duperret, M. Sibuet, J.P. Foucher, and A. Fiala-Médioni. 1996. "Structure and Distribution of Cold Seep Communities along the Peruvian Active Margin: Relationship to Geological and Fluid Patterns." *Marine Ecology Progress Series* 132 (1–3): 109–25. https://doi.org/10.3354/meps132109.

Orme, C.D.L., D.L.J. Quicke, J.M. Cook, and A. Purvis. 2002. "Body Size Does Not Predict Species Richness among the Metazoan Phyla." *Journal of Evolutionary Biology* 15 (2): 235–47. https://doi.org/10.1046/j.1420-9101.2002.00379.x.

Orsi, W., J.F. Biddle, and V. Edgcomb. 2013. "Deep Sequencing of Subseafloor Eukaryotic RRNA Reveals Active Fungi across Marine Subsurface Provinces." Edited by López-García, P. *PLoS ONE* 8 (2): e56335. https://doi.org/10.1371/journal.pone.0056335.

Palmer, M., E. Bernhardt, E. Chornesky, S. Collins, A. Dobson, C. Duke, B. Gold, et al. 2004. "Ecology for a Crowded Planet." *Science* 304 (5675): 1251–52. https://doi.org/10.1126/science.1095780.

Pansu, J., C. Giguet-Covex, F. Ficetola, L. Gielly, F. Boyer, E. Coissac, I. Domaizon, L. Zinger, J. Poulenard, and F. Arnaud. 2015. "Environmental DNA Metabarcoding to Investigate Historic Changes in Biodiversity." *Genome* 58 (5): 264.

Pante, E., N. Puillandre, A. Viricel, S. Arnaud-Haond, D. Aurelle, M. Castelin, A. Chenuil, et al. 2015. "Species Are Hypotheses: Avoid Connectivity Assessments Based on Pillars of Sand." *Molecular Ecology* 24 (3): 525–44. https://doi.org/10.1111/mec.13048.

Pape, E., T.N. Bezerra, D.O.B.B. Jones, and A. Vanreusel. 2013. "Unravelling the Environmental Drivers of Deep-Sea Nematode Biodiversity and Its Relation with Carbon Mineralisation along a Longitudinal Primary Productivity Gradient." *Biogeosciences* 10 (5): 3127–43. https://doi.org/10.5194/bg-10-3127-2013.

Parada, A.E., D.M. Needham, and J.A. Fuhrman. 2016. "Every Base Matters: Assessing Small Subunit RRNA Primers for Marine Microbiomes with Mock Communities, Time Series and Global Field Samples." *Environ Microbiol* 18 (5): 1403–14. https://doi.org/10.1111/1462-2920.13023.

Patarnello, T., F.A.M.J. Volckaert, and R. Castilho. 2007. "Pillars of Hercules: Is the Atlantic-Mediterranean Transition a Phylogeographical Break?" *Molecular Ecology* 16 (21): 4426–44. https://doi.org/10.1111/j.1365-294X.2007.03477.x.

Pawlowski, J.W., R. Christen, B. Lecroq, D. Bachar, H.R. Shahbazkia, L. Amaral-Zettler, and

L. Guillou. 2011. "Eukaryotic Richness in the Abyss: Insights from Pyrotag Sequencing." *PLoS One* 6 (4). https://doi.org/e1816910.1371/journal.pone.0018169.

Pawlowski, J.W., P. Esling, F. Lejzerowicz, T. Cedhagen, and T.A. Wilding. 2014. "Environmental Monitoring through Protist Next-Generation Sequencing Metabarcoding: Assessing the Impact of Fish Farming on Benthic Foraminifera Communities." *Mol Ecol Resour* 14 (6): 1129–40. https://doi.org/10.1111/1755-0998.12261.

Pawlowski, J.W., P. Esling, F. Lejzerowicz, T. Cordier, J.A. Visco, C.I.M. Martins, A. Kvalvik, K. Staven, and T. Cedhagen. 2016a. "Benthic Monitoring of Salmon Farms in Norway Using Foraminiferal Metabarcoding." *Aquaculture Environment Interactions* 8: 371–86. https://doi.org/10.3354/aei00182.

Pawlowski, J.W., M. Kelly-Quinn, F. Altermatt, L. Apothéloz-Perret-Gentil, P. Beja, A. Boggero, A. Borja, et al. 2018. "The Future of Biotic Indices in the Ecogenomic Era: Integrating (e)DNA Metabarcoding in Biological Assessment of Aquatic Ecosystems." *Science of the Total Environment* 637–638 (October): 1295–1310. https://doi.org/10.1016/j.scitotenv.2018.05.002.

Pawlowski, J.W., F. Lejzerowicz, L. Apotheloz-Perret-Gentil, J. Visco, and P. Esling. 2016b. "Protist Metabarcoding and Environmental Biomonitoring: Time for Change." *European Journal of Protistology* 55: 12–25. https://doi.org/10.1016/j.ejop.2016.02.003.

Pawlowski, J.W., F. Lejzerowicz, P. Esling, L. Apotheloz-Perret-Gentil, J. V Amorim, L. Pillet, R. Vivien, and A. Cordonier. 2015. "Inferring Biotic Indices from Metabarcoding Data: Promises and Challenges." *Genome* 58 (5): 265.

Pei, A.Y., W.E. Oberdorf, C.W. Nossa, A. Agarwal, P. Chokshi, E.A. Gerz, Z. Jin, et al. 2010. "Diversity of 16S RRNA Genes within Individual Prokaryotic Genomes." *Applied and Environmental Microbiology* 76 (12): 3886–97. https://doi.org/10.1128/AEM.02953-09.

Peng, G., R. Bellerby, F. Zhang, X. Sun, and D. Li. 2020. "The Ocean's Ultimate Trashcan: Hadal Trenches as Major Depositories for Plastic Pollution." *Water Research* 168 (January). https://doi.org/10.1016/j.watres.2019.115121.

Pernice, M.C., I. Forn, A. Gomes, E. Lara, L. Alonso-Saez, J.M. Arrieta, F. del Carmen Garcia, et al. 2015a. "Global Abundance of Planktonic Heterotrophic Protists in the Deep Ocean." *Isme Journal* 9 (3): 782–92. https://doi.org/10.1038/ismej.2014.168.

Pernice, M.C., C.R. Giner, R. Logares, J. Perera-Bel, S.G. Acinas, C.M. Duarte, J.M. Gasol, and R. Massana. 2015b. "Large Variability of Bathypelagic Microbial Eukaryotic Communities across the World's Oceans." *The ISME Journal*, no. 10: 945–58. https://doi.org/10.1038/ismej.2015.170.

Pfannkuche, O., and H. Thiel. 1987. "Meiobenthic Stocks and Benthic Activity on the NE-Svalbard Shelf and in the Nansen Basin." *Polar Biology* 7 (5): 253–66. https://doi.org/10.1007/BF00443943.

Phillips, J.D., D.J. Gillis, and R.H. Hanner. 2019. "Incomplete Estimates of Genetic Diversity within Species: Implications for DNA Barcoding." *Ecology and Evolution*. John Wiley & Sons, Ltd. https://doi.org/10.1002/ece3.4757.

Pimm, S.L. 2012. "Biodiversity: Not Just Lots of Fish in the Sea." *Current Biology* 22 (23): R996–97. https://doi.org/10.1016/j.cub.2012.09.028.

Pinol, J., G. Mir, P. Gomez-Polo, and N. Agusti. 2015. "Universal and Blocking Primer Mismatches Limit the Use of High-Throughput DNA Sequencing for the Quantitative Metabarcoding of Arthropods." *Molecular Ecology Resources* 15 (4): 819–30. https://doi.org/10.1111/1755-0998.12355.

Piñol, J., M.A. Senar, and W.O.C. Symondson. 2019. "The Choice of Universal Primers and the Characteristics of the Species Mixture Determine When DNA Metabarcoding Can Be Quantitative." *Molecular Ecology* 28 (2): 407–19. https://doi.org/10.1111/mec.14776.

Plouviez, S., T.M. Shank, B. Faure, C. Daguin-Thiebaut, F. Viard, F.H. Lallier, and D. Jollivet. 2009. "Comparative Phylogeography among Hydrothermal Vent Species along the East Pacific Rise Reveals Vicariant Processes and Population Expansion in the South." *Molecular Ecology* 18 (18): 3903–17. https://doi.org/10.1111/j.1365-294X.2009.04325.x.

Pochon, X., A. Zaiko, L.M. Fletcher, O. Laroche, and S.A. Wood. 2017. "Wanted Dead or Alive? Using Metabarcoding of Environmental DNA and RNA to Distinguish Living Assemblages for Biosecurity Applications." Edited by Doi, H. *PLoS ONE* 12 (11): e0187636. https://doi.org/10.1371/journal.pone.0187636.

Pomerantz, A., N. Peñafiel, A. Arteaga, L. Bustamante, F. Pichardo, L.A. Coloma, C.L. Barrio-Amorós, D. Salazar-Valenzuela, and S. Prost. 2018. "Real-Time DNA Barcoding in a Rainforest Using Nanopore Sequencing: Opportunities for Rapid Biodiversity Assessments and Local Capacity Building." *GigaScience* 7 (4). https://doi.org/10.1093/gigascience/giy033.

Pont, D., M. Rocle, A. Valentini, R. Civade, P. Jean, A. Maire, N. Roset, M. Schabuss, H. Zornig, and T. Dejean. 2018. "Environmental DNA Reveals Quantitative Patterns of Fish Biodiversity in Large Rivers despite Its Downstream Transportation." *Scientific Reports* 8 (1): 10361. https://doi.org/10.1038/s41598-018-28424-8.

Poulin, R., and G. Pérez-Ponce de León. 2017. "Global Analysis Reveals That Cryptic Diversity Is Linked with Habitat but Not Mode of Life." *Journal of Evolutionary Biology* 30 (3): 641–49. https://doi.org/10.1111/jeb.13034.

Probandt, D., K. Knittel, H.E. Tegetmeyer, S. Ahmerkamp, M. Holtappels, and R. Amann. 2017. "Permeability Shapes Bacterial Communities in Sublittoral Surface Sediments." *Environmental Microbiology* 19 (4): 1584–99. https://doi.org/10.1111/1462-2920.13676.

Purvis, A., and A. Hector. 2000. "Getting the Measure of Biodiversity." *Nature* 405 (6783). https://www.nature.com/nature/journal/v405/n6783/pdf/405212a0.pdf.

Pusceddu, A., C. Gambi, D. Zeppilli, S. Bianchelli, and R. Danovaro. 2009. "Organic Matter Composition, Metazoan Meiofauna and Nematode Biodiversity in Mediterranean Deep-Sea Sediments." *Deep-Sea Research Part II: Topical Studies in Oceanography* 56 (11–12): 755–62. https://doi.org/10.1016/j.dsr2.2008.10.012.

Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F.O. Glöckner. 2012. "The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools." *Nucleic Acids Research* 41 (D1): D590–96. https://doi.org/10.1093/nar/gks1219.

Queiroz, K. de. 2005. "Ernst Mayr and the Modern Concept of Species." *Proceedings of the National Academy of Sciences* 102 (Supplement 1): 6600–6607. https://doi.org/10.1073/pnas.0502030102.

R. Hessler, R., and H. L. Sanders. 1967. "Faunal Diversity in the Deep-Sea." *Deep-Sea Research and Oceanographic Abstracts* 14 (1). https://doi.org/10.1016/0011-7471(67)90029-0.

R Core Team. 2018. "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/.

Ramirez-Llodra, E., A. Brandt, R. Danovaro, B. De Mol, E. Escobar, C.R. German, L.A. Levin, et al. 2010. "Deep, Diverse and Definitely Different: Unique Attributes of the World's Largest Ecosystem." *Biogeosciences* 7 (9): 2851–99. https://doi.org/10.5194/bg-7-2851-2010.

Ramirez-Llodra, E., P.A. Tyler, M.C. Baker, O.A. Bergstad, M.R. Clark, E. Escobar, L.A. Levin, et al. 2011. "Man and the Last Great Wilderness: Human Impact on the Deep Sea." Edited by Roopnarine, P. *PLoS ONE* 6 (7): e22588. https://doi.org/10.1371/journal.pone.0022588.

Ramírez, G.A., S.L. Jørgensen, R. Zhao, and S. D'Hondt. 2018. "Minimal Influence of Extracellular DNA on Molecular Surveys of Marine Sedimentary Communities." *Frontiers in Microbiology* 9 (December): 2969. https://doi.org/10.3389/fmicb.2018.02969.

Renaud, P.E., W.G. Ambrose, A. Vanreusel, and L.M. Clough. 2006. "Nematode and Macrofaunal Diversity in Central Arctic Ocean Benthos." In *Journal of Experimental Marine Biology and Ecology*, 330:297–306. https://doi.org/10.1016/j.jembe.2005.12.035.

Rex, M.A. 1981. "Community Structure in the Deep-Sea Benthos." *Annual Review of Ecology and Systematics* 12 (1): 331–53. https://doi.org/10.1146/annurev.es.12.110181.001555.

Rex, M.A., R.J. Etter, J.S. Morris, J. Crouse, C.R. McClain, N.A. Johnson, C.T. Stuart, J.W. Deming, R. Thies, and R. Avery. 2006. "Global Bathymetric Patterns of Standing Stock and Body Size in the Deep-Sea Benthos." *Marine Ecology Progress Series* 317 (July): 1–8. https://doi.org/10.3354/meps317001.

Rex, M.A., C.R. McClain, N.A. Johnson, R.J. Etter, J.A. Allen, P. Bouchet, and A. Warén. 2005. "A Source-Sink Hypothesis for Abyssal Biodiversity." *The American Naturalist* 165 (2): 163–78. https://doi.org/10.1086/427226.

Rex, M.A., C.T. Stuart, R.R. Hessler, J.A. Allen, H.L. Sanders, and G.D.F. Wilson. 1993. "Global-Scale Latitudinal Patterns of Species Diversity in the Deep-Sea Benthos." *Nature* 365 (6447): 636–39. https://doi.org/10.1038/365636a0.

Risso, A. 1816. *Histoire Naturelle Des Crustaces Des Environs de Nice.*

Ritari, J., J. Salojärvi, L. Lahti, and W.M. de Vos. 2015. "Improved Taxonomic Assignment of Human Intestinal 16S RRNA Sequences by a Dedicated Reference Database." *BMC Genomics* 16 (1): 1056. https://doi.org/10.1186/s12864-015-2265-y.

Rodríguez-Martínez, R., G. Leonard, D.S. Milner, S. Sudek, M. Conway, K. Moore, T. Hudson, et al. 2020. "Controlled Sampling of Ribosomally Active Protistan Diversity in Sediment-Surface Layers Identifies Putative Players in the Marine Carbon Sink." *ISME Journal* 14 (4): 984–98. https://doi.org/10.1038/s41396-019-0581-y.

Rognes, T., T. Flouri, B. Nichols, C. Quince, and F. Mahé. 2016. "VSEARCH: A Versatile Open Source Tool for Metagenomics." *PeerJ* 4 (October): e2584. https://doi.org/10.7717/peerj.2584.

Rohr, J.R., D.J. Civitello, F.W. Halliday, P.J. Hudson, K.D. Lafferty, C.L. Wood, and E.A. Mordecai. 2020. "Towards Common Ground in the Biodiversity–Disease Debate." *Nature Ecology and Evolution*. Nature Research. https://doi.org/10.1038/s41559-019-1060-6.

Rosli, N., D. Leduc, A.A. Rowden, M.R. Clark, P.K. Probert, K. Berkenbusch, and C. Neira. 2016. "Differences in Meiofauna Communities with Sediment Depth Are Greater than Habitat Effects on the New Zealand Continental Margin: Implications for Vulnerability to Anthropogenic Disturbance." *PeerJ* 4 (July): e2154. https://doi.org/10.7717/peerj.2154.

Rosli, N., D. Leduc, A.A. Rowden, and P.K. Probert. 2017. "Review of Recent Trends in Ecological Studies of Deep-Sea Meiofauna, with Focus on Patterns and Processes at Small to Regional Spatial Scales." *Marine Biodiversity*, 1–22. https://doi.org/10.1007/s12526-017-0801-5.

Rossberg, A.G., T. Rogers, and A.J. Mckane. 2013. "Are There Species Smaller than 1 Mm?" *Proceedings of the Royal Society B: Biological Sciences* 280 (1767): 20131248. https://doi.org/10.1098/rspb.2013.1248.

Rosselló-Mora, R., and R. Amann. 2001. "The Species Concept for Prokaryotes." *FEMS Microbiology Reviews* 25 (1): 39–67. https://doi.org/10.1111/j.1574-6976.2001.tb00571.x.

Roussel, E., C. Konn, J. Charlou, J. Donval, J. Querellou, D. Prieur, M.C. Bonavita, et al. 2011. "Comparison of Microbial Communities Associated with Three Atlantic Ultramafic

Hydrothermal Systems." *FEMS Microbiology Ecology* 77 (3): 647–65. https://doi.org/10.1111/j.1574-6941.2011.01161.x . hal-01145008.

Saeedi, H., J.D. Reimer, M.I. Brandt, P.O. Dumais, A.M. Jażdżewska, N.W. Jeffery, P.M. Thielen, and M.J. Costello. 2019. "Global Marine Biodiversity in the Context of Achieving the Aichi Targets: Ways Forward and Addressing Data Gaps." *PeerJ* 7 (10): e7221. https://doi.org/10.7717/peerj.7221.

Salazar, G., F.M. Cornejo-Castillo, V. Benitez-Barrios, E. Fraile-Nuez, X.A. Alvarez-Salgado, C.M. Duarte, J.M. Gasol, and S.G. Acinas. 2016. "Global Diversity and Biogeography of Deep-Sea Pelagic Prokaryotes." *Isme Journal* 10 (3): 596–608. https://doi.org/10.1038/ismej.2015.137.

Sassoubre, L.M., K.M. Yamahara, L.D. Gardner, B.A. Block, and A.B. Boehm. 2016. "Quantification of Environmental DNA (EDNA) Shedding and Decay Rates for Three Marine Fish." *Environmental Science and Technology* 50 (19): 10456–64. https://doi.org/10.1021/acs.est.6b03114.

Sawaya, N.A., A. Djurhuus, C.J. Closek, M. Hepner, E. Olesin, L. Visser, C. Kelble, K. Hubbard, and M. Breitbart. 2019. "Assessing Eukaryotic Biodiversity in the Florida Keys National Marine Sanctuary through Environmental DNA Metabarcoding." *Ecology and Evolution* 9 (3): 1029–40. https://doi.org/10.1002/ece3.4742.

Schabacker, J.C., S.J. Amish, B.K. Ellis, B. Gardner, D.L. Miller, E.A. Rutledge, A.J. Sepulveda, and G. Luikart. 2020. "Increased EDNA Detection Sensitivity Using a Novel High-volume Water Sampling Method." *Environmental DNA*, edn3.63. https://doi.org/10.1002/edn3.63.

Schmidt, C., and P. Martínez Arbizu. 2015. "Unexpectedly Higher Metazoan Meiofauna Abundances in the Kuril-Kamchatka Trench Compared to the Adjacent Abyssal Plains." *Deep-Sea Research Part II: Topical Studies in Oceanography* 111 (January): 60–75. https://doi.org/10.1016/j.dsr2.2014.08.019.

Schnell, I.B., K. Bohmann, and M.T.P. Gilbert. 2015. "Tag Jumps Illuminated - Reducing Sequence-to-Sample Misidentifications in Metabarcoding Studies." *Molecular Ecology Resources* 15 (6): 1289–1303. https://doi.org/10.1111/1755-0998.12402.

Schratzberger, M., and J. Ingels. 2017. "Meiofauna Matters: The Roles of Meiofauna in Benthic Ecosystems." *Journal of Experimental Marine Biology and Ecology*, 2017. https://doi.org/10.1016/j.jembe.2017.01.007.

Shank, T.M., M.B. Black, K.M. Halanych, R.A. Lutz, and R.C. Vrijenhoek. 1999. "Miocene Radiation of Deep-Sea Hydrothermal Vent Shrimp (Caridea: Bresiliidae): Evidence from Mitochondrial Cytochrome Oxidase Subunit I." *Molecular Phylogenetics and Evolution* 13 (2): 244–54. https://doi.org/10.1006/mpev.1999.0642.

Shapiro, B.J., J.B. Leducq, and J. Mallet. 2016. "What Is Speciation?" Edited by Matic, I. *PLoS Genetics* 12 (3): e1005860. https://doi.org/10.1371/journal.pgen.1005860.

Shearer, T.L., and M.A. Coffroth. 2008. "Barcoding Corals: Limited by Interspecific Divergence, Not Intraspecific Variation." *Molecular Ecology Resources* 8 (2): 247–55. https://doi.org/10.1111/j.1471-8286.2007.01996.x.

Shearer, T.L., M.J.H. Van Oppen, S.L. Romano, and G. Wörheide. 2002. "Slow Mitochondrial DNA Sequence Evolution in the Anthozoa (Cnidaria)." *Molecular Ecology*. https://doi.org/10.1046/j.1365-294X.2002.01652.x.

Shepherd, E., E.J. Milner-Gulland, A.T. Knight, M.A. Ling, S. Darrah, A. van Soesbergen, and N.D. Burgess. 2016. "Status and Trends in Global Ecosystem Services and Natural Capital: Assessing Progress Toward Aichi Biodiversity Target 14." *Conservation Letters* 9 (6): 429–37. https://doi.org/10.1111/conl.12320.

Shirayama, Y. 1984. "Vertical Distribution of Meiobenthos in the Sediment Profile in Bathyal,

Abyssal and Hadal Deep Sea Systems of the Western Pacific." *Oceanological Acta* 7 (1): 123–29. http://archimer.ifremer.fr/doc/00113/22413/20110.pdf.

Sibuet, M., and K. Olu. 1998. "Biogeography, Biodiversity and Fluid Dependence of Deep-Sea Cold-Seep Communities at Active and Passive Margins." *Deep-Sea Research Part II: Topical Studies in Oceanography* 45 (1–3): 517–67. https://doi.org/10.1016/S0967-0645(97)00074-X.

Simpson, G.G. 1951. "The Species Conept." *Evolution* 5 (4): 285–98. https://doi.org/10.1111/j.1558-5646.1951.tb02788.x.

Singer, G.A.C., N.A. Fahner, J.G. Barnes, A. McCarthy, and M. Hajibabaei. 2019. "Comprehensive Biodiversity Analysis via Ultra-Deep Patterned Flow Cell Technology: A Case Study of EDNA Metabarcoding Seawater." *Scientific Reports* 9 (1): 1–12. https://doi.org/10.1038/s41598-019-42455-9.

Sinniger, F., J.W. Pawlowski, S. Harii, A.J. Gooday, H. Yamamoto, P. Chevaldonné, T. Cedhagen, G. Carvalho, and S. Creer. 2016. "Worldwide Analysis of Sedimentary DNA Reveals Major Gaps in Taxonomic Knowledge of Deep-Sea Benthos." *Frontiers in Marine Science* 3 (June): 92. https://doi.org/10.3389/FMARS.2016.00092.

Slon, V., C. Hopfe, C.L. Weiß, F. Mafessoni, M. De La Rasilla, C. Lalueza-Fox, A. Rosas, et al. 2017. "Neandertal and Denisovan DNA from Pleistocene Sediments." *Science* 356 (6338): 605–8. https://doi.org/10.1126/science.aam9695.

Smet, B. De, E. Pape, T. Riehl, P. Bonifácio, L. Colson, and A. Vanreusel. 2017. "The Community Structure of Deep-Sea Macrofauna Associated with Polymetallic Nodules in the Eastern Part of the Clarion-Clipperton Fracture Zone." *Frontiers in Marine Science* 4 (April): 103. https://doi.org/10.3389/fmars.2017.00103.

Smith, C.R., and A.R. Baco. 2003. "Ecology of Whale Falls at the Deep-Sea Floor." In *Oceanography and Marine Biology: An Annual Review*, edited by Gibson, R.N. and Atkinson, R.J.A., 41:311–54. Taylor & Francis Group.

Smith, C.R., J. Drazen, and S.L. Mincks. 2006. "Deep-Sea Biodiversity and Biogeography : Perspectives from the Abyss." *International Seabed Authority Seamount Biodiversity Symposium*, no. March: 1–13.

Smith, C.R., F.C. De Leo, A.F. Bernardino, A.K. Sweetman, and P.M. Arbizu. 2008. "Abyssal Food Limitation, Ecosystem Structure and Climate Change." *Trends in Ecology and Evolution* 23 (9): 518–28. https://doi.org/10.1016/j.tree.2008.05.002.

Smith, C.R., and P.V.R. Snelgrove. 2002. "A Riot of Species in An Environmental Calm." In , 311–42. CRC Press. https://doi.org/10.1201/9780203180594.ch6.

Smith, K.L., H.A. Ruhl, B.J. Bett, D.S.M. Billett, R.S. Lampitt, and R.S. Kaufmann. 2009. "Climate, Carbon Cycling, and Deep-Ocean Ecosystems." *Proceedings of the National Academy of Sciences of the United States of America* 106 (46): 19211–18. https://doi.org/10.1073/pnas.0908322106.

Smith, K.L., H.A. Ruhl, M. Kahru, C.L. Huffard, and A.D. Sherman. 2013. "Deep Ocean Communities Impacted by Changing Climate over 24 y in the Abyssal Northeast Pacific Ocean." *Proceedings of the National Academy of Sciences of the United States of America* 110 (49): 19838–41. https://doi.org/10.1073/pnas.1315447110.

Snelgrove, P.V.R. 1999. "Getting to the Bottom of Marine Biodiversity: Sedimentary Habitats." *BioScience* 49 (2): 129. https://doi.org/10.2307/1313538.

Snider, L.J., B.R. Burnett, and R.R. Hessler. 1984. "The Composition and Distribution of Meiofauna and Nanobiota in a Central North Pacific Deep-Sea Area." *Deep Sea Research Part A, Oceanographic Research Papers* 31 (10): 1225–49. https://doi.org/10.1016/0198-0149(84)90059-1.

Soetaert, K., J. Vanaverbeke, C. Heip, P.M.J. Herman, J.J. Middelburg, A. Sandee, and G.

Duineveld. 1997. "Nematode Distribution in Ocean Margin Sediments of the Goban Spur (Northeast Atlantic) in Relation to Sediment Geochemistry." *Deep-Sea Research Part I: Oceanographic Research Papers* 44 (9–10): 1671–83. https://doi.org/10.1016/S0967-0637(97)00043-5.

Sokal, R.R., and T.J. Crovello. 1970. "The Biological Species Concept : A Critical Evaluation." *The American Naturalist* 104 (936): 127–53.

Soltwedel, T. 2000. "Metazoan Meiobenthos along Continental Margins: A Review." *Progress in Oceanography* 46 (1): 59–84. https://doi.org/10.1016/S0079-6611(00)00030-6.

Song, H., J.E. Buhay, M.F. Whiting, and K.A. Crandall. 2008. "Many Species in One: DNA Barcoding Overestimates the Number of Species When Nuclear Mitochondrial Pseudogenes Are Coamplified." *Proceedings of the National Academy of Sciences of the United States of America* 105 (36): 13486–91. https://doi.org/10.1073/pnas.0803076105.

Stat, M., M.J. Huggett, R. Bernasconi, J.D. Dibattista, T.E. Berry, S.J. Newman, E.S. Harvey, and M. Bunce. 2017. "Ecosystem Biomonitoring with EDNA: Metabarcoding across the Tree of Life in a Tropical Marine Environment." *Scientific Reports* 7. https://doi.org/10.1038/s41598-017-12501-5.

Staudigel, H., and D.A. Clague. 2010. "The Geological History of Deep-Sea Volcanoes: Biosphere, Hydrosphere, and Lithosphere Interactions." *Oceanography* 23 (1): 58–71. https://doi.org/10.5670/oceanog.2010.62.

Stefanni, S., D. Stanković, D. Borme, A. de Olazabal, T. Juretić, A. Pallavicini, and V. Tirelli. 2018. "Multi-Marker Metabarcoding Approach to Study Mesozooplankton at Basin Scale." *Scientific Reports* 8 (1): 12085. https://doi.org/10.1038/s41598-018-30157-7.

Stefanoudis, P. V., B.J. Bett, and A.J. Gooday. 2016. "Abyssal Hills: Influence of Topography on Benthic Foraminiferal Assemblages." *Progress in Oceanography* 148 (November): 44–55. https://doi.org/10.1016/j.pocean.2016.09.005.

Steyaert, M., J. Vanaverbeke, A. Vanreusel, C. Barranguet, C. Lucas, and M. Vincx. 2003. "The Importance of Fine-Scale, Vertical Profiles in Characterising Nematode Community Structure." *Estuarine, Coastal and Shelf Science* 58 (2): 353–66. https://doi.org/10.1016/S0272-7714(03)00086-6.

Stoeck, T., D. Bass, M. Nebel, R. Christen, M.D.M. Jones, H.W. Breiner, and T.A. Richards. 2010. "Multiple Marker Parallel Tag Environmental DNA Sequencing Reveals a Highly Complex Eukaryotic Community in Marine Anoxic Water." *Molecular Ecology* 19 (SUPPL. 1): 21–31. https://doi.org/10.1111/j.1365-294X.2009.04480.x.

Stokke, R., H. Dahle, I. Roalkvam, J. Wissuwa, F.L. Daae, A. Tooming-Klunderud, I.H. Thorseth, R.B. Pedersen, and I.H. Steen. 2015. "Functional Interactions among Filamentous Epsilonproteobacteria and Bacteroidetes in a Deep-Sea Hydrothermal Vent Biofilm." *Environmental Microbiology* 17 (10): 4063–77. https://doi.org/10.1111/1462-2920.12970.

Stokstad, E. 2005. "Taking the Pulse of Earth's Life-Support Systems." *Science* 308 (5718). http://science.sciencemag.org/content/308/5718/41.full.

Stork, N.E. 1993. "How Many Species Are There?" *Biodiversity and Conservation* 2 (3): 215–32. https://doi.org/10.1007/BF00056669.

Sunagawa, S., L.P. Coelho, S. Chaffron, J.R. Kultima, K. Labadie, G. Salazar, B. Djahanschiri, et al. 2015. "Structure and Function of the Global Ocean Microbiome." *Science* 348 (6237). https://doi.org/10.1126/science.1261359.

Taberlet, P., E. Coissac, M. Hajibabaei, and L.H. Rieseberg. 2012a. "Environmental DNA." *Molecular Ecology* 21 (8): 1789–93. https://doi.org/10.1111/j.1365-294X.2012.05542.x.

Taberlet, P., S.M. Prud'Homme, E. Campione, J. Roy, C. Miquel, W. Shehzad, L. Gielly, et al. 2012b. "Soil Sampling and Isolation of Extracellular DNA from Large Amount of Starting

Material Suitable for Metabarcoding Studies." *Molecular Ecology* 21 (8): 1816–20. https://doi.org/10.1111/j.1365-294X.2011.05317.x.

Tanabe, A.S., S. Nagai, K. Hida, M. Yasuike, A. Fujiwara, Y. Nakamura, Y. Takano, and S. Katakura. 2016. "Comparative Study of the Validity of Three Regions of the 18S-RRNA Gene for Massively Parallel Sequencing-Based Monitoring of the Planktonic Eukaryote Community." *Molecular Ecology Resources* 16 (2): 402–14. https://doi.org/10.1111/1755-0998.12459.

Tang, C.Q., F. Leasi, U. Obertegger, A. Kieneke, T.G. Barraclough, and D. Fontaneto. 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 (40): 16208–12. https://doi.org/10.1073/pnas.1209160109.

Tedersoo, L., M. Bahram, T. Cajthaml, S. Polme, I. Hiiesalu, S. Anslan, H. Harend, et al. 2016. "Tree Diversity and Species Identity Effects on Soil Fungi, Protists and Animals Are Context Dependent." *Isme Journal* 10 (2): 346–62. https://doi.org/10.1038/ismej.2015.116.

Teixeira, S., K. Olu, C. Decker, R.L. Cunha, S. Fuchs, S. Hourdez, E.A. Serrão, and S. Arnaud-Haond. 2013. "High Connectivity across the Fragmented Chemosynthetic Ecosystems of the Deep Atlantic Equatorial Belt: Efficient Dispersal Mechanisms or Questionable Endemism?" *Molecular Ecology* 22 (18): 4663–80. https://doi.org/10.1111/mec.12419.

Teixeira, S., E.A. Serrão, and S. Arnaud-Haond. 2012. "Panmixia in a Fragmented and Unstable Environment: The Hydrothermal Shrimp Rimicaris Exoculata Disperses Extensively along the Mid-Atlantic Ridge." Edited by López-García, P. *PLoS ONE* 7 (6): e38521. https://doi.org/10.1371/journal.pone.0038521.

Tempestini, A., S. Rysgaard, and F. Dufresne. 2018. "Species Identification and Connectivity of Marine Amphipods in Canada's Three Oceans." Edited by Cimmaruta, R. *PLoS ONE* 13 (5): e0197174. https://doi.org/10.1371/journal.pone.0197174.

Thiel, H. 1983. "Meiobenthos and Nanobenthos of the Deep-Sea." In *The Sea, Vol. 8*, edited by Rowe, G.T., 167–230. John Wiley and Sons.

Thomas, A.C., J. Howard, P.L. Nguyen, T.A. Seimon, and C.S. Goldberg. 2018. "ANDe™: A Fully Integrated Environmental DNA Sampling System." *Methods in Ecology and Evolution* 9 (6): 1379–85. https://doi.org/10.1111/2041-210X.12994.

Tietjen, J.H. 1992. "Abundance and Biomass of Metazoan Meiobenthos in the Deep Sea." In *Deep-Sea Food Chains and the Global Carbon Cycle*, 45–62. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-011-2452-2_4.

Tittensor, D.P., C. Mora, W. Jetz, H.K. Lotze, D. Ricard, E. Vanden Berghe, and B. Worm. 2010. "Global Patterns and Predictors of Marine Biodiversity across Taxa." *Nature* 466 (7310): 1098–1101. https://doi.org/http://www.nature.com/nature/journal/v466/n7310/abs/nature09329.html#supplementary-information.

Torti, A., M.A. Lever, and B.B. Jørgensen. 2015. "Origin, Dynamics, and Implications of Extracellular DNA Pools in Marine Sediments." *Marine Genomics*. https://doi.org/10.1016/j.margen.2015.08.007.

Tuomisto, H. 2010. "A Consistent Terminology for Quantifying Species Diversity? Yes, It Does Exist." *Oecologia* 164 (4): 853–60. https://doi.org/10.1007/s00442-010-1812-0.

Turon, X., A. Antich, C. Palacín, K. Præbel, and O.S. Wangensteen. 2020. "From Metabarcoding to Metaphylogeography: Separating the Wheat from the Chaff." *Ecological Applications* 30 (2). https://doi.org/10.1002/eap.2036.

Vacher, C., A. Tamaddoni-Nezhad, S. Kamenova, N. Peyrard, Y. Moalic, R. Sabbadin, L.

Schwaller, et al. 2016. “Learning Ecological Networks from Next-Generation Sequencing Data.” In *Advances in Ecological Research*, 54:1–39. Academic Press Inc. https://doi.org/10.1016/bs.aecr.2015.10.004.

Valen, L. Van. 1976. “Ecological Species, Multispecies, and Oaks.” *Taxon* 25 (2–3): 233–39. https://doi.org/10.2307/1219444.

Valentini, A., F. Pompanon, and P. Taberlet. 2009. “DNA Barcoding for Ecologists.” *Trends in Ecology and Evolution*. Elsevier Current Trends. https://doi.org/10.1016/j.tree.2008.09.011.

Valentini, A., P. Taberlet, C. Miaud, R.R. Civade, J. Herder, P.F. Thomsen, E. Bellemain, et al. 2016. “Next-Generation Monitoring of Aquatic Biodiversity Using Environmental DNA Metabarcoding.” *Molecular Ecology* 25 (4): 929–42. https://doi.org/10.1111/mec.13428.

Vanreusel, A., A.C. Andersen, A. Boetius, D. Connelly, M.R. Cunha, C. Decker, A. Hilario, et al. 2009. “Biodiversity of Cold Seep Ecosystems along the European Margins.” *Oceanography* 22 (1): 110–27. https://doi.org/10.5670/oceanog.2009.12.

Vanreusel, A., G. Fonseca, R. Danovaro, M.C. Da Silva, A.M. Esteves, T. Ferrero, G. Gad, et al. 2010a. “The Contribution of Deep-Sea Macrohabitat Heterogeneity to Global Nematode Diversity.” *Marine Ecology* 31 (1): 6–20. https://doi.org/10.1111/j.1439-0485.2009.00352.x.

Vanreusel, A., A. de Groote, S. Gollner, and M. Bright. 2010b. “Ecology and Biogeography of Free-Living Nematodes Associated with Chemosynthetic Environments in the Deep Sea: A Review.” Edited by Unsworth, R.K.F. *PLoS ONE* 5 (8): e12449. https://doi.org/10.1371/journal.pone.0012449.

Vargas, C., S. Audic, N. Henry, J. Decelle, F. Mahé, R. Logares, E. Lara, et al. 2015. “Eukaryotic Plankton Diversity in the Sunlit Ocean.” *Science* 348 (6237). https://doi.org/10.1126/science.1261605.

Visco, J.A., L. Apotheloz-Perret-Gentil, A. Cordonier, P. Esling, L. Pillet, and J.W. Pawlowski. 2015. “Environmental Monitoring: Inferring the Diatom Index from Next-Generation Sequencing Data.” *Environmental Science & Technology* 49 (13): 7597–7605. https://doi.org/10.1021/es506158m.

Vivien, R., S. Wyler, M. Lafont, and J.W. Pawlowski. 2015. “Molecular Barcoding of Aquatic Oligochaetes: Implications for Biomonitoring.” *PLoS One* 10 (4). https://doi.org/10.1371/journal.pone.0125485.

Wangensteen, O.S., and X. Turon. 2016. “Metabarcoding Techniques for Assessing Biodiversity of Marine Animal Forests.” In *Marine Animal Forests*, edited by Rossi, S., Bramanti, L., Gori, A., and Orejas Saco del Valle, C., 1–29. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-17001-5_53-1.

Washington, H.G. 1984. “Diversity, Biotic and Similarity Indicies. A Review with Specieal Relevence to Aquatic Systems.” *Water Research* 18 (6): 653–94.

Watling, L., J. Guinotte, M.R. Clark, and C.R. Smith. 2013. “A Proposed Biogeography of the Deep Ocean Floor.” *Progress in Oceanography* 111: 91–112. https://doi.org/10.1016/j.pocean.2012.11.003.

Weber, A.A.-T., and J.W. Pawlowski. 2013. “Can Abundance of Protists Be Inferred from Sequence Data: A Case Study of Foraminifera.” Edited by López-García, P. *PLoS ONE* 8 (2): e56739. https://doi.org/10.1371/journal.pone.0056739.

Webster, G., C.J. Newberry, J.C. Fry, and A.J. Weightman. 2003. “Assessment of Bacterial Community Structure in the Deep Sub-Seafloor Biosphere by 16S RDNA-Based Techniques: A Cautionary Tale.” *Journal of Microbiological Methods* 55 (1): 155–64. https://doi.org/10.1016/S0167-7012(03)00140-4.

Whittaker, R.H. 1960. “Vegetation of the Siskiyou Mountains, Oregon and California.”

Ecological Monographs 30 (3): 279–338. https://doi.org/10.2307/1943563.

Wilcox, T.M., K.E. Zarn, M.P. Piggott, M.K. Young, K.S. McKelvey, and M.K. Schwartz. 2018. "Capture Enrichment of Aquatic Environmental DNA: A First Proof of Concept." *Molecular Ecology Resources* 18 (6): 1392–1401. https://doi.org/10.1111/1755-0998.12928.

Wilson, M. 2007. *Igneous Petrogenesis: A Global Tectonic Approach*. 1st ed. Springer.

Wood, S.A., X. Pochon, O. Laroche, U. von Ammon, J. Adamson, and A. Zaiko. 2019. "A Comparison of Droplet Digital Polymerase Chain Reaction (PCR), Quantitative PCR and Metabarcoding for Species-Specific Detection in Environmental DNA." *Molecular Ecology Resources* 19 (6): 1407–19. https://doi.org/10.1111/1755-0998.13055.

Woods, D.R., and J.H. Tietjen. 1985. "Horizontal and Vertical Distribution of Meiofauna on Sandy Beaches." *Marine Geology* 68: 233–41. https://ac.els-cdn.com/0025322785900143/1-s2.0-0025322785900143-main.pdf?_tid=c787c266-d032-11e7-ad14-00000aacb361&acdnat=1511430397_6fcfdeee79df5a23c1d71a18437fd6fb.

Woolley, S.N.C., D.P. Tittensor, P.K. Dunstan, G. Guillera-Arroita, J.J. Lahoz-Monfort, B.A. Wintle, B. Worm, and T.D. O'Hara. 2016. "Deep-Sea Diversity Patterns Are Shaped by Energy Availability." *Nature* 533: 393–96. https://doi.org/10.1038/nature17937.

Worm, B., E.B. Barbier, N. Beaumont, J.E. Duffy, C. Folke, B.S. Halpern, J.B.C. Jackson, et al. 2006. "Impacts of Biodiversity Loss on Ocean Ecosystem Services." *Science* 314 (5800).

Xiong, W., and A. Zhan. 2018. "Testing Clustering Strategies for Metabarcoding-Based Investigation of Community–Environment Interactions." *Molecular Ecology Resources* 18 (6): 1326–38. https://doi.org/10.1111/1755-0998.12922.

Yarza, P., P. Yilmaz, E. Pruesse, F.O. Glöckner, W. Ludwig, K.H. Schleifer, W.B. Whitman, J. Euzéby, R. Amann, and R. Rosselló-Móra. 2014. "Uniting the Classification of Cultured and Uncultured Bacteria and Archaea Using 16S RRNA Gene Sequences." *Nature Reviews Microbiology* 12 (9): 635–45. https://doi.org/10.1038/nrmicro3330.

Yoccoz, N.G., K.A. Brathen, L. Gielly, J. Haile, M.E. Edwards, T. Goslar, H. von Stedingk, et al. 2012. "DNA from Soil Mirrors Plant Taxonomic and Growth Form Diversity." *Molecular Ecology* 21 (15): 3647–55. https://doi.org/10.1111/j.1365-294X.2012.05545.x.

Yu, D.W., Y. Ji, B.C. Emerson, X. Wang, C. Ye, C. Yang, and Z. Ding. 2012. "Biodiversity Soup: Metabarcoding of Arthropods for Rapid Biodiversity Assessment and Biomonitoring." *Methods in Ecology and Evolution* 3 (4): 613–23. https://doi.org/10.1111/j.2041-210X.2012.00198.x.

Zanol, J., K.M. Halanych, T.H. Struck, and K. Fauchald. 2010. "Phylogeny of the Bristle Worm Family Eunicidae (Eunicida, Annelida) and the Phylogenetic Utility of Noncongruent 16S, COI and 18S in Combined Analyses." *Molecular Phylogenetics and Evolution* 55 (2): 660–76. https://doi.org/10.1016/j.ympev.2009.12.024.

Zeppilli, D., L. Bongiorni, A. Cattaneo, R. Danovaro, and R.S. Santos. 2013. "Meiofauna Assemblages of the Condor Seamount (North-East Atlantic Ocean) and Adjacent Deep-Sea Sediments." *Deep-Sea Research Part II: Topical Studies in Oceanography* 98 (PA): 87–100. https://doi.org/10.1016/j.dsr2.2013.08.009.

Zeppilli, D., L. Bongiorni, R.S. Santos, and A. Vanreusel. 2014. "Changes in Nematode Communities in Different Physiographic Sites of the Condor Seamount (North-East Atlantic Ocean) and Adjacent Sediments." Edited by Bianchi, C.N. *PLoS ONE* 9 (12): e115601. https://doi.org/10.1371/journal.pone.0115601.

Zeppilli, D., M. Canals, and R. Danovaro. 2012. "Pockmarks Enhance Deep-Sea Benthic Biodiversity: A Case Study in the Western Mediterranean Sea." *Diversity and*

Distributions 18 (8): 832–46. https://doi.org/10.1111/j.1472-4642.2011.00859.x.

Zeppilli, D., D. Leduc, C. Fontanier, D. Fontaneto, S. Fuchs, A.J. Gooday, A. Goineau, et al. 2018. "Characteristics of Meiofauna in Extreme Marine Ecosystems: A Review." *Marine Biodiversity* 48 (1): 35–71. https://doi.org/10.1007/s12526-017-0815-z.

Zeppilli, D., M. Mea, C. Corinaldesi, and R. Danovaro. 2011. "Mud Volcanoes in the Mediterranean Sea Are Hot Spots of Exclusive Meiobenthic Species." *Progress in Oceanography* 91 (3): 260–72. https://doi.org/10.1016/j.pocean.2011.01.001.

Zeppilli, D., A. Pusceddu, F. Trincardi, and R. Danovaro. 2016. "Seafloor Heterogeneity Influences the Biodiversity-Ecosystem Functioning Relationships in the Deep Sea." *Scientific Reports* 6 (1): 26352. https://doi.org/10.1038/srep26352.

Zhan, A., S.A. Bailey, D.D. Heath, and H.J. Macisaac. 2014. "Performance Comparison of Genetic Markers for High-Throughput Sequencing-Based Biodiversity Assessment in Complex Communities." *Molecular Ecology Resources* 14 (5): 1049–59. https://doi.org/10.1111/1755-0998.12254.

Zhang, J., Q.L. Sun, Z.G. Zeng, S. Chen, and L. Sun. 2015. "Microbial Diversity in the Deep-Sea Sediments of Iheya North and Iheya Ridge, Okinawa Trough." *Microbiological Research* 177 (August): 43–52. https://doi.org/10.1016/j.micres.2015.05.006.

Zhang, L., M. Kang, J. Xu, J. Xu, Y. Shuai, X. Zhou, Z. Yang, and K. Ma. 2016. "Bacterial and Archaeal Communities in the Deep-Sea Sediments of Inactive Hydrothermal Vents in the Southwest India Ridge." *Scientific Reports* 6 (May). https://doi.org/10.1038/srep25982.

Zhao, F., S. Filker, T. Stoeck, and K. Xu. 2017. "Ciliate Diversity and Distribution Patterns in the Sediments of a Seamount and Adjacent Abyssal Plains in the Tropical Western Pacific Ocean." *BMC Microbiology* 17 (1): 192. https://doi.org/10.1186/s12866-017-1103-6.

Zhao, F., S. Filker, K. Xu, P. Huang, and S. Zheng. 2020. "Microeukaryote Communities Exhibit Phyla-Specific Distance-Decay Patterns and an Intimate Link between Seawater and Sediment Habitats in the Western Pacific Ocean." *Deep-Sea Research Part I: Oceanographic Research Papers* 160: 103279. https://doi.org/10.1016/j.dsr.2020.103279.

Zinger, L., L.A. Amaral-Zettler, J.A. Fuhrman, M.C. Horner-Devine, S.M. Huse, D.B.M. Welch, J.B.H. Martiny, M. Sogin, A. Boetius, and A. Ramette. 2011. "Global Patterns of Bacterial Beta-Diversity in Seafloor and Seawater Ecosystems." *PLoS ONE* 6 (9): e24570. https://doi.org/10.1371/journal.pone.0024570.

Zinger, L., J. Chave, E. Coissac, A. Iribar, E. Louisanna, S. Manzi, V. Schilling, H. Schimann, G. Sommeria-Klein, and P. Taberlet. 2016. "Extracellular DNA Extraction Is a Fast, Cheap and Reliable Alternative for Multi-Taxa Surveys Based on Soil DNA." *Soil Biology and Biochemistry* 96: 16–19. https://doi.org/10.1016/j.soilbio.2016.01.008.

Acknowledgements

I don't know where to start, really… so I guess I will start with where it all began. Thank you mom and dad for having shared with me your love of Nature, and especially Water. I guess I would have never been this interested in the oceans without all those hours spent swimming and snorkelling in the sea. Thank you mom, for having introduced me to scuba diving when I was 12. Passions arise with youth, and it is great to have people around that enable you to live them. Thank you Alejandro, for having introduced me to molecular biology, and showed me that research can be fun. Thank you Tim, for having introduced me to the Maldives and their exceptional biodiversity, for the English corrections, and for the motivation your presence instigated in the last months. I guess this PhD is in part a result of our "Maldivian expedition", seeing the incredible amounts of organisms in the water column made me wonder how on earth could we ever understand all this diversity without having to spend a lifetime describing each organism (something you pretty much did!). Than you Nicole and Thomas for having been my mentors and supervisors during my MSc studies in Bremen. I knew I wanted to keep doing marine ecology, but I also was very intrigued by the molecular underpinnings of organismal interactions. I somehow did not know how to combine the two. I guess I have found out now.

Thank you Sophie, for having put into the world a project like ABYSS, and Daniela for having opened my eyes to the wonderful world of meiofauna. Thank you both, for your presence, advice, and support during the last, quite intense, years.

Merci à tous les ardéchois et ardéchoises, pour vos cœurs si fidèles, votre présence et votre joie de vivre m'ont été d'une grande aide pour mieux vivre ces années de thèse. Merci à mes quatre super nanas : Maman, Margaux, Aurore, et Gipsy, pour votre constant soutien moral et émotionnel. Merci à toi Nicolas pour ta présence, ton amour, ton humour ces (presque 4) dernières années. Elles ne furent pas des plus faciles pour moi, et je te remercie du fond du cœur pour ta remarquable tolérance et ton soutien infaillible.

Last but not least, thank you to all my colleagues, past and present, that have made my life in Sète more enjoyable. Your supportive presence has definitely lightened up my PhD years!

Appendix

Résumé substantiel en français

L'humain, même s'il a, au cours de son évolution, continuellement cherché à se protéger de la nature, dépend fortement de l'équilibre des écosystèmes naturels qui l'entourent. En effet, les systèmes naturels apportent de grands bénéfices aux sociétés humaines, allant de la production de nourriture et d'eau potable, à la régulation du climat ou à l'atténuation de maladies.

Beaucoup d'études ont montré que la biodiversité, c'est-à-dire la variété des formes de vies sur notre planète, est une composante clef de la santé écosystémique. Le bon fonctionnement des écosystèmes, donc la manière dont ils stockent les ressources, produisent de la biomasse, décomposent et recyclent les nutriments, est étroitement lié à la biodiversité qu'ils abritent. Leur stabilité et leur capacité d'adaptation aux changements globaux aussi.

L'importance de la biodiversité repose donc dans les milliers de rôles que jouent toutes ces formes de vies, et qui rendent possible des systèmes biotiques complexes. Malgré une prise de conscience de la nécessité de protéger le monde naturel, et d'accords internationaux comme la déclaration de Rio en 1992, des études globales montrent le déclin continu de la biodiversité, principalement dû à l'augmentation des pressions anthropiques (Cardinale et al. 2012). La conséquence primordiale de ces facteurs de pression est la perte d'espèces ou de populations locales, en forte augmentation au 20^{ie} siècle. Cette diminution de la biodiversité au cours du siècle dernier a été mise en lien avec une productivité diminuée des écosystèmes et une diminution des quantités d'eau potable et des ressources naturelles à l'échelle mondiale. La biodiversité est donc un facteur clef des changements globaux actuels, et sa connaissance est nécessaire à la compréhension du système Terre.

Les méthodes de métabarcoding d'ADN environnemental (ADNe) ont révolutionné notre façon d'étudier la biodiversité et révélé la richesse spécifique considérable des écosystèmes, des sols alpins (Yoccoz et al. 2012) aux forêts tropicales (Lopes et al. 2017), en passant par les sédiments marins (Fonseca, V. G. et al. 2010; Forster et al. 2016; Pawlowski et al. 2011; Cowart et al. 2015; Sinniger et al. 2016; Cordier et al. 2019a) ou les milieux aquatiques (Deiner et al. 2016; Sunagawa et al. 2015; Vargas et al. 2015; Ibarbalz et al. 2019; Boussarie et al. 2018).

Le metabarcoding environnemental apparait comme une stratégie idéale pour l'étude d'écosystèmes vastes et difficiles d'accès: il permet une analyse rapide d'échantillons variés, tout en permettant de cibler plusieurs compartiments taxonomiques en parallèle. Cependant, de

nombreux obstacles restent à surmonter pour appliquer les méthodes basées sur l'ADNe de manière reproductible et fiable.

En particulier, l'application de ces méthodes chez les métazoaires reste compliquée car il est difficile de définir des unités taxonomiques opérationnelles moléculaires (OTUs) décrivant correctement la diversité au niveau spécifique. En effet, une source importante d'erreur dans les inventaires moléculaires de métazoaires provient du fait que ces-derniers sont des organismes multicellulaires, et que les marqueurs génétiques utilisés par le métabarcoding sont présents en copies multiples dans leurs génomes, et ce à des taux différents selon les taxons (Fig. 4, chapitre 1). Les erreurs de séquençage, de PCR, et les importantes variations génétiques intraspécifiques résultent dans le fait qu'une espèce, voire un individu, produit plusieurs unités taxonomiques opérationnelles (OTUs). Or, comme ces OTUs sont le proxy moléculaire pour la description d'espèces, il est important que la correspondance OTU-espèce soit maintenue afin de conserver la fiabilité de l'inventaire moléculaire de biodiversité.

Les pipelines bioinformatiques pour l'analyse de données de métabarcoding par ADNe sont en développement constant, et des avancées récentes permettent maintenant de corriger les séquences ADN (DADA2), de regrouper les séquences plus précisément en construisant des réseaux haplotypiques (swarm v2), et de filtrer les OTU en comparant leur taux d'identité et de cooccurrence (LULU), ce qui permet d'éviter les filtres d'abondance relative arbitraires.

En utilisant des communautés artificielles et 42 échantillons de sédiments profonds, le **deuxième chapitre** de la thèse cherche à évaluer ces nouveaux outils d'analyse de séquences ADN et de mettre en place un pipeline bioinformatique pour améliorer la qualité et la fiabilité des inventaires moléculaires de biodiversité.

Le pipeline développé est basé sur la correction des séquences Illumina avec DADA2, et permet d'analyser des données de métabarcoding produites à partir de marqueurs ribosomaux et mitochondriaux, de compartiments taxonomiques procaryotes et eucaryotes. Nous avons implémenté l'option de regrouper les variantes génétiques identiques (« amplicon sequence variant », ASV) produites par DADA2 en unités taxonomiques opérationnelles (OTU) avec swarm v2, un algorithme de regroupement (« clustering ») basé sur la théorie des réseaux, plus sensible aux données. En effet, les algorithmes de clustering, regroupant des séquences proches mais non-identiques, ont été développés pour diminuer le biais engendré par les erreurs produites durant le séquençage et la PCR, mais aussi pour diminuer le biais de la démultiplication des OTUs dû aux variations intraspécifiques. Le clustering reste donc potentiellement une étape importante dans le processus d'analyse des données de

métabarcoding, en particulier chez les métazoaires. Pour finir, les ASVs / OTUs finales peuvent être filtrées en fonction de leurs taux d'identité et de cooccurrence en utilisant LULU.

Les résultats montrent que des inventaires de diversité fiables peuvent être obtenus en utilisant les algorithmes de correction DADA2 et LULU, mais soulignent que le clustering des ASVs en OTUs, combiné à la filtration additionnelle de LULU, est nécessaire pour produire des inventaires de biodiversité métazoaire fiables. Aussi, les seuils d'identités de LULU sont à choisir soigneusement selon la variabilité du marqueur utilisé. Pour les marqueurs mitochondriaux, le seuil défaut de 84% était approprié mais trop bas pour les marqueurs ribosomaux tel que 18S où il conduisait à la perte d'espèces dans les communautés artificielles et a donc dû être augmenté à 90%. Enfin, deux méthodes d'assignation taxonomique des ASVs/OTUs ont été implémentées dans le pipeline : le classificateur bayésien (RDP) et BLAST, un algorithme basé sur l'identité des séquences. La comparaison de BLAST et du classificateur RDP a souligné le potentiel de ce dernier à fournir de très bonnes assignations, mais a mis en évidence la nécessité d'un effort concerté par la communauté scientifique pour développer des bases de données exhaustives et spécifiques aux communautés étudiées.

L'utilisation des nouvelles générations de séquençage (NGS) combiné à l'ADNe constitue une avancée considérable pour la recherche en environnement profond car les fonds abyssaux restent très peu étudiés du fait de leur difficulté d'accès, et constituent potentiellement un grand réservoir de biodiversité. En effet, ils recouvrent plus de 50% de la planète Terre, et ils ont d'importants rôles écosystémiques, comme le recyclage du carbone, ou la production secondaire de matière organique (Smith, K. L. et al. 2009; Bik et al. 2012b). Mais ces habitats dits 'profonds' subissent l'impact grandissant des activités humaines, allant de menaces indirectes dû au changement climatique affectant l'équilibre physico-chimiques des océans, aux menaces directes tel que le stockage de déchets, la pollution, ou l'exploitation de ressources naturelles, minières, et pétrolières (Ramirez-Llodra et al. 2010; 2011; Levin, L. A. et al. 2016).

Une meilleure connaissance de la biodiversité marine profonde et des facteurs biotiques et abiotiques influençant sa distribution est donc nécessaire, afin de mettre en place des stratégies de protection et de gestion de ces écosystèmes. Même si le metabarcoding d'ADNe est une méthode efficace dans ces écosystèmes vastes et difficiles d'accès, son application sur des sédiments profonds est potentiellement biaisée par la présence d'ADN archivé provenant d'organismes morts. Or, l'inclusion de cet ADN ancien (ADNa) aboutirait à des inventaires de biodiversité passée plutôt que présente. Ainsi, le second objectif de cette thèse est d'évaluer des

protocoles de métabarcoding d'ADNe afin de sélectionner l'approche permettant de décrire au mieux les communautés vivantes.

Dans le **troisième chapitre**, nous avons pour cela utilisé cinq sites profonds couvrant des habitats allant de monts sous-marins à des sources hydrothermales et des volcans de boue, et nous avons ciblé en parallèle les procaryotes (16S-V4V5), les protistes (18S-V4), et les métazoaires (18S-V1, COI).

Dans un premier temps, des inventaires basés sur l'ADN furent comparés à ceux révélés par l'ARN. En effet, ce-dernier, étant produit uniquement par des organismes vivants, a été présenté comme une approche plus pratique pour cibler la partie active des communautés. Parallèlement, l'ADN ancien étant principalement constitué de petits fragments, nous avons aussi évalué l'effet de l'élimination de fragments d'ADN courts par sélection de taille et reconcentration par éthanol.

Les résultats montrent que l'élimination de fragments d'ADN courts n'affecte pas les estimations de la diversité alpha et bêta dans aucun des compartiments biologiques étudiés. Les résultats confirment également les doutes quant à la possibilité de mieux décrire les communautés vivantes en utilisant l'ARN environnemental (ARNe). Sur les marqueurs ribosomaux, l'ARN, tout en résolvant des schémas spatiaux similaires à l'ADN co-extrait, a entraîné des estimations de richesse significativement plus élevées, soutenant les hypothèses de persistance accrue de l'ARN ribosomal (ARNr) dans l'environnement, et l'existence d'un biais additionnel et non mesuré en raison de la surabondance d'ARNr dans l'environnement et d'ARN sécrété à taux variables en fonction de l'activité métabolique des organismes. Sur le locus mitochondrial, l'ARN a détecté une richesse métazoaire inférieure tout en résolvant moins de différences écologiques que l'ADN co-extrait, reflétant la grande labilité de l'ARN messager. Les résultats soulignent également l'importance d'utiliser de grandes quantités de sédiments (≥ 10 g) pour étudier avec précision la diversité eucaryote. Dans le **troisième chapitre**, nous montrons donc que l'ADN est plus pertinent que l'ARN pour des études logistiquement réalistes, reproductibles, et fiables. Nous confirmons aussi que des quantités de sédiments plus grandes (≥ 10 g) fournissent des évaluations plus complètes et précises de la biodiversité eucaryote benthique et qu'il faut favoriser l'augmentation du nombre de réplicas biologiques plutôt que techniques pour déduire des patrons écologiques fiables.

Bien que le tri par taille des organismes moyennant tamisage des sédiments ait permis, comme attendu (Elbrecht et al.2017), de détecter une diversité métazoaire plus élevée (**chapitre**

4), des ségrégrations spatiales et des compositions taxonomiques similaires ont été obtenus dans des échantillons tamisés et non tamisés, indiquant que le temps considérable associé au tamisage n'est pas essentiel pour faire des évaluations écologiques robustes. Cela est confirmé par le fait que le lavage et le tamisage des échantillons peuvent entraîner une perte substantielle d'organismes (Montagna et al.2017), ajouté au risque accru de contamination par de l'ADN allochtone, en particulier si les compartiments de taille protiste et procaryote présentent un intérêt d'étude.

Les protocoles de métabarcoding optimisés dans les chapitres précédents ont été utilisés dans le **cinquième chapitre** pour réévaluer la biodiversité profonde dans la zone de transition Atlantique-Méditerranée, offrant une preuve de concept pour l'étude de la biodiversité des grands fonds marins à large échelle à travers l'ADNe.

En effet, alors que les habitats sédimentaires représentent la grande majorité des habitats dans les abysses, et qu'ils sont un grand réservoir de biodiversité encore largement non décrite, largement moins de 1% des grands fonds a été étudié à ce jour. Dans de tels écosystèmes vastes et difficiles d'accès, le pouvoir de détection élevé du métabarcoding d'ADN environnemental (ADNe), sur des échantillons plus faciles à recueillir que les collections de spécimens morphologiques, offre de nouvelles perspectives pour l'investigation standardisée de biodiversité et biogéographie à grande échelle.

En combinant le marqueur génétique mitochondrial COI et la région V1-V2 de l'ARN ribosomal 18S (ARNr), nous avons étudié la biodiversité métazoaire à petite et à grande échelle dans la zone de transition Atlantique-Méditerranée, à l'aide d'ADNe extrait de sédiments profonds provenant de 13 sites allant de la Méditerranée centrale à la dorsale médio-atlantique. Nous avons évalué l'influence de la couche de sédiments, de la taille des grains de sédiments, de la teneur en matière organique ainsi que des communautés procaryotes (16S V4-V5), sur l'étendue et la structure de la biodiversité métazoaire dans cette région.

Nos résultats soulignent que les facteurs à petite échelle (centimètres) affectent fortement la richesse des métazoaires des grands fonds marins et la composition des communautés. Une diminution significative de la richesse en OTU a été observée avec chaque couche de sédiments, allant de 1 cm à 15 cm de profondeur, et une ségrégation verticale importante dans la structure des communautés a été révélée dans toutes les régions pour la méiofaune et la macrofaune. Les premiers cinq centimètres de sédiment abritaient la plupart des OTU métazoaires (94% pour

18S, 98% pour COI), avec des nombres d'OTU allant de 2 à 168 par échantillon pour 18S et de 81 à 1259 pour COI.

Les facteurs à grande échelle (> 100 km) ont davantage affecté la diversité bêta que la diversité alpha. Le pourcentage de matière organique et la taille des grains de sédiments montrèrent une forte variation à l'échelle régionale, avec une teneur en matière organique plus élevée dans les sédiments méditerranéens et des particules de plus grande taille dans l'Atlantique. Ces deux variables contribuèrent de manière significative à expliquer les différences de composition des communautés entre sites. La méio et la macrofaune ont également montré une forte relation avec le compartiment procaryote (RV = 0,5-0,65), ceci peut être dû à une dépendance similaire aux paramètres abiotiques, ainsi qu'à des relations biotiques directes ou indirectes. Enfin, le détroit de Gibraltar était un facteur supplémentaire expliquant les très fortes différences régionales dans la composition des communautés, soutenant une influence combinée de facteurs historiques et de mouvements actuels des masses d'eau sur la distribution de la diversité benthique.

En 2010, la Conférence des Parties à la Convention sur la Diversité Biologique a établi le Plan Stratégique pour la Biodiversité 2011-2020, et a délimité cinq « objectifs stratégiques » qui ont été divisés en 20 objectifs (Aichi Targets). Chaque « Objectif d'Aichi » a été conçu pour mieux comprendre et prévoir la biodiversité, en particulier, comment la diversité biologique sous-tend le fonctionnement des écosystèmes et comment les services écosystémiques sont essentiels pour le bien-être humain. L'accomplissement de ces objectifs d'Aichi garantit en fin de compte nos moyens de subsistance ainsi que le développement économique, et est essentielle pour le maintien de la biodiversité et la réduction de la pauvreté.

Les lacunes dans les connaissances sur la biodiversité sont particulièrement sévères dans les abysses, où > 90% des étendues restent inexplorées. Cette thèse a évalué et optimisé les protocoles de métabarcoding par ADNe pour les sédiments des grands fonds à un niveau de traitement bioinformatique, moléculaire et d'échantillonnage, et en ciblant de multiples compartiments du vivant, ouvrant la porte à des études à grande échelle sur la biodiversité dans les grands fonds marins, contribuant ainsi à une meilleure compréhension de la biodiversité, de la biogéographie et du fonctionnement des écosystèmes dans cet univers vaste et insaisissable. Ce travail ouvre également la voie à l'établissement de protocoles de biomonitoring efficaces qui sont de plus en plus nécessaires dans les zones ciblées par les industries minières et pétrolières.

Supplementary material Chapter II.

Supplemental tables

Table S1. Taxonomic and relative composition of the deep-sea metazoan mock communities used in this study.

Taxonomic group	Species	Mock 3 (%)	Mock 5 (%)
Polychaeta; Eunicida	*Eunice norvegica*	40	80
Crustacea; Malacostraca	*Chorocaris* sp. (now *Rimicaris* sp. or *M. fortunata*)	3	0.7
Crustacea; Malacostraca	*Alvinocaris muricola*	3	0.7
Crustacea; Malacostraca	*Munidopsis* sp.	3	0.7
Anthozoa;Alcyonacea	*Acanella arbuscula*	20	10
Anthozoa; Scleractinia;Caryophylliidae	*Desmophyllum dianthus*	3	0.7
Bivalvia;Veneroida; Vesicomyidae	*Calyptogena pacifica*	3	0.7
Bivalvia;Veneroida; Vesicomyidae	*Christineconcha regab* (formerly *Calyptogena* sp.)	3	0.7
Bivalvia;Veneroida; Vesicomyidae	*Vesicomya gigas*	3	0.7
Gastropoda; Patellogastropoda	*Paralepetopsis* sp.	20	5

Table S2. Sampling sites and their GPS locations and associated habitats

Station	Cruise	Depth (m)	Latitude	Longitude	Habitat	Region
ESN	EssNaut	2,400	42.9423	6.7423	Abyssal plain	Gulf of Lyon, Mediterranean
PCT-FA	PEACETIME	2,800	37.9467	2.9167	Abyssal plain	Western Mediterranean
MDW-ST179	MEDWAVES	729	36.4808	-2.8945	Seamount	Western Mediterranean
MDW-ST201	MEDWAVES	381	36.546	-2.8135	Seamount	Western Mediterranean
MDW-ST215	MEDWAVES	554	36.5157	-2.7942	Seamount	Western Mediterranean
MDW-ST22	MEDWAVES	470	36.5598	-6.9492	Mud volcano	Gibraltar Strait
MDW-ST23	MEDWAVES	470	36.5605	-6.9498	Mud volcano	Gibraltar Strait
MDW-ST38	MEDWAVES	1,920	36.8442	-11.3025	Seamount	North Atlantic
MDW-ST68	MEDWAVES	1,245	37.2837	-24.7873	Seamount	North Atlantic
MDW-ST117	MEDWAVES	1,325	37.34	-24.7552	Seamount	North Atlantic
MRM-ST35	MarMine	2,683	73.4643	7.1975	Hydrothermal vent	Arctic
MRM-ST38	MarMine	2,684	73.4639	7.1984	Hydrothermal vent	Arctic
MRM-ST48	MarMine	2,826	73.4598	7.2184	Hydrothermal vent	Arctic
CHR	CANHROV	2,490	42.7167	6.1333	Marine canyon	Gulf of Lyon, Mediterranean

Table S3. ABYSS metabarcoding pipeline.

Process	**Software**	**Script(s) and command(s)**
Raw reads preprocessing for ligation data	Abyss-preprocessing: separate forward and reverse reads in each run, and re-pair reads	extract.sh using extractR1R2.py with cutadapt v1.18 (-e 0.14-0.17 for 18S,16S i.e. 3 nt mismatches and 0.27 for COI , -O length of primer -1) and BBMAP Repair v38.22
Read quality-filtering	Dada2 v.1.10	filterAndTrim() in dada2main.R maxEE=2, maxN=0, truncQ=11, truncLen=220 (18S, 16S) or 200 (COI)
Read error learning	Dada2 v.1.10	learnErrors() in dada2main.R nbases=1e8, multithread=TRUE, randomize=TRUE
Read dereplicating	Dada2 v.1.10	derepFastq() in dada2main.R
Read correction	Dada2 v.1.10	dada() in dada2main.R
Read merging	Dada2 v.1.10	mergePairs() in dada2main.R minOverlap=12, maxMismatch=0
Make sequence table and filter by length	Dada2 v.1.10	makeSequenceTable() in dada2main.R seqtab[,nchar(colnames(seqtab)) %in% seq(lengthMin,lengthMax)] lengthMin= 330 (18S-V1), 300 (COI), 350 (18S-V4), 87 (18S-V9), 350 (16S) lengthMax= 390 (18S-V1), 326 (COI), 410 (18S-V4), 186 (18S-V9), 390 (16S)
Chimera removal	Dada2 v.1.10	removeBimeraDenovo() in dada2main.R
Taxonomy assignement with RDP Classifier	Dada2 v.1.10	assignTaxonomy () in dada2outputfiles.R minBoot=50, outputBootstraps=TRUE
Taxonomy assignment with BLAST+	blastn (megablast) v.2.6.0	blast.pbs -outfmt 11 -qcov_hsp_perc 80 -perc_identity 70 -max_hsps 1, -evalue 1e-5, then merge BLAST and RDP taxonomies using concat_blast_rdp_tax.pbs
Clustering (optional), chimera removal, taxonomic assignement of OTUs	FROGS v.2.0.0	frogs.pbs using clustering.py, then remove_chimera.py, and affiliation_OTU_identite_couverture.py
Blank correction	Rscript	Data_refining.Rmd using packages decontam v.1.2.1 and phyloseq v.1.26.0
Removal of unassigned and non-target clusters		
Deletion of defective samples (< 10,000 target reads)		
Tag-switching renormalisation	Rscript	owi_renormalize.R
LULU curation	LULU v.0.1	lulu() in lulu.R using minimum_ratio_type = "min", minimum_ratio = 1/100/1000, minimum_match =84/90, minimum_relative_cooccurence =0.95

Table S4. DADA2 read-track table. Number of reads obtained in samples after each processing step. Data refining was performed in R (decontamination, renormalisation, removal of non-target taxa, and clusters unassigned at phylum-level or with unreliable phylum-level assignments), based on BLAST assignments obtained using the Silva training set available on the DADA2 website for 18S and 16S, and on the MIDORI database for COI.

Sample type	Number initial of samples	Raw reads	Quality-filtered reads	Merged reads	Length-filtered reads	Non chimeric reads Dada2	% reads retained	Number of samples after refining	Number of target reads after all refining setps
LOCUS									
COI									
Control Sample	16	2,146,476	1,053,997	1,024,547	1,015,821	1,015,700	47	0	-
Mock Sample	2	1,482,785	1,261,045	1,252,908	1,251,994	1,224,795	83	2	689,755
Environmental Sample	40	31,010,653	26,011,238	25,287,002	22,197,457	22,004,407	71	40	4,771,366
18S-V1									
Control Sample	14	6,141,567	2,508,908	2,441,821	2,200,132	2,186,230	36	0	-
Mock Sample	2	2,096,631	1,607,219	1,436,773	1,430,823	1,289,608	62	2	1,287,871
Environmental Sample	42	37,590,781	26,828,194	24,826,430	22,636,689	22,297,846	59	42	8,212,578
16S-V4V5									
Control Sample	10	3,531,226	2,889,163	2,634,536	2,619,479	2,618,729	74	0	-
Environmental Sample	42	12,875,651	9,307,729	7,122,154	7,114,195	6,827,513	53	42	6,699,094

Two sediment samples failed amplification for the COI marker gene (PCT_FA_CT2_0_1 and CHR_CT1_0_1). For metazoans, less reads were retained after bioinformatic processing in negative controls (36% for 18S, 47% for COI) compared to true samples (~60% for 18S, ~70% for COI), while the opposite was observed for 16S (74% of reads retained in control samples against 53% in true samples). Negative control samples (field, extraction, and PCR controls) contained 2,186,230 (~8%) 18S reads, 1,015,700 (~4%) COI reads, and 2,618,729 (28%) 16S reads. These reads were mostly originating from the field controls for metazoans (48% for 18S, 55% for COI) and extractions controls for 16S (50%). After blank correction, data refining, and abundance renormalization, rarefaction curves showed that a plateau was achieved for all samples in both clustered and non-clustered datasets, suggesting an overall sequencing depth adequate to capture the diversity presen

Table S5. Relative read abundance (%) detected per species in the mock communities using different bioinformatic pipelines. Taxonomy is given up to the lowest common rank assigned to OTUs from mock species. "Others" represents unexpected OTUs, i.e. with assignments not related to any species in the mocks. These may represent contamination or symbionts of the mock species. °Bivalvia was common rank for P. kilmeri and C. regab for all pipelines with swarm clustering. *Bivalvia was common rank for pipelines with $d > 1$.

	COI minimum-ratio = 1	**DADA2 +**	**Ø**	**LULU**		**swarm**					**swarm+LULU 90%**					**swarm+LULU 84%**				
		DNA in %		**90%**	**84%**	**d1**	**d5**	**d6**	**d7**	**d13**	**d1**	**d5**	**d6**	**d7**	**d13**	**d1**	**d5**	**d6**	**d7**	**d13**
Mock3	*Acanella arbuscula*	20	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11	11
	Hexacorallia; *D.dianthus*	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
	Alvinocaris ; *A. muricola*	3	5	5	5	5	5	5	5	5	5	5	5	5	5	4	4	4	4	4
	Chorocaris sp.	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Munidopsis sp.	3	15	15	15	15	15	15	15	15	15	16	16	16	16	15	15	15	15	15
	Gastropoda; *Paralepetopsis* sp.	20	28	27	27	28	28	28	28	28	27	27	27	27	27	27	27	27	27	27
	Bivalvia; *C. regab°*	3	0.2	0.2	0.3	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
	Phreagena kilmeri°	3	8	8	7															
	Vesicomya gigas	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	Polychaeta; *E.norvegica*	40	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26	26
	Others	0	0.1	0.1	0.1	0	0	0	0	0	0	0	0	0	0	0.01	0.01	0.01	0.01	0.01
Mock5	*Acanella arbuscula*	10	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12	12
	Hexacorallia; *D.dianthus*	0.7	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	Alvinocaris ; *A. muricola*	0.7	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	Chorocaris sp.	0.7	0.5	0.6	0.7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Munidopsis sp.	0.7	9	9	10	9	9	9	9	9	9	9	9	9	9	10	10	10	10	10
	Gastropoda; *Paralepetopsis* sp.	5	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18	18
	Bivalvia; *C. regab°*	0.7	0.1	0.1	0.1	7	7	7	7	7	6	6	6	6	6	6	6	6	6	6
	Phreagena kilmeri°	0.7	7	6	6															
	Vesicomya gigas	0.7	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	Polychaeta; *E.norvegica*	80	46	46	45	46	46	46	46	46	46	46	46	46	46	45	45	45	45	45
	Others	0	0.1	0.1	0.1	0	0	0	0	0	0	0	0	0	0	0.01	0.01	0.01	0.01	0.006
	18S minimum-ratio = 100			**90%**	**84%**	**d1**	**d3**	**d4**	**d5**	**d11**	**d1**	**d3**	**d4**	**d5**	**d11**	**d1**	**d3**	**d4**	**d5**	**d11**
Mock3	Alcyonacea; *A.arbuscula*	20	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55	55
	Caryophylliidae; *D.dianthus*	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	Alvinocaris muricola	3	1	1	1	1	2	2	2	2	1	2	2	2	2	1	2	2	2	2
	Chorocaris sp.	3	1	1	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0
	Munidopsis sp.	3	8	8	9	8	8	8	8	8	8	8	8	8	8	9	8	8	8	8
	Gastropoda; *Paralepetopsis* sp.	20	0.01	0.01	0	0.01	0.03	0.03	0.03	0.01	0.02	0.02	0.02	0.02	0.02	0	0	0	0	0
	Vesicomyidae; *P. kilmeri/C. regab/V. gigas**	9	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19	19
	Polychaeta; *E.norvegica*	40	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14
	Others	0	0.05	0.04	0.04	0.06	0.04	0.05	0	0.03	0.04	0.06	0.06	0.06	0.06	0.05	0.04	0.06	0.06	0.06
Mock5	Alcyonacea; *A.arbuscula*	10	49	50	49	50	49	49	49	50	50	50	50	50	50	50	50	50	50	50
	Caryophylliidae; *D.dianthus*	0.7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Alvinocaris muricola	0.7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Chorocaris sp.	0.7	0.3	0.3	0.3	0.3	0	0	0	0	0.3	0	0	0	0	0.3	0	0	0	0
	Munidopsis sp.	0.7	4	4	5	4	4	4	4	4	4	4	4	4	4	5	4	4	5	5
	Gastropoda; *Paralepetopsis* sp.	5	0.01	0.01	0	0.01	0	0	0	0	0.01	0.01	0.01	0.01	0.01	0	0	0	0	0
	Vesicomyidae; *P. kilmeri/C. regab/V. gigas**	2.1	12	12	12	12	12	12	12	12	11	11	11	11	12	12	12	12	12	12
	Polychaeta; *E.norvegica*	80	32	33	32	33	32	32	32	32	33	32	32	32	32	32	32	32	32	32
	Others	0	0.03	0.02	0.02	0.03	0.01	0.01	0.02	0.01	0.03	0.01	0.01	0.01	0	0.03	0.01	0.01	0.01	0.01

Table S6. The effect of LULU minimum match and minimum ratio parameters. Number of 18S ASVs/OTUs detected per species in the mock communities using DADA2 with or without swarm clustering, and LULU curation at two different minimum match (84% and 90%) and minimum ratio (1 and 1000) parameters. Taxonomy is given up to the lowest common rank assigned to OTUs from mock species. "Others" represents unexpected OTUs, i.e. with assignments not related to any species in the mocks. These may represent contamination or symbionts of the mock species. *Bivalvia was common rank for pipelines with $d > 1$.

	18S minimum-ratio = 1	**DADA2+**	**Ø**	**LULU**		**swarm**					**swarm+LULU 90%**					**swarm+LULU 84%**				
		DNA in %		**90%**	**84%**	**d1**	**d3**	**d4**	**d5**	**d11**	**d1**	**d3**	**d4**	**d5**	**d11**	**d1**	**d3**	**d4**	**d5**	**d11**
Mock3	Alcyonacea; *A.arbuscula*	20	64	1	1	29	11	6	9	7	1	1	1	1	1	1	1	1	1	1
	Caryophylliidae; *D.dianthus*	3	2	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1
	Alvinocaris muricola	3	1	1	1	2	2	1	2	2	1	1	1	1	1	1	1	1	1	1
	Chorocaris sp.	3	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	Munidopsis sp.	3	4	1	1	5	4	2	3	3	1	1	1	1	1	1	1	1	1	1
	Gastropoda; *Paralepetopsis* sp.	20	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**0**	**0**	**0**	**0**
	Vesicomyidae; *P. kilmeri/C. regab/V. gigas**	9	8	1	1	4	4	2	4	4	1	1	1	1	1	1	1	1	1	1
	Polychaeta; *E.norvegica*	40	8	3	2	5	4	3	2	2	3	2	2	2	2	2	1	2	2	2
	Others	0	4	2	2	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Mock5	Alcyonacea; *A.arbuscula*	10	55	1	1	28	11	6	9	7	1	1	1	1	1	1	1	1	1	1
	Caryophylliidae; *D.dianthus*	0.7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Alvinocaris muricola	0.7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Chorocaris sp.	0.7	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
	Munidopsis sp.	0.7	3	1	1	3	2	2	2	2	1	1	1	1	1	1	1	1	1	1
	Gastropoda; *Paralepetopsis* sp.	5	**1**	**1**	**0**	**1**	**0**	**1**	**0**	**0**	**1**	**1**	**1**	**1**	**1**	**0**	**0**	**0**	**0**	**0**
	Vesicomyidae; *P. kilmeri/C. regab/V. gigas**	2.1	5	1	1	5	4	2	4	4	1	1	1	1	1	1	1	1	1	1
	Polychaeta; *E.norvegica*	80	10	3	2	5	4	3	4	4	3	2	2	2	2	2	1	2	2	2
	Others	0	5	2	2	6	3	2	3	3	5	3	3	3	1	4	2	2	2	2
	18S minimum-ratio = 1000			**90%**	**84%**	**d1**	**d3**	**d4**	**d5**	**d11**	**d1**	**d3**	**d4**	**d5**	**d11**	**d1**	**d3**	**d4**	**d5**	**d11**
Mock3	Alcyonacea; *A.arbuscula*	20	64	16	16	29	11	6	9	7	6	4	4	3	2	6	4	4	3	2
	Caryophylliidae; *D.dianthus*	3	2	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1
	Alvinocaris muricola	3	1	3	1	2	2	1	2	2	2	2	2	2	2	1	2	2	2	2
	Chorocaris sp.	3	**1**	**1**	**1**	**1**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**
	Munidopsis sp.	3	4	1	1	5	4	2	3	3	2	2	2	2	1	2	1	1	1	1
	Gastropoda; *Paralepetopsis* sp.	20	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**0**
	Vesicomyidae; *P. kilmeri/C. regab/V. gigas**	9	8	4	4	4	4	2	4	4	3	2	2	2	1	3	2	2	2	1
	Polychaeta; *E.norvegica*	40	8	4	3	5	4	3	2	2	3	2	2	2	2	2	1	2	2	2
	Others	0	4	2	2	3	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Mock5	Alcyonacea; *A.arbuscula*	10	55	15	15	28	11	6	9	7	6	4	4	3	2	6	4	4	3	2
	Caryophylliidae; *D.dianthus*	0.7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Alvinocaris muricola	0.7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	Chorocaris sp.	0.7	**1**	**1**	**1**	**1**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**
	Munidopsis sp.	0.7	3	2	2	3	2	2	2	2	2	2	2	2	1	2	2	2	2	1
	Gastropoda; *Paralepetopsis* sp.	5	**1**	**1**	**1**	**1**	**0**	**1**	**0**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**0**
	Vesicomyidae; *P. kilmeri/C. regab/V. gigas**	2.1	5	3	3	5	4	2	4	4	3	2	2	2	1	3	2	2	2	1
	Polychaeta; *E.norvegica*	80	10	5	4	5	4	3	4	4	3	2	2	2	2	2	1	1	1	2
	Others	0	5	2	3	6	3	2	3	3	4	2	2	2	1	4	3	3	3	2

Table S7. Number of raw, refined, and LULU-curated molecular clusters obtained for each pipeline in the three datasets. Data refining was performed in R (decontamination, renormalisation, removal of non-target taxa, and clusters unassigned at phylum-level or with unreliable phylum-level assignments), based on BLAST assignments obtained using the Silva v132 database for 18S and 16S, and on the MIDORI database for COI. LULU curation was performed at minimum ratio = 100 for 18S, and minimum ratio = 1 for COI (default).

Locus	Pipeline	Number of raw ASVs/OTUs	Number of target ASVs/OTUs after all refining steps	Number of target OTUs after LULU 90%	Number of target OTUs after LULU 84%	ASV to OTU ratio	ASV to OTU ratio (LULU 90%)	ASV to OTU ratio (LULU 84%)
COI	Dada2	78,785	13,397	10,028	5,539	-	-	-
	Dada2+d=1	64,669	5,563	4,422	2,263	2.4	2.3	2.4
	Dada2+d=5	53,749	4,503	3,761	2,016	3.0	2.7	2.7
	Dada2+d=6	52,216	4,324	3,656	1,971	3.1	2.7	2.8
	Dada2+d=7	50,919	4,190	3,568	1,922	3.2	2.8	2.9
	Dada2+d=13	44,684	3,518	3,158	1,758	3.8	3.2	3.2
18S V1-V2	Dada2	57,661	8,280	6,909	6,518	-	-	-
	Dada2+d=1	44,948	6,015	5,063	4,677	1.4	1.4	1.4
	Dada2+d=3	34,569	4,194	3,676	3,322	2.0	1.9	2.0
	Dada2+d=4	31,509	3,716	3,278	2,945	2.2	2.1	2.2
	Dada2+d=5	28,764	3,313	2,949	2,636	2.5	2.3	2.5
	Dada2+d=11	19,504	1,869	1,676	1,469	4.4	4.1	4.4
16S V4-V5	Dada2	56,577	53,815	-	-	-	-	-
	Dada2+d=1	41,746	38,972	-	-	1.4	-	-
	Dada2+d=3	29,023	26,676	-	-	2.0	-	-
	Dada2+d=4	25,406	23,165	-	-	2.3	-	-
	Dada2+d=5	22,841	20,697	-	-	2.6	-	-
	Dada2+d=11	14,631	12,800	-	-	4.2	-	-

Table S8. Read and cluster abundance with data refining based on BLAST and RDP taxonomy. Number of ASVs/OTUs obtained in datasets when refining was performed based on BLAST or RDP assignments (blast / rdp). Metazoan datasets (COI and 18S) were clustered (swarm with d=1) and curated with LULU at 90% for 18S (min-ratio=100) and 84% for COI (min-ratio=1), while ASVs were used in the prokaryote dataset. Taxonomic affiliations were obtained using the Silva v132 database for 18S and 16S, and the MIDORI-UNIQUE database subsampled for marine taxa for COI. BLAST assignments were performed with minimal hit identity of 70%. Phylum-level taxonomy filter was performed by keeping only clusters with BLAST identities ≥ 86% for ribosomal loci and ≥ 80% for COI, or with phylum bootstrap ≥ 80%. Genus-level taxonomy filter was performed by keeping only clusters with BLAST identities > 95% for ribosomal loci and > 93% for COI, or with genus bootstrap ≥80%.

LOCUS	Number of raw ASVs/OTUs	% raw ASVs/OTUs assigned at phylum-level (BLAST / RDP)	Number of target clusters before taxonomy quality filter (BLAST / RDP)	Number of target clusters after phylum-level quality-filter (BLAST / RDP)	Number of target clusters after genus-level quality-filter (BLAST / RDP)
COI	64,669	33% / 99%	10,113 / 39,269	3,041 / 242	105 / 112
18S V1-V2	44,948	76% / 97%	5,634 / 7,034	5,063 / 5,410	1,916 / 4,187
16S V4-V5	56,577	97% / 95%	55,129 / 53,715	53,815 / 51,474	35,614 / 40,827

Supplemental figures

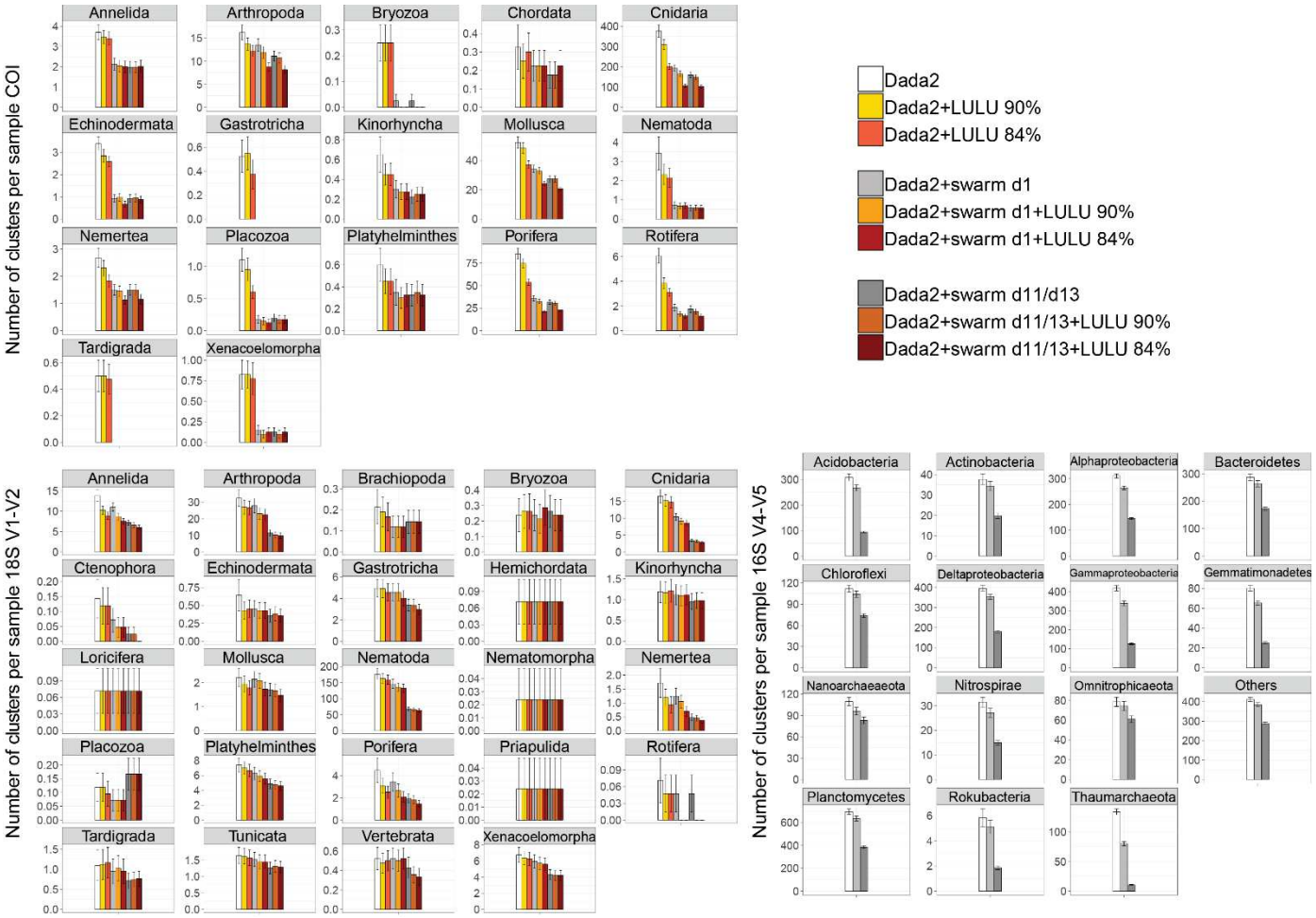


Figure S1. Mean number of metazoan (COI, 18S) and prokaryote (16S) clusters detected per taxon in ASV vs OTU-centred datasets. Cluster numbers from sediment samples of 14 deep-sea sites were calculated from datasets rarefied to same depth. ASVs detected with the DADA2 metabarcoding pipeline are compared with OTU numbers obtained after clustering with swarm v2 (*d = 1*, *d = 11* for 18S and 16S, and d = *13* for COI) and/or after curation with LULU at 84% and 90% *minimum match*. LULU curation was performed with *minimum ratio* = 100 for 18S and *minimum ratio* = 1 for COI. Error bars represent standard errors.

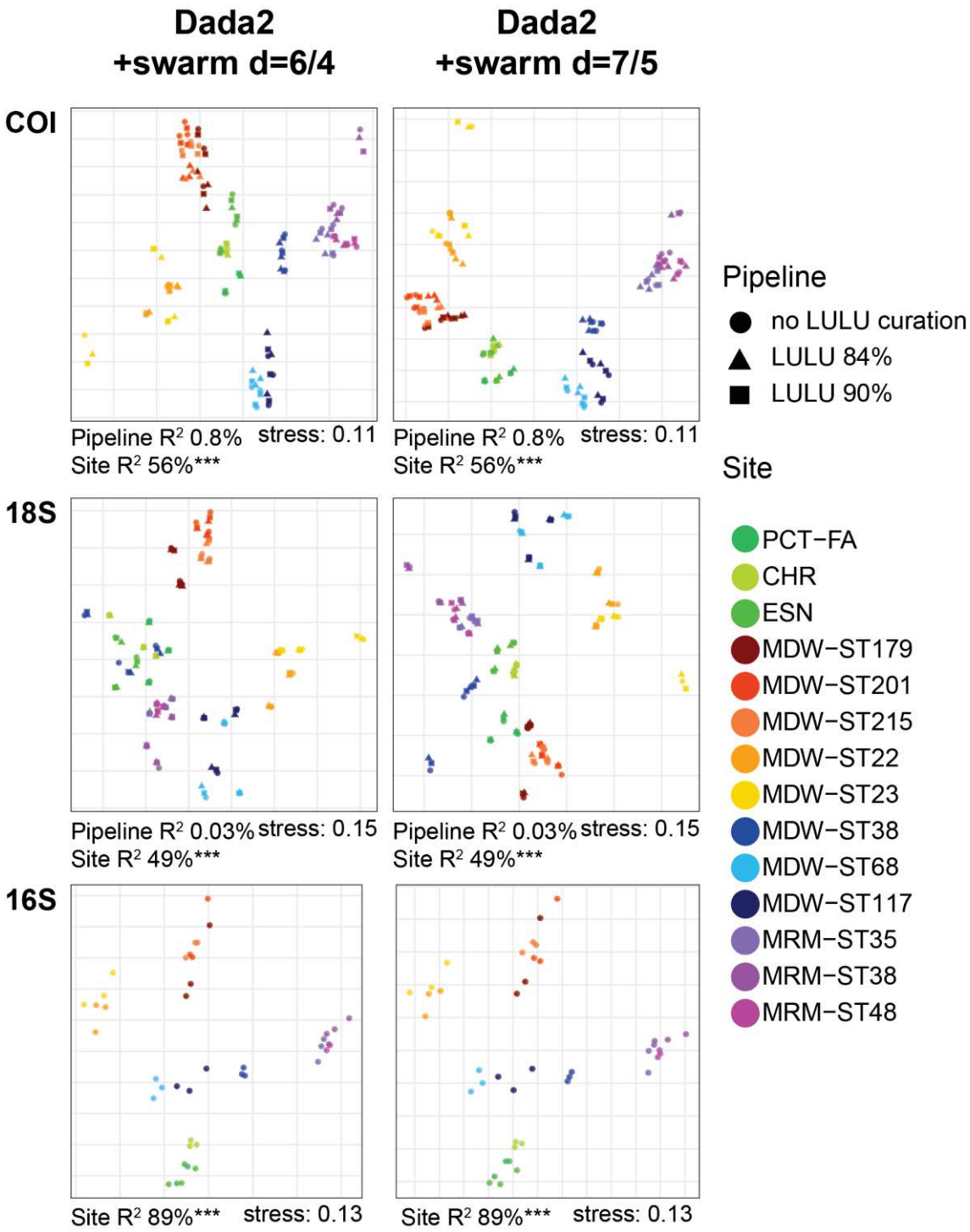


Figure S2. Beta-diversity patterns in OTU-centred datasets at swarm clustering levels of *d = 4-7*. Nonmetric multidimensional scaling (NMDS) ordinations showing community differentiation observed between sites with different clustering scenarios. ASVs obtained with DADA2 were clustered with swarm at *d* = 6-7 (COI) and *d* = 4-5 (18S, 16S). Metazoan OTUs were curated with LULU at 84% and 90% *minimum match*. LULU curation was performed with *minimum ratio* = 100 for 18S and *minimum ratio* = 1 for COI. R^2 values and associated p-values obtained in PERMANOVAs are shown under the ordination plots. Significance codes: ***: p<0.001; **: p<0.01; *: p<0.05. Site colour codes: Green: Mediterranean > 1,000 m; Red: Mediterranean Gibraltar Strait 300-1,000 m; Yellow: Atlantic Gibraltar Strait 300-1,000 m; Blue: North Atlantic > 1,000 m; Purple: Arctic > 1,000 m.

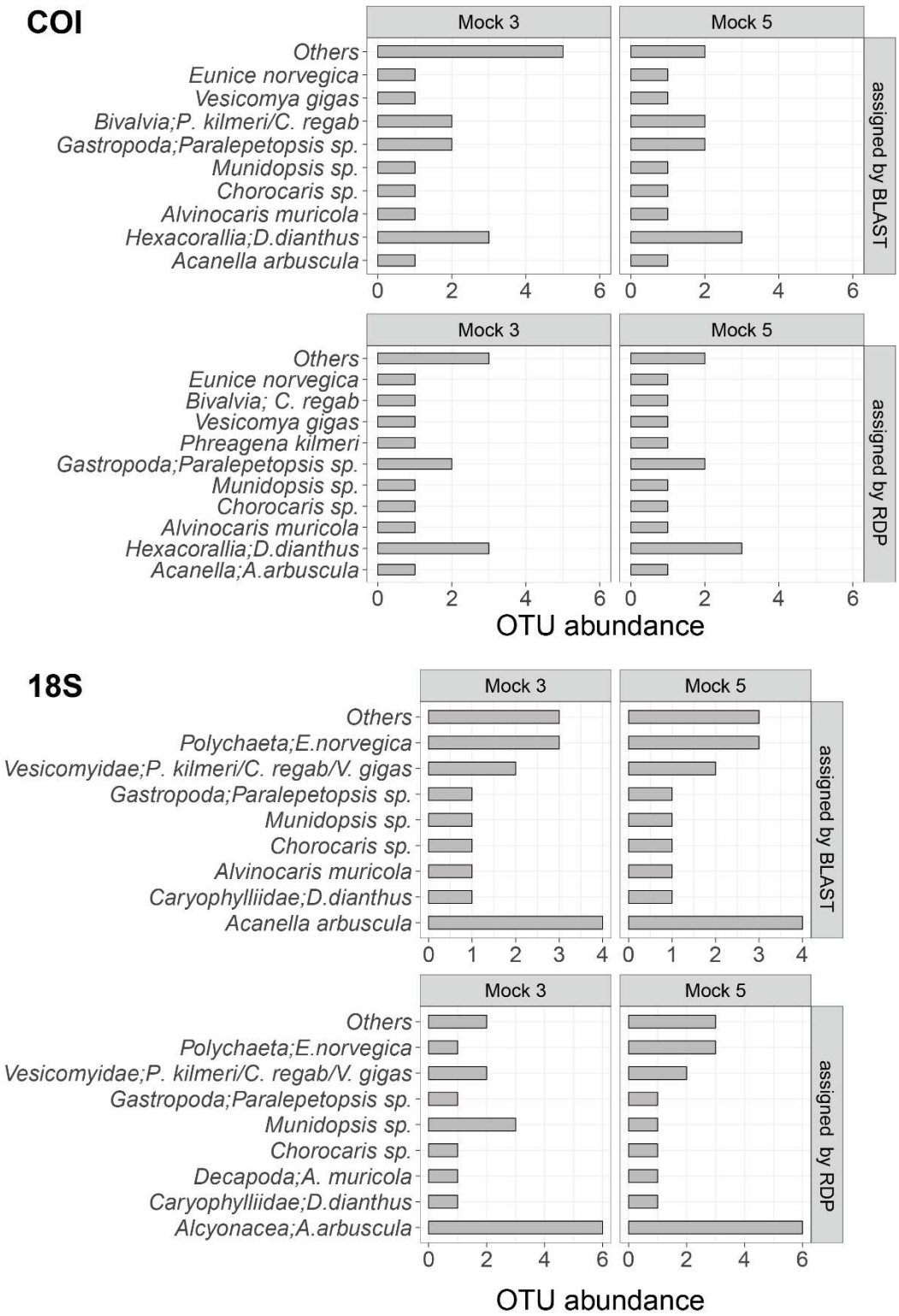


Figure S3. Performance of RDP and BLAST taxonomic assignments methods on two mock communities of ten deep-sea species. The two mock samples were analysed within a dataset of 42 deep-sea samples, processed via the DADA2 metabarcoding pipeline, clustered with swarm at *d=1*, abundance-renormalized, and curated with LULU at 90% (*minimum ratio* = 100) for 18S and 84% (*minimum ratio* = 1) for COI. Only taxonomic assignments reliable at phylum-level were retained during refining (BLAST hit identity ≥ 86% for 18S and ≥ 80% for COI, RDP assignments phylum-level bootstraps ≥ 80%). Silva132 was used as a reference database for 18S and MIDORI-UNIQUE, subsampled to marine-only taxa was used for COI. Cluster abundances were calculated on rarefied datasets.

Supplementary material Chapter III.

Supplemental materials and methods

Eukaryotic 18S-V1 rRNA gene amplicon generation

Eukaryotic 18S-V1V2 barcodes were generated using the SSUF04 (5'-GCTTGTCTCAAAGATTAAGCC-3') and SSUR22*mod* (5'-CCTGCTGCCTTCCTTRGA-3') primers (Sinniger et al. 2016) and the *Phusion* High Fidelity PCR Master Mix with GC buffer (ThermoFisher Scientific, Waltham, MA, USA). The PCR reactions (25 µL final volume) contained 2.5 ng or less of DNA template with 0.4 µM concentration of each primer, 3% of DMSO, and 1X Phusion Master Mix.

PCR amplifications (98 °C for 30 s; 25 cycles of 10 s at 98 °C, 30 s at 45 °C, 30 s at 72 °C; and 72 °C for 10 min) of all samples were carried out in triplicate in order to smooth the intra-sample variance while obtaining sufficient amounts of amplicons for Illumina sequencing. Amplicon triplicates were pooled and PCR products were purified using 1X AMPure XP beads (Beckman Coulter, Brea, CA, USA) cleanup. Aliquots of purified amplicons were run on an Agilent Bioanalyzer using the DNA High Sensitivity LabChip kit (Agilent Technologies, Santa Clara, CA, USA) to check their lengths, and quantified with a Qubit fluorometer (Invitrogen, Carlsbad, CA, USA).

Eukaryotic 18S-V4 rRNA gene amplicon generation

Eukaryotic 18S-V4 barcodes were generated using the TAReukF1 (5'-CCAGCASCYGCGGTAATTCC-3') and TAReukR (5'-ACTTTCGTTCTTGATYRA-3') primers (Stoeck et al. 2010). Triplicate PCR reactions were prepared as described above, but amplification was performed by a nested PCR with the first annealing temperature being 53°C for 10 cycles, followed by 48°C for 15 cycles. After PCR product cleanup using 1X AMPure XP beads, amplicon lengths and amounts were checked as described above.

Prokaryotic 16S-V4V5 rRNA gene amplicon generation

Prokaryotic barcodes were generated using the 515F-Y (5′-GTGYCAGCMGCCGCGGTAA-3′) and 926R (5′- CCGYCAATTYMTTTRAGTTT-3′) primers (Parada et al. 2016). Triplicate PCR reactions were prepared as described above for 18S-V1V2, but annealing temperature was at 53 °C. After PCR product cleanup using 1X AMPure XP beads, amplicon lengths and amounts were checked as described above.

Eukaryotic COI gene amplicon generation

Metazoan COI barcodes were generated using the mlCOIintF 5'-GGWACWGGWTGAACWGTWTAYCCYCC-3'and jgHCO2198 5'-TAIACYTCIGGRTGICCRAARAAYCA-3' primers (Leray et al. 2013). The PCR reactions (20 µL final volume) contained 2.5 ng or less of total DNA template with 0.5 µM final concentration of each primer, 3% of DMSO, 0.175 mM final concentration of dNTPs, and 1X Advantage 2 Polymerase Mix (Takara Bio, Kusatsu, Japan). Nested PCR amplifications were carried out in triplicates and consisted of an initial denaturation at 95 °C for 10 min, and 16 cycles of 10 s at 95°C, 30 s at 62 °C (−1°C per cycle), 60 s at 68 °C followed by 15 cycles of 95 °C for 10 s, 30 s at 46°C, 68 °C for 60 s, and a final extension of 68 °C for 7 min.

Amplicon library preparation

One hundred ng of amplicons were directly end-repaired, A-tailed and ligated to Illumina adapters on a Biomek FX Laboratory Automation Workstation (Beckman Coulter, Brea, CA, USA). Library amplification was performed using a Kapa Hifi HotStart NGS library Amplification kit (Kapa Biosystems, Wilmington, MA, USA) with the same cycling conditions applied for all metagenomic libraries and purified using 1X AMPure XP beads.

Sequencing library quality control

Libraries were quantified by Quant-iT dsDNA HS assay kits using a Fluoroskan Ascent microplate fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and then by qPCR with the KAPA Library Quantification Kit for Illumina Libraries (Kapa Biosystems, Wilmington, MA, USA) on a MxPro instrument (Agilent Technologies, Santa Clara, CA, USA). Library profiles were assessed using a high-throughput microfluidic capillary electrophoresis system (LabChip GX, Perkin Elmer, Waltham, MA, USA).

Sequencing procedure

Library concentrations were normalized to 10 nM by addition of 10 mM Tris-Cl (pH 8.5) and applied to cluster generation according to the Illumina Cbot User Guide (Part # 15006165). Amplicon libraries are characterized by low diversity sequences at the beginning of the reads due to the presence of the primer sequence. Low-diversity libraries can interfere in correct cluster identification, resulting in a drastic loss of data output. Therefore, loading concentrations

of libraries were decreased (8–9 pM instead of 12–14 pM for standard libraries) and PhiX DNA spike-in was increased (20% instead of 1%) in order to minimize the impacts on the run quality.

Libraries were sequenced on HiSeq2500 (System User Guide Part # 15035786) instruments (Illumina, San Diego, CA, USA) in 250 base pairs paired-end mode.

Supplemental tables

Table S1. Sampling sites and their GPS locations and associated habitats. MEDWAVES: MEDiterranean outflow WAter and Vulnerable EcosystemS.

Site	Cruise	Depth (m)	Latitude	Longitude	Habitat	Region
MDW-ST179; Seco de los Olivos	MEDWAVES	729	36.4808	-2.8945	Seamount	Western Mediterranean
MDW-ST23; Gazul	MEDWAVES	470	36.5605	-6.9498	Mud volcano	Gibraltar Strait
MDW-ST38; Ormonde	MEDWAVES	1,920	36.8442	-11.3025	Seamount	North Atlantic
MDW-ST117; Formigas	MEDWAVES	1,325	37.34	-24.7552	Seamount	North Atlantic
MRM-ST48; Mohn's Treasure	MarMine	2,826	73.4598	7.2184	Hydrothermal vent	Arctic

Table S2. Nucleic acid concentrations of samples from each of the five moleccular processing methods evaluated in this study. Original extracts were normalised to 0.25 ng/µL and 10 µL of standardized samples were used in PCR.

Sample name	Extraction kit	Sample type	DNA preparation method	Sediment amount for extraction	Concentration of original extract (ng/µL)
Mock community 3	CTAB extraction	mock DNA sample	mass-balanced mix of 10 species	na	5.0
Mock community 5	CTAB extraction	mock DNA sample	mass-balanced mix of 10 species	na	5.0
MDW_ST179_CT1_0_1	PowerMax Soil DNA Isolation Kit	DNA	crude extract	10g	8.1
MDW_ST179_CT2_0_1	PowerMax Soil DNA Isolation Kit	DNA	crude extract	10g	11.0
MDW_ST179_CT3_0_1	PowerMax Soil DNA Isolation Kit	DNA	crude extract	10g	9.7
MDW_ST117_CT1_0_1	PowerMax Soil DNA Isolation Kit	DNA	crude extract	10g	0.9
MDW_ST117_CT2_0_1	PowerMax Soil DNA Isolation Kit	DNA	crude extract	10g	2.2
MDW_ST117_CT3_0_1	PowerMax Soil DNA Isolation Kit	DNA	crude extract	10g	0.9
MDW_ST38_CT1_0_1	PowerMax Soil DNA Isolation Kit	DNA	crude extract	9.5g	0.2
MDW_ST38_CT2_0_1	PowerMax Soil DNA Isolation Kit	DNA	crude extract	10g	4.7
MDW_ST38_CT3_0_1	PowerMax Soil DNA Isolation Kit	DNA	crude extract	10g	4.3
MDW_ST23_CT1_0_1	PowerMax Soil DNA Isolation Kit	DNA	crude extract	10g	1.4
MDW_ST23_CT2_0_1	PowerMax Soil DNA Isolation Kit	DNA	crude extract	10g	9.2
MDW_ST23_CT3_0_1	PowerMax Soil DNA Isolation Kit	DNA	crude extract	9g	0.7
MRM_ST48_PC09_0_1	PowerMax Soil DNA Isolation Kit	DNA	crude extract	10g	6.5
MRM_ST48_PC15_0_1	PowerMax Soil DNA Isolation Kit	DNA	crude extract	10g	8.2
MRM_ST48_PC16_0_1	PowerMax Soil DNA Isolation Kit	DNA	crude extract	8.7g	4.5
MDW_ST179_CT1_0_1	PowerMax Soil DNA Isolation Kit	DNA	EtOH-reconcentrated	10g	30.3
MDW_ST179_CT2_0_1	PowerMax Soil DNA Isolation Kit	DNA	EtOH-reconcentrated	10g	6.4
MDW_ST179_CT3_0_1	PowerMax Soil DNA Isolation Kit	DNA	EtOH-reconcentrated	10g	16.3
MDW_ST117_CT1_0_1	PowerMax Soil DNA Isolation Kit	DNA	EtOH-reconcentrated	10g	8.9
MDW_ST117_CT2_0_1	PowerMax Soil DNA Isolation Kit	DNA	EtOH-reconcentrated	10g	15.5
MDW_ST117_CT3_0_1	PowerMax Soil DNA Isolation Kit	DNA	EtOH-reconcentrated	10g	7.0
MDW_ST38_CT1_0_1	PowerMax Soil DNA Isolation Kit	DNA	EtOH-reconcentrated	9.5g	1.0
MDW_ST38_CT2_0_1	PowerMax Soil DNA Isolation Kit	DNA	EtOH-reconcentrated	10g	23.7
MDW_ST38_CT3_0_1	PowerMax Soil DNA Isolation Kit	DNA	EtOH-reconcentrated	10g	23.6
MDW_ST23_CT1_0_1	PowerMax Soil DNA Isolation Kit	DNA	EtOH-reconcentrated	10g	7.3
MDW_ST23_CT2_0_1	PowerMax Soil DNA Isolation Kit	DNA	EtOH-reconcentrated	10g	38.7
MDW_ST23_CT3_0_1	PowerMax Soil DNA Isolation Kit	DNA	EtOH-reconcentrated	9g	3.1
MRM_ST48_PC09_0_1	PowerMax Soil DNA Isolation Kit	DNA	EtOH-reconcentrated	10g	31.4
MRM_ST48_PC15_0_1	PowerMax Soil DNA Isolation Kit	DNA	EtOH-reconcentrated	10g	32.7
MRM_ST48_PC16_0_1	PowerMax Soil DNA Isolation Kit	DNA	EtOH-reconcentrated	8.7g	26.4
MDW_ST179_CT1_0_1	PowerMax Soil DNA Isolation Kit	DNA	size selected	10g	32.6
MDW_ST179_CT2_0_1	PowerMax Soil DNA Isolation Kit	DNA	size selected	10g	35.3
MDW_ST179_CT3_0_1	PowerMax Soil DNA Isolation Kit	DNA	size selected	10g	33.0
MDW_ST117_CT1_0_1	PowerMax Soil DNA Isolation Kit	DNA	size selected	10g	3.2
MDW_ST117_CT2_0_1	PowerMax Soil DNA Isolation Kit	DNA	size selected	10g	9.0
MDW_ST117_CT3_0_1	PowerMax Soil DNA Isolation Kit	DNA	size selected	10g	2.4
MDW_ST38_CT1_0_1	PowerMax Soil DNA Isolation Kit	DNA	size selected	9.5g	0.2
MDW_ST38_CT2_0_1	PowerMax Soil DNA Isolation Kit	DNA	size selected	10g	19.3
MDW_ST38_CT3_0_1	PowerMax Soil DNA Isolation Kit	DNA	size selected	10g	14.4
MDW_ST23_CT1_0_1	PowerMax Soil DNA Isolation Kit	DNA	size selected	10g	2.7
MDW_ST23_CT2_0_1	PowerMax Soil DNA Isolation Kit	DNA	size selected	10g	14.0
MDW_ST23_CT3_0_1	PowerMax Soil DNA Isolation Kit	DNA	size selected	9g	1.2
MRM_ST48_PC09_0_1	PowerMax Soil DNA Isolation Kit	DNA	size selected	10g	18.9
MRM_ST48_PC15_0_1	PowerMax Soil DNA Isolation Kit	DNA	size selected	10g	23.1
MRM_ST48_PC16_0_1	PowerMax Soil DNA Isolation Kit	DNA	size selected	8.7g	16.5
MDW_ST179_CT1_0_1	RNeasy PowerSoil DNA elution kit	DNA	crude extract	4g	33.2
MDW_ST179_CT2_0_1	RNeasy PowerSoil DNA elution kit	DNA	crude extract	3.9g	33.3
MDW_ST179_CT3_0_1	RNeasy PowerSoil DNA elution kit	DNA	crude extract	4g	33.3
MDW_ST117_CT1_0_1	RNeasy PowerSoil DNA elution kit	DNA	crude extract	3.2g	0.9
MDW_ST117_CT2_0_1	RNeasy PowerSoil DNA elution kit	DNA	crude extract	4.7g	0.4
MDW_ST117_CT3_0_1	RNeasy PowerSoil DNA elution kit	DNA	crude extract	4g	0.1
MDW_ST38_CT1_0_1	RNeasy PowerSoil DNA elution kit	DNA	crude extract	4g	1.9
MDW_ST38_CT2_0_1	RNeasy PowerSoil DNA elution kit	DNA	crude extract	3.7g	3.0
MDW_ST38_CT3_0_1	RNeasy PowerSoil DNA elution kit	DNA	crude extract	4.5g	5.4
MDW_ST23_CT1_0_1	RNeasy PowerSoil DNA elution kit	DNA	crude extract	4g	36.7
MDW_ST23_CT2_0_1	RNeasy PowerSoil DNA elution kit	DNA	crude extract	4g	31.9
MDW_ST23_CT3_0_1	RNeasy PowerSoil DNA elution kit	DNA	crude extract	5g	34.7
MRM_ST48_PC09_0_1	RNeasy PowerSoil DNA elution kit	DNA	crude extract	4g	36.7
MRM_ST48_PC15_0_1	RNeasy PowerSoil DNA elution kit	DNA	crude extract	4g	35.3
MRM_ST48_PC16_0_1	RNeasy PowerSoil DNA elution kit	DNA	crude extract	4.5g	34.7
MDW_ST179_CT1_0_1	RNeasy PowerSoil	cDNA	crude extract	4g	6.07
MDW_ST179_CT2_0_1	RNeasy PowerSoil	cDNA	crude extract	3.9g	10.60
MDW_ST179_CT3_0_1	RNeasy PowerSoil	cDNA	crude extract	4g	5.39
MDW_ST117_CT1_0_1	RNeasy PowerSoil	cDNA	crude extract	3.2g	0.927
MDW_ST117_CT2_0_1	RNeasy PowerSoil	cDNA	crude extract	4.7g	0.7
MDW_ST117_CT3_0_1	RNeasy PowerSoil	cDNA	crude extract	4g	0.56
MDW_ST38_CT1_0_1	RNeasy PowerSoil	cDNA	crude extract	4g	2.94
MDW_ST38_CT2_0_1	RNeasy PowerSoil	cDNA	crude extract	3.7g	3.35
MDW_ST38_CT3_0_1	RNeasy PowerSoil	cDNA	crude extract	4.5g	5.97
MDW_ST23_CT1_0_1	RNeasy PowerSoil	cDNA	crude extract	4g	8.2
MDW_ST23_CT2_0_1	RNeasy PowerSoil	cDNA	crude extract	4g	4.44
MDW_ST23_CT3_0_1	RNeasy PowerSoil	cDNA	crude extract	5g	6.8
MRM_ST48_PC09_0_1	RNeasy PowerSoil	cDNA	crude extract	4g	24.1
MRM_ST48_PC15_0_1	RNeasy PowerSoil	cDNA	crude extract	4g	15.1
MRM_ST48_PC16_0_1	RNeasy PowerSoil	cDNA	crude extract	4.5g	10.4
MOBIO_10g_DNA_extraction_blank	PowerMax Soil DNA Isolation Kit	extraction and field blank	na	na	0.0
MOBIO_2g_DNA_extraction_blank	RNeasy PowerSoil DNA elution kit	extraction blank	na	na	0.0
MOBIO_2g_RNA_extraction_blank	RNeasy PowerSoil	extraction blank	na	na	0.2

Table S3. Primers used in this study, targeting metazoans with the COI and 18S-V1V2 loci, micro-eukaryotes with the 18S-V4 barcode, and prokaryotes with the 16S-V4V5 marker.

Locus	**Target and *specificity***	**Primer forward Primer reverse**	**Short name**	**Sequence (5'-3')**	**Amplicon size (bp)**	**Reference**
COI	Eukaryotes	mlCOIintF	COI-F	GGWACWGGWTGAACWGTWTAYCCYCC	313	Leray et al., 2013
	pref. metazoans	jgHCO2198	COI-R	TAIACYTCIGGRTGICCRAARAAYCA		
18S-V1V2	Eukaryotes	SSUF04	18S-V1F	GCTTGTCTCAAAGATTAAGCC	330-390	Sinniger et al., 2016
	pref. metazoans	SSURmod	18S-V1-R	CCTGCTGCCTTCCTTRGA		
18S-V4	Eukaryotes	V4F (TAReukFWD1)	18S-V4-F	CCAGCASCYGCGGTAATTCC	350-410	Stoeck et al., 2010
	all	V4R (TAReukREV3)	18S-V4-R	ACTTTCGTTCTTGATYRA		
16S-V4V5	Prokaryotes	515f	16S-F	GTGYCAGCMGCCGCGGTAA	350-390	Parada et al., 2016
	Pref. Eubacteria	926r	16S-R	CCGYCAATTYMTTTRAGTTT		

Table S4. ABYSS metabarcoding pipeline.

Process	**Software**	**Script(s) and command(s)**
Raw reads preprocessing for ligation data	Abyss-preprocessing: separate forward and reverse reads in each run, and re-pair reads	extractR1R2.pbs using cutadapt v1.18 (-e 0.17 for 18S-V1 and 0.27 for COI , -O length of primer -1) and BBMAP Repair v38.22
Read quality-filtering	Dada2 v.1.10	filterAndTrim() in dada2main.R maxEE=2, maxN=0, truncQ=11, truncLen=220 (18S, 16S) or 200 (COI)
Read error learning	Dada2 v.1.10	learnErrors() in dada2main.R nbases=1e8, multithread=TRUE, randomize=TRUE
Read dereplicating	Dada2 v.1.10	derepFastq() in dada2main.R
Read correction	Dada2 v.1.10	dada() in dada2main.R
Read merging	Dada2 v.1.10	mergePairs() in dada2main.R minOverlap=12, maxMismatch=0
Make sequence table and filter by length	Dada2 v.1.10	makeSequenceTable() in dada2main.R seqtab[,nchar(colnames(seqtab)) %in% seq(lengthMin,lengthMax)] lengthMin= 330 (18S-V1), 300 (COI), 350 (18S-V4), 350 (16S) lengthMax= 390 (18S-V1), 326 (COI), 410 (18S-V4), 390 (16S)
Chimera removal	Dada2 v.1.10	removeBimeraDenovo() in dada2main.R
Taxonomic assignment with RDP Classifier	Dada2 v.1.10	assignTaxonomy () in dada2outputfiles.R minBoot=50, outputBootstraps=TRUE
Taxonomic assignment with BLAST+	blastn (megablast) v.2.6.0	blast.pbs -outfmt 11 -qcov_hsp_perc 80 -perc_identity 70 -max_hsps 1, -evalue 1e-5, then merge BLAST and RDP taxonomies using concat_blast_rdp_tax.pbs
Clustering (optional)	FROGS v.2.0.0	clustering.py with d=4 for 18S-V1V2 and d=6 for COI, remova_chimera.py, affiliation_OTU_identite_couverture.py
Blank correction	Rscript	Data_refining.Rmd using packages decontam v.1.2.1 and phyloseq v.1.26.0
Removal of unassigned and non-target clusters		
Deletion of defective samples (< 10,000 target reads)		
Tag-switching renormalisation	Rscript	owi_renormalize.R
LULU curation	LULU v.0.1	lulu() in lulu_final.R minimum_ratio_type = "min", minimum_ratio = 1, minimum_match =84, minimum_relative_cooccurence =0.93

Table S5. Tests for homogeneity of multivariate dispersions for the 4 genes studied. The tests for performed with 9999 permutations on Jaccard distances for 18S-V1 and COI, and on Bray-Curtis distances for 18S-V4 and 16S. Significant p values are in bold. For pairwise comparisons, DNA 10g comprises all processing methods based on DNA extracted from ~10g of sediment, and significance codes are: p<0.001: '***'; p<0.01: '**'; p<0.05: '*'. In cases of significantly different dispersions, PERMANOVAs were performed on balanced datasets.

LOCUS	***df***	**SS**	**F-value**	***p-value***	**significant parwise comparisons**
18S-V1					
Molecular processing	4	0.00270	1.26	0.3	NA
Residuals	67	0.0361			
COI					
Molecular processing	4	0.00349	3.26	**0.017**	RNA 2g / DNA 10g * DNA 2g / DNA 10g S-S *
Residuals	63	0.0169			
18S-V4					
Molecular processing	4	0.00222	0.78	0.54	NA
Residuals	64	0.0450			
16S-V4V5					
Molecular processing	4	0.00875	0.93	0.44	NA
Residuals	69	0.161			

Table S6. Number of reads and clusters (ASVs for 18S-V4 and 16S, OTUs for 18S-V1 and COI) obtained at different analysis steps, depending on molecular processing category (DNA 10g: crude DNA extracts from ~10 g of sediment with the PowerMax Soil kit; DNA 10g EtOH rec.: ethanol reconcentrated 10g DNA extracts; DNA 10g S-S: size-selected 10g DNA extracts; DNA/RNA 2g: crude DNA/RNA extracts from ~2g of sediment with the RNeasy PowerSoil kit). Data refining was performed in R, based on BLAST assignments obtained using the Silva v132 database for 18S and 16S loci, and on the MIDORI database for COI. Final number of target reads represent the number of target-taxa reads after data refining (abundance renormalisation for 18S and 16S loci, abundance renormalisation and LULU curation for COI). Final number of target clusters are the corresponding ASVs for 18S-V4 and 16S, and the corresponding OTUs for 18S-V1 and COI.

Sample type	Number of samples	Raw reads	Quality-filtered reads	Merged reads	Length-filtered reads	Non chimeric reads	% chimeras	% reads retained	Number of clusters before refining	Number of samples after refining	Final number of target reads	Final number of target clusters
LOCUS												
18S-V1												
Control	5	2,921,651	1,654,366	1,601,427	1,380,613	1,379,141	0.11	47	42,876	0	16,157,973	6,031
DNA 10g	15	12,969,202	9,978,581	9,141,929	8,577,414	8,463,482	1.33	65		15		
DNA 10g EtOH rec.	15	13,646,370	10,577,221	9,757,954	9,271,161	9,129,915	1.52	67		15		
DNA 10g S-S	15	11,735,871	8,938,926	7,990,574	7,403,206	7,328,224	1.01	62		15		
DNA 2g	14	8,476,073	6,402,605	5,840,025	5,215,391	5,168,422	0.9	61		14		
Positive Control (Metazoa only)	2	2,096,631	1,607,219	1,438,424	1,432,399	1,293,985	9.66	62		2		
RNA 2g	14	18,130,054	14,322,872	13,311,099	12,180,021	11,591,521	4.83	64		13		
COI												
Control	7	642,571	414,329	410,866	410,260	410,189	0.02	64	45,508	0	10,977,614	4,333
DNA 10g	15	13,804,664	11,871,147	11,645,233	10,515,311	10,437,446	0.74	76		15		
DNA 10g EtOH rec.	15	12,735,345	10,966,940	10,758,928	9,634,212	9,560,863	0.76	75		15		
DNA 10g S-S	15	13,172,416	11,357,075	11,141,979	10,019,802	9,948,428	0.71	76		15		
DNA 2g	14	10,992,972	8,962,857	8,748,635	7,478,953	7,439,814	0.52	68		14		
Positive Control (Metazoa only)	2	1,482,785	1,261,045	1,253,408	1,252,485	1,226,728	2.06	83		2		
RNA 2g	9	8,085,884	6,548,055	6,405,869	5,780,511	5,749,188	0.54	71		9		
18S-V4												
Control	5	38,028	1,088	1,005	786	786	0	2	65,832	0	8,654,710	40,868
DNA 10g	15	5,108,793	3,852,156	3,244,507	3,081,831	3,073,436	0.27	60		15		
DNA 10g EtOH rec.	13	4,812,187	3,622,684	3,014,540	2,884,756	2,876,930	0.27	60		13		
DNA 10g S-S	15	3,675,283	2,779,263	2,334,265	2,222,758	2,216,696	0.27	60		15		
DNA 2g	13	2,569,170	1,853,940	1,539,838	1,422,958	1,419,415	0.25	55		13		
RNA 2g	13	14,024,345	10,695,784	8,260,707	6,978,244	6,876,335	1.46	49		13		
16S-V4V5												
Control	5	1,100,024	815,505	692,998	687,737	686,244	0.22	62	148,797	0	21,740,351	138,478
DNA 10g	15	6,228,145	4,351,718	3,445,745	3,436,831	3,311,742	3.64	53		15		
DNA 10g EtOH rec.	15	7,400,388	5,167,853	4,033,571	4,024,943	3,861,002	4.07	52		15		
DNA 10g S-S	15	7,163,763	5,039,022	4,045,902	4,037,814	3,891,760	3.62	54		15		
DNA 2g	15	7,788,742	5,409,390	4,582,930	4,573,875	4,428,719	3.17	57		15		
RNA 2g	14	15,248,792	10,719,709	7,858,978	7,854,365	7,438,875	5.29	49		14		

Table S7 Community differentiation between RNA and DNA molecular processing methods, using either RNA/DNA extracted jointly from ~2 g of sediment (RNA 2g/DNA 2g) or DNA extracted from ~10g of sediment (DNA 10g) in five deep-sea sites using four barcode markers targeting metazoans (COI, 18S-V1), micro-eukaryotes (18S-V4), and prokaryotes (16S-V4V5). PERMANOVAs were calculated on normalised datasets by permuting 10,000 times with Site as a blocking factor, using Jaccard distances for 18S-V1 and COI, and Bray-Curtis distances for 18S-V4 and 16S. Significant p values are in bold.

LOCUS	**18S-V1**		**COI**		**18S-V4**		**16S-V4V5**	
	R^2	***p-value***	**R^2**	***p-value***	**R^2**	***p-value***	**R^2**	***p-value***
DNAvsRNA	0.04	**< 0.001**	0.04	**< 0.01**	0.05	**< 0.001**	0.05	**< 0.001**
Kit	0.03	**< 0.001**	0.038	0.09	0.027	**< 0.01**	0.01	0.11
Site	0.23	**< 0.001**	0.18	**< 0.001**	0.33	**< 0.001**	0.54	**< 0.001**

Supplemental figures

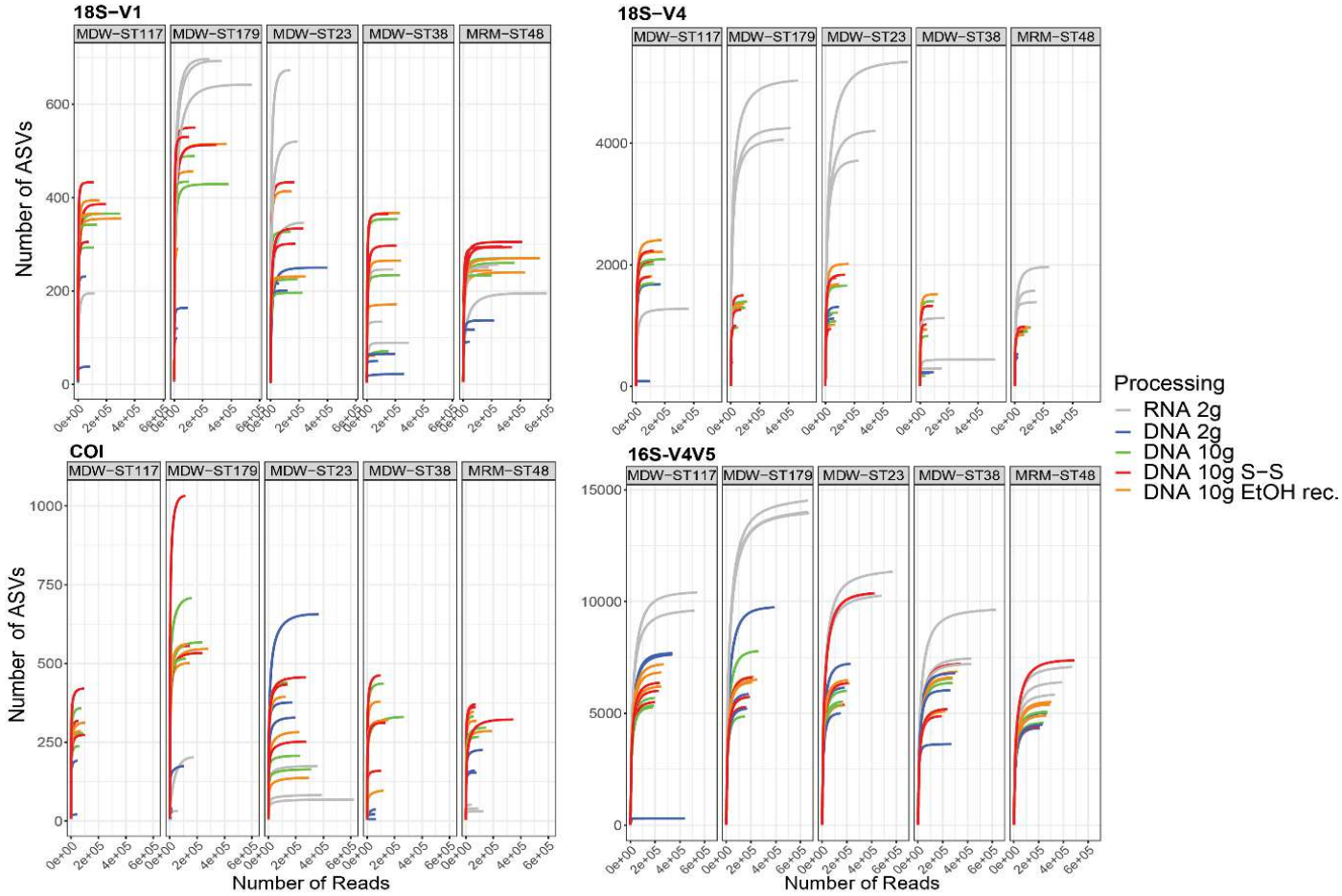


Figure S1. Rarefaction curves in deep-sea sediment samples from 5 sampling sites, processed with 5 molecular methods for producing metabarcoding inventories of metazoans (18S-V1, COI), micro-eukaryotes (18S-V4), and prokaryotes (16S-V4V5), showing a plateau is reached in all samples. DNA 10g: crude DNA extracts from ~10 g of sediment with the PowerMax Soil kit; DNA 10g EtOH rec.: ethanol reconcentrated 10g DNA extracts; DNA 10g S-S: size-selected 10g DNA extracts; DNA/RNA 2g: crude DNA/RNA extracts from ~2g of sediment with the RNeasy PowerSoil kit.

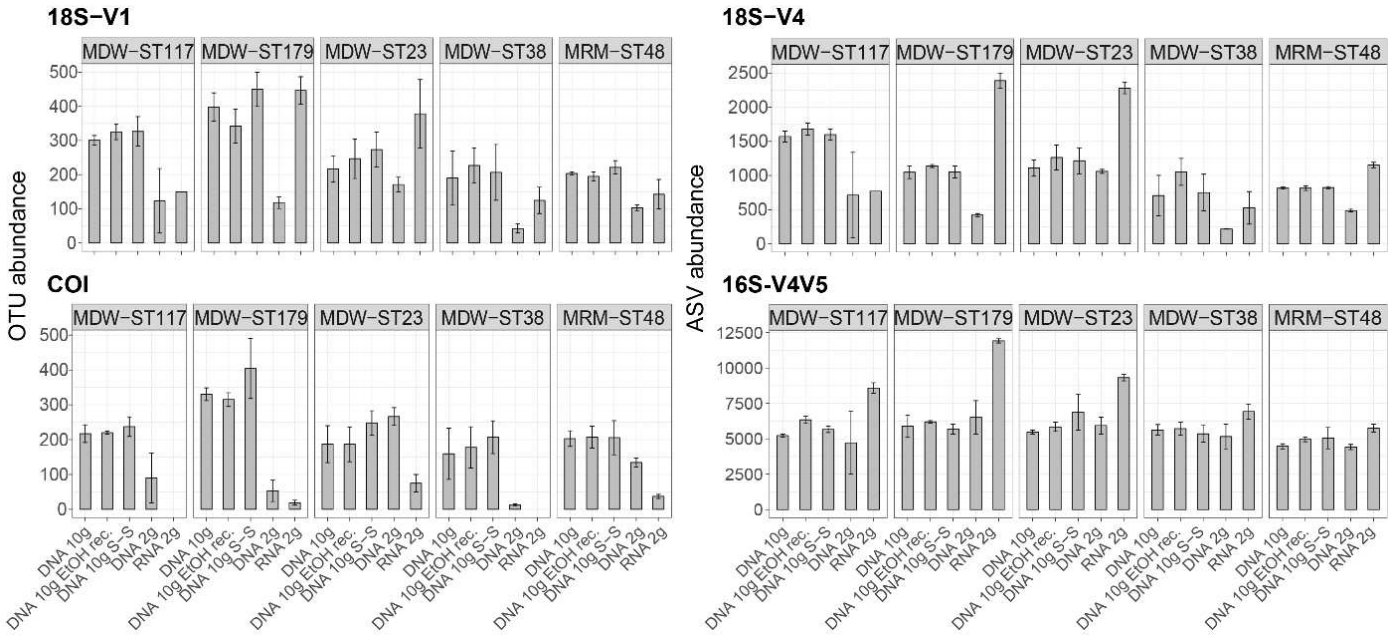


Figure S2. Mean number of metazoan OTUs (18S-V1, COI), protist (18S-V4) and prokaryote ASVs (16S-V4V5) recovered in each of the five sampling sites by the five molecular processing methods evaluated in this study (DNA 10g: crude DNA extracts from ~10 g of sediment with the PowerMax Soil kit; DNA 10g EtOH rec.: ethanol reconcentrated 10g DNA extracts; DNA 10g S-S: size-selected 10g DNA extracts; DNA/RNA 2g: crude DNA/RNA extracts from ~2g of sediment with the RNeasy PowerSoil kit). Cluster numbers were calculated on the rarefied datasets. Error bars represent standard errors.

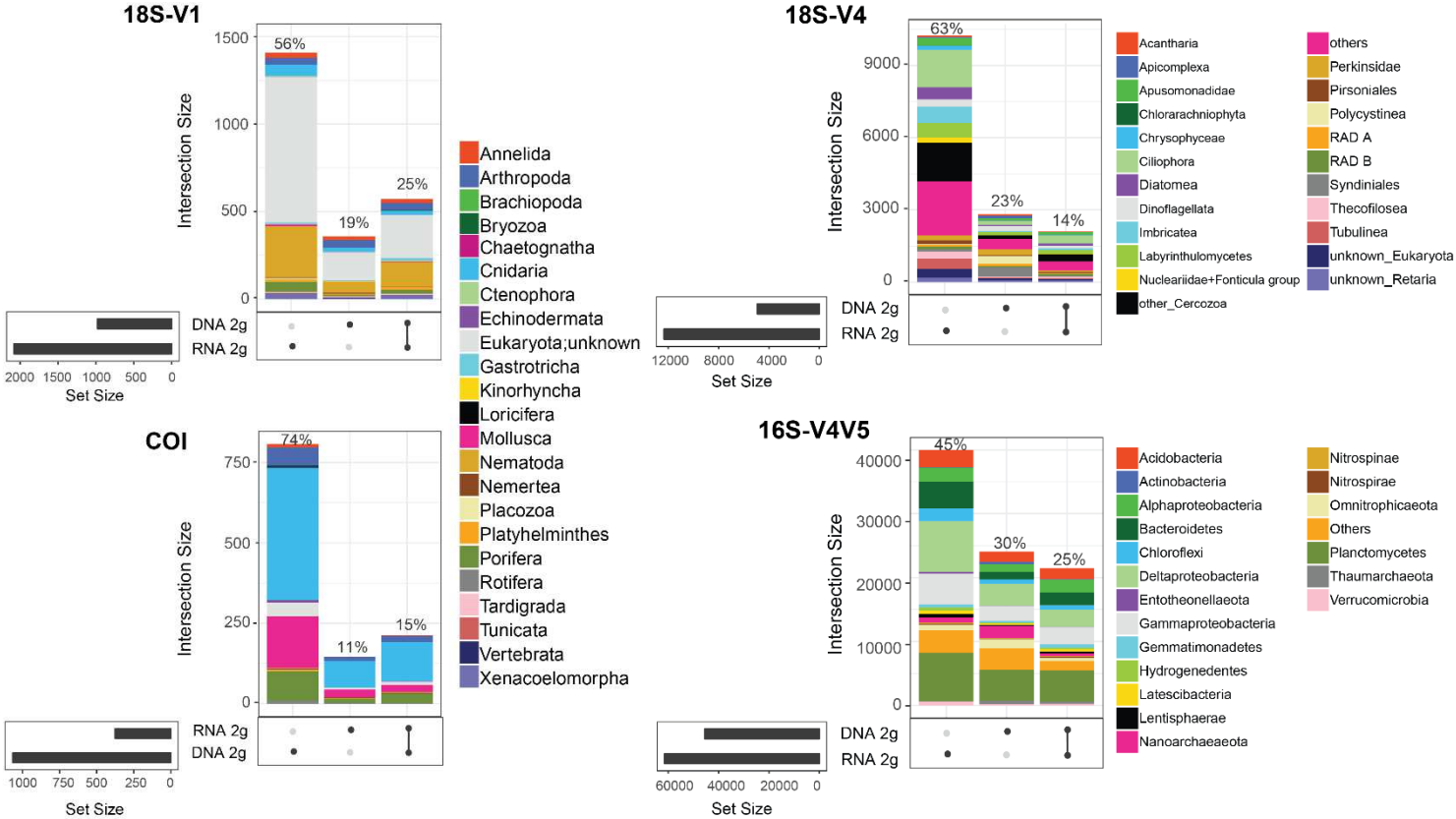


Figure S3. Shared and unique metazoan OTUs (18S-V1, COI), protozoan ASVs (18S-V4), and prokaryote ASVs (16S-V4V5) among the joint DNA and RNA datasets (DNA/RNA 2g: crude DNA/RNA extracts from ~2g of sediment with the RNeasy PowerSoil kit). Numbers were calculated on the rarefied datasets.

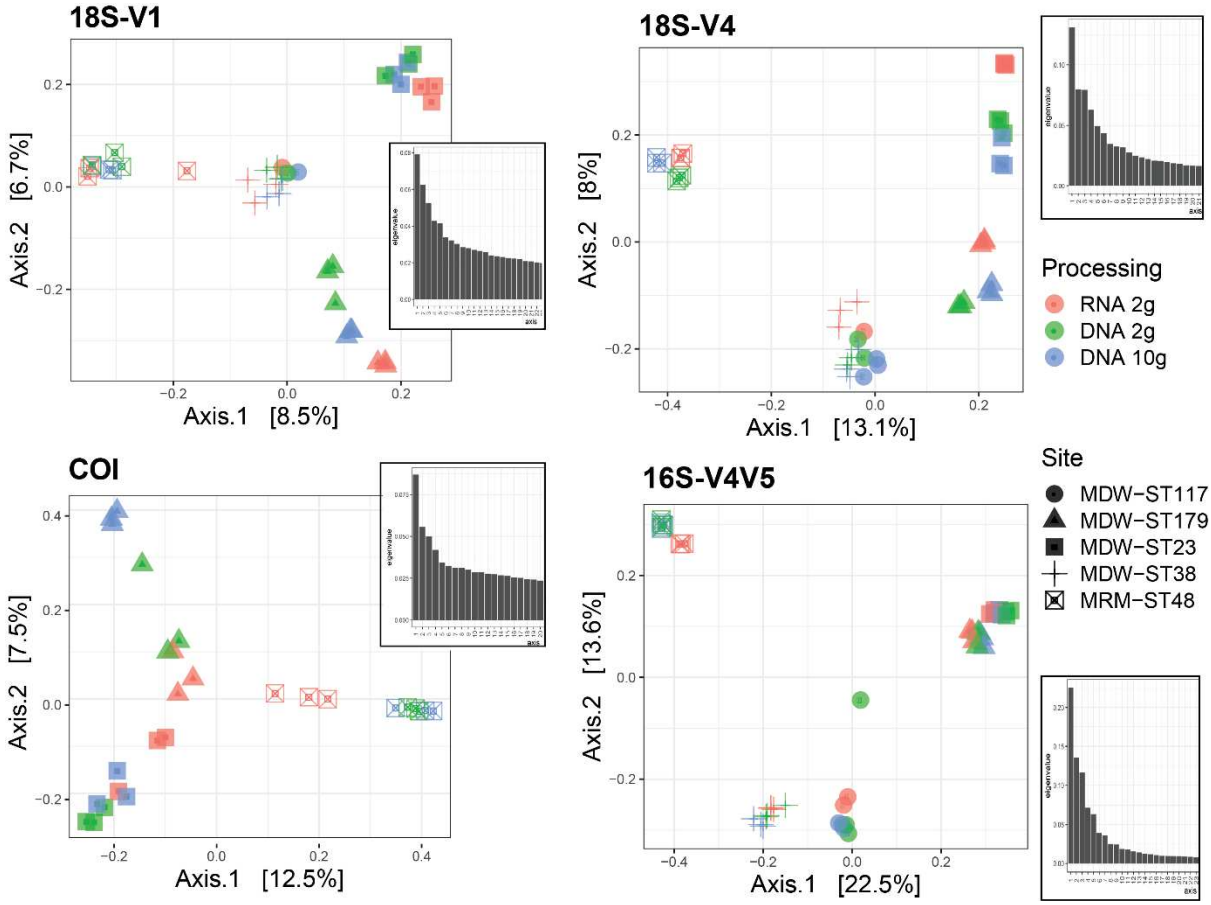


Figure S4. Community differences between RNA and DNA molecular processing methods using either RNA/DNA extracted jointly from ~2 g of sediment (RNA/DNA 2g) or DNA extracted from ~10g of sediment (DNA 10g) in five deep-sea sites using four barcode markers targeting metazoans (18S-V1, COI), micro-eukaryotes (18S-V4), and prokaryotes (16S-V4V5). PCoAs were calculated using Jaccard dissimilarities for metazoans and Bray-Curtis dissimilarities for unicellular organisms. The first two axes of the PCoAs shown here capture the main source of variation, the site variation. Scree plots of each ordination are shown in inserts, indicating that variation due to processing method is captured by secondary axes.

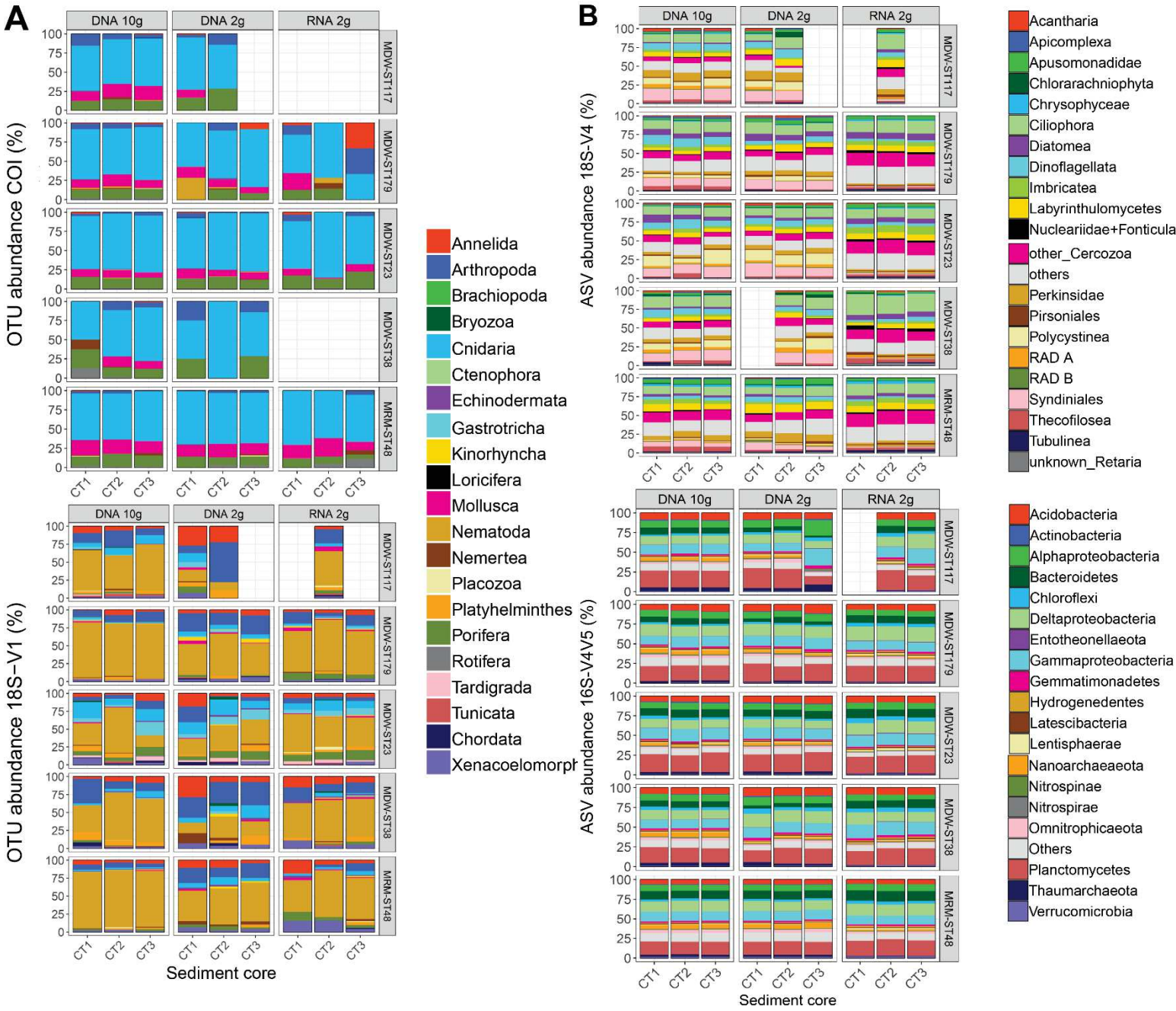


Figure S5. Patterns of relative cluster abundance resolved by metabarcoding results in triplicate sediment cores from five deep-sea sites by RNA and DNA molecular processing methods using RNA/DNA extracted jointly from ~2 g of sediment (RNA/DNA 2g) or DNA extracted from ~10g of sediment (DNA 10g), using four barcode markers targeting metazoans (**A**: 18S-V1, COI), micro-eukaryotes (**B**: 18S-V4), and prokaryotes (**B**: 16S-V4V5).

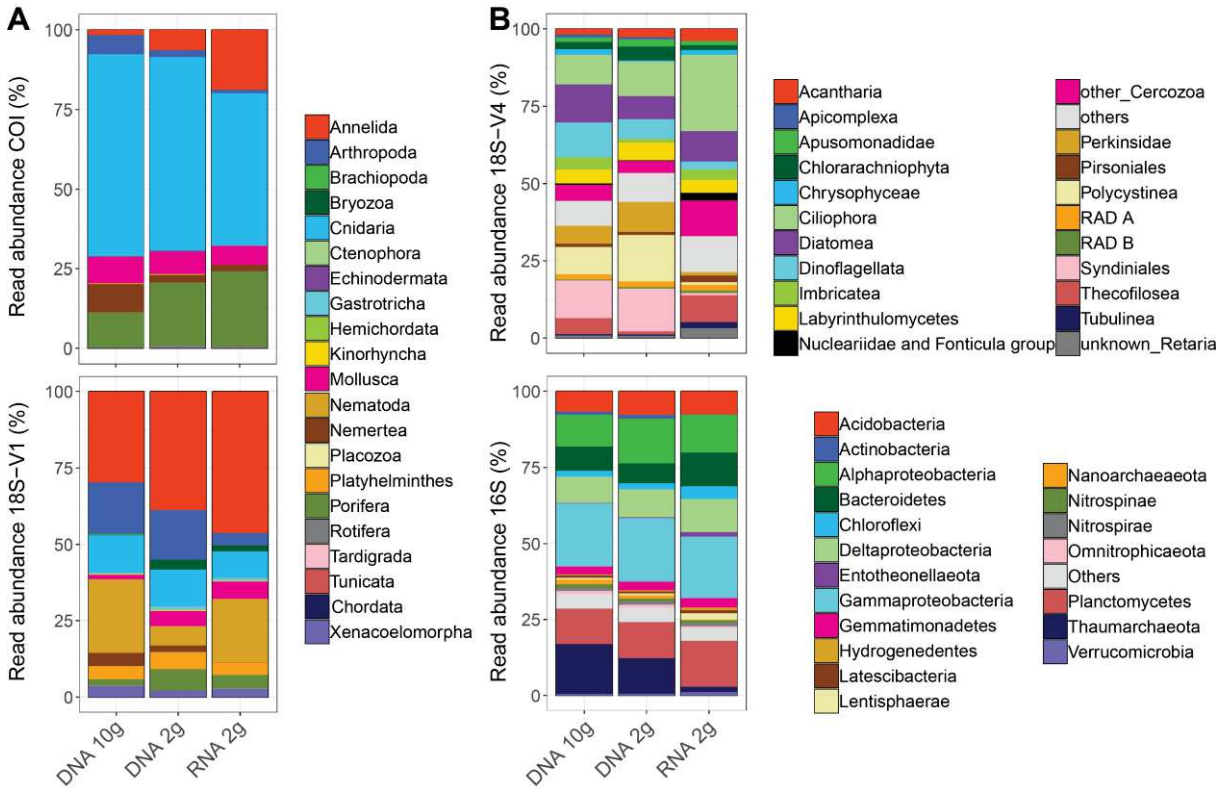


Figure S6. Patterns of relative read abundance resolved by metabarcoding results in five deep-sea sites by RNA and DNA molecular processing methods using RNA/DNA extracted jointly from ~2 g of sediment (RNA/DNA 2g) or DNA extracted from ~10g of sediment (DNA 10g), using four barcode markers targeting metazoans (**A**: 18S-V1, COI), micro-eukaryotes (**B**: 18S-V4), and prokaryotes (**B**: 16S). Values were calculated on balanced datasets.

Supplementary material Chapter IV.

Supplemental materials and methods

Eukaryotic 18S-V1V2 rRNA gene amplicon generation

Eukaryotic 18S-V1V2 barcodes were generated using the SSUF04 (5'-GCTTGTCTCAAAGATTAAGCC-3') and SSUR22*mod* (5'-CCTGCTGCCTTCCTTRGA-3') primers (Sinniger et al. 2016) and the *Phusion* High Fidelity PCR Master Mix with GC buffer (ThermoFisher Scientific, Waltham, MA, USA). The PCR reactions (25 µL final volume) contained 2.5 ng or less of DNA template with 0.4 µM concentration of each primer, 3% of DMSO, and 1X Phusion Master Mix.

PCR amplifications (98 °C for 30 s; 25 cycles of 10 s at 98 °C, 30 s at 45 °C, 30 s at 72 °C; and 72 °C for 10 min) of all samples were carried out in triplicate in order to smooth the intra-sample variance while obtaining sufficient amounts of amplicons for Illumina sequencing. Amplicon triplicates were pooled and PCR products were purified using 1X AMPure XP beads (Beckman Coulter, Brea, CA, USA) cleanup. Aliquots of purified amplicons were run on an Agilent Bioanalyzer using the DNA High Sensitivity LabChip kit (Agilent Technologies, Santa Clara, CA, USA) to check their lengths, and quantified with a Qubit fluorometer (Invitrogen, Carlsbad, CA, USA).

Eukaryotic 18S-V4 rRNA gene amplicon generation

Eukaryotic 18S-V4 barcodes were generated using the TAReukF1 (5'-CCAGCASCYGCGGTAATTCC-3') and TAReukR (5'-ACTTTCGTTCTTGATYRA-3') primers (Stoeck et al. 2010). Triplicate PCR reactions were prepared as described above, but amplification was performed by a nested PCR with the first annealing temperature being 53°C for 10 cycles, followed by 48°C for 15 cycles. After PCR product cleanup using 1X AMPure XP beads, amplicon lengths and amounts were checked as described above.

Prokaryotic 16S-V4V5 rRNA gene amplicon generation

Prokaryotic barcodes were generated using the 515F-Y (5′-GTGYCAGCMGCCGCGGTAA-3′) and 926R (5′- CCGYCAATTYMTTTRAGTTT-3′) primers (Parada et al. 2016). Triplicate PCR reactions were prepared as described above for

18S-V1V2, but annealing temperature was at 53 °C. After PCR product cleanup using 1X AMPure XP beads, amplicon lengths and amounts were checked as described above.

Eukaryotic COI gene amplicon generation

Metazoan COI barcodes were generated using the mlCOIintF 5'-GGWACWGGWTGAACWGTWTAYCCYCC-3'and jgHCO2198 5'-TAIACYTCIGGRTGICCRAARAAYCA-3' primers (Leray et al. 2013). The PCR reactions (20 µL final volume) contained 2.5 ng or less of total DNA template with 0.5 µM final concentration of each primer, 3% of DMSO, 0.175 mM final concentration of dNTPs, and 1X Advantage 2 Polymerase Mix (Takara Bio, Kusatsu, Japan). Nested PCR amplifications were carried out in triplicates and consisted of an initial denaturation at 95 °C for 10 min, and 16 cycles of 10 s at 95°C, 30 s at 62 °C (−1°C per cycle), 60 s at 68 °C followed by 15 cycles of 95 °C for 10 s, 30 s at 46°C, 68 °C for 60 s, and a final extension of 68 °C for 7 min.

Amplicon library preparation

Triplicate PCR reactions were pooled and 100 ng were directly end-repaired, A-tailed and ligated to Illumina adapters on a Biomek FX Laboratory Automation Workstation (Beckman Coulter, Brea, CA, USA). Library amplification was performed using a Kapa Hifi HotStart NGS library Amplification kit (Kapa Biosystems, Wilmington, MA, USA) with the same cycling conditions applied for all metagenomic libraries and purified using 1X AMPure XP beads.

Sequencing library quality control

Libraries were quantified by Quant-iT dsDNA HS assay kits using a Fluoroskan Ascent microplate fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and then by qPCR with the KAPA Library Quantification Kit for Illumina Libraries (Kapa Biosystems, Wilmington, MA, USA) on a MxPro instrument (Agilent Technologies, Santa Clara, CA, USA). Library profiles were assessed using a high-throughput microfluidic capillary electrophoresis system (LabChip GX, Perkin Elmer, Waltham, MA, USA).

Sequencing procedure

Library concentrations were normalized to 10 nM by addition of 10 mM Tris-Cl (pH 8.5) and applied to cluster generation according to the Illumina Cbot User Guide (Part # 15006165). Amplicon libraries are characterized by low diversity sequences at the beginning of the reads due to the presence of the primer sequence. Low-diversity libraries can interfere in correct cluster identification, resulting in a drastic loss of data output. Therefore, loading concentrations of libraries were decreased (8–9 pM instead of 12–14 pM for standard libraries) and PhiX DNA spike-in was increased (20% instead of 1%) in order to minimize the impacts on the run quality. Libraries were sequenced on HiSeq2500 (System User Guide Part # 15035786) instruments (Illumina, San Diego, CA, USA) in 250 base pairs paired-end mode.

Supplemental tables

Table S1. Sampling sites, their GPS locations, and associated habitats. Sieved sediment was sieved through five mesh sizes (1,000; 500; 250; 40; 20 µm), and DNA was extracted from each size fraction separately. An equimolar pool of the five DNA extracts of each size fraction was then made for PCR and sequencing of the sieved samples. Volume for PCR was always 10 µl, for template stock standardized at ≤0.25 ng/µl.

Sample name	ENA sample alias	Extraction kit	Sample type	Size fraction (µm)	Sample volume for DNA extraction	Concentration of original extract (ng/µL)	Depth (m)	Latitude	Longitude	Habitat	Region
ESSNAUT_PL06_CT2_0_1_rep1	eDNAB0000081	PowerMax Soil DNA Isolation Kit	DNA	NA	2 g	1.1	2,417	42.9422	6.7422	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL06_CT2_0_1_rep2	eDNAB0000081	PowerMax Soil DNA Isolation Kit	DNA	NA	2 g	1	2,417	42.9422	6.7422	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL07_CT2_0_1_rep1	eDNAB0000122	PowerMax Soil DNA Isolation Kit	DNA	NA	5.1 g	3.2	2,415	42.9423	6.7426	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL07_CT2_0_1_rep2	eDNAB0000122	PowerMax Soil DNA Isolation Kit	DNA	NA	5.1 g	3	2,415	42.9423	6.7426	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL11_CT2_0_1_rep1	eDNAB0000189	PowerMax Soil DNA Isolation Kit	DNA	NA	4 g	2.7	2,418	42.9423	6.7423	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL11_CT2_0_1_rep2	eDNAB0000189	PowerMax Soil DNA Isolation Kit	DNA	NA	4 g	2.6	2,418	42.9423	6.7423	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL06_CT4_0_1_pool_rep1	eDNAB0003158	PowerMax Soil DNA Isolation Kit	DNA	>1,000; 500-1,000; 250-500; 40-250; 20-40	2.2+0.7+1.5+2+2.2=7.6 g	0.4	2,417	42.9422	6.7422	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL06_CT4_0_1_pool_rep2	eDNAB0003158	PowerMax Soil DNA Isolation Kit	DNA	>1,000; 500-1,000; 250-500; 40-250; 20-40	2.2+0.7+1.5+2+2.2=7.6 g	0.6	2,417	42.9422	6.7422	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL07_CT4_0_1_pool_rep1	eDNAB0003159	PowerMax Soil DNA Isolation Kit	DNA	>1,000; 500-1,000; 250-500; 40-250; 20-40	1.4+1.2+1.5+10+10=24.1 g	0.9	2,415	42.9423	6.7426	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL07_CT4_0_1_pool_rep2	eDNAB0003159	PowerMax Soil DNA Isolation Kit	DNA	>1,000; 500-1,000; 250-500; 40-250; 20-40	1.4+1.2+1.5+10+10=24.1 g	1.2	2,415	42.9423	6.7426	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL11_CT4_0_1_pool_rep1	eDNAB0003160	PowerMax Soil DNA Isolation Kit	DNA	>1,000; 500-1,000; 250-500; 40-250; 20-40	0.7+0.4+5.4+6.9+10=23.4 g	0.3	2,418	42.9423	6.7423	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL11_CT4_0_1_pool_rep2	eDNAB0003160	PowerMax Soil DNA Isolation Kit	DNA	>1,000; 500-1,000; 250-500; 40-250; 20-40	0.7+0.4+5.4+6.9+10=23.4 g	0.3	2,418	42.9423	6.7423	Continental slope	Gulf of Lyon, Western Mediterranean
EssNaut_DNA_extraction blank	eDNAB0003157	PowerMax Soil DNA Isolation Kit	DNA	NA	NA	0.2	NA	NA	NA	Continental slope	Gulf of Lyon, Western Mediterranean
Water_extraction_blank	eDNAB0003316	Tara Oceans extraction protocol	DNA	NA	NA	0.2	NA	NA	NA	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL11_Salsa3Bol2_20	eDNAB0000222	Tara Oceans extraction protocol	DNA	> 20	6,300 L	3.02	2,417	42.9425	6.7440	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL11_Salsa3Bol4_20	eDNAB0000224	Tara Oceans extraction protocol	DNA	> 20	6,300 L	3.22	2,417	42.9425	6.7440	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL7_Salsa1Bol3_20	eDNAB0000149	Tara Oceans extraction protocol	DNA	> 20	4,740 L	6.96	2,417	42.9423	6.7426	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL7_Salsa1Bol5_20	eDNAB0000150	Tara Oceans extraction protocol	DNA	> 20	4,740 L	5	2,417	42.9423	6.7426	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL9_Salsa2Bol2_20	eDNAB0000155	Tara Oceans extraction protocol	DNA	> 20	5,400 L	16.3	1,152	43.2237	6.8876	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL9_Salsa2Bol4_20	eDNAB0000157	Tara Oceans extraction protocol	DNA	> 20	5,400 L	6.46	1,152	43.2237	6.8876	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL10_PBT2_0.2	eDNAB0002506	Tara Oceans extraction protocol	DNA	0.2-2.0	7.5 L	3.6	2,420	42.9425	6.7444	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL10_PBT2_2	eDNAB0002507	Tara Oceans extraction protocol	DNA	2.0-20	7.5 L	0.01	2,420	42.9425	6.7444	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL10_PBT2_20	eDNAB0002508	Tara Oceans extraction protocol	DNA	> 20	7.5 L	0.01	2,420	42.9425	6.7444	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL7_PBT2_20	eDNAB0002505	Tara Oceans extraction protocol	DNA	> 20	7.5 L	0.01	2,417	42.9423	6.7426	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL7_PBT2_2	eDNAB0002504	Tara Oceans extraction protocol	DNA	2.0-20	7.5 L	0.01	2,417	42.9423	6.7426	Continental slope	Gulf of Lyon, Western Mediterranean
ESSNAUT_PL7_PBT2_0.2	eDNAB0002503	Tara Oceans extraction protocol	DNA	0.2-2.0	7.5 L	1	2,417	42.9423	6.7426	Continental slope	Gulf of Lyon, Western Mediterranean

Table S2. Primers used in this study, targeting metazoans with the COI and 18S V1-V2 loci, unicellular eukaryotes with the 18S V4 locus, and prokaryotes with the 16S V4-V5 marker.

Locus	**Target and *specificity***	**Primer forward Primer reverse**	**Short name**	**Sequence (5'-3')**	**Amplicon size (bp)**	**Reference**
COI	Eukaryotes	mlCOIintF	COI-F	GGWACWGGWTGAACWGTWTAYCCYCC	313	Leray et al., 2013
	pref. metazoans	jgHCO2198	COI-R	TAIACYTCIGGRTGICCRAARAAYCA		
18S V1-V2	Eukaryotes	SSUF04	18S-V1F	GCTTGTCTCAAAGATTAAGCC	330-390	Sinniger et al., 2016
	pref. metazoans	SSURmod	18S-V1-R	CCTGCTGCCTTCCTTRGA		
18S V4	Eukaryotes	V4F (TAReukFWD1)	18S-V4-F	CCAGCASCYGCGGTAATTCC	350-410	Stoeck et al., 2010
	all	V4R (TAReukREV3)	18S-V4-R	ACTTTCGTTCTTGATYRA		
16S V4-V5	Prokaryotes	515f	16S-F	GTGYCAGCMGCCGCGGTAA	350-390	Parada et al., 2016
	Pref. Eubacteria	926r	16S-R	CCGYCAATTYMTTTRAGTTT		

Table S3. ABYSS metabarcoding pipeline

Process	**Software**	**Script(s) and command(s)**
Raw reads preprocessing for ligation data	Abyss-preprocessing: separate forward and reverse reads in each run, and re-pair reads	extract.sh using extractR1R2.py with cutadapt v1.18 (-e 0.14-0.17 for 18S,16S i.e. 3 nt mismatches and 0.27 for COI , -O length of primer -1) and BBMAP Repair v38.22
Read quality-filtering	Dada2 v.1.10	filterAndTrim() in dada2main.R maxEE=2, maxN=0, truncQ=11, truncLen=220 (18S, 16S) or 200 (COI)
Read error learning	Dada2 v.1.10	learnErrors() in dada2main.R nbases=1e8, multithread=TRUE, randomize=TRUE
Read dereplicating	Dada2 v.1.10	derepFastq() in dada2main.R
Read correction	Dada2 v.1.10	dada() in dada2main.R
Read merging	Dada2 v.1.10	mergePairs() in dada2main.R minOverlap=12, maxMismatch=0
Make sequence table and filter by length	Dada2 v.1.10	makeSequenceTable() in dada2main.R seqtab[,nchar(colnames(seqtab)) %in% seq(lengthMin,lengthMax)] lengthMin= 330 (18S-V1), 300 (COI), 350 (18S-V4), 350 (16S) lengthMax= 390 (18S-V1), 326 (COI), 410 (18S-V4), 390 (16S)
Chimera removal	Dada2 v.1.10	removeBimeraDenovo() in dada2main.R
Taxonomic assignment with RDP Classifier	Dada2 v.1.10	assignTaxonomy() in dada2outputfiles.R minBoot=50, outputBootstraps=TRUE
Taxonomic assignment of ASVs with BLAST+	blastn (megablast) v.2.6.0	blast.pbs -outfmt 11 -qcov_hsp_perc 80 -perc_identity 70 -max_hsps 1, -evalue 1e-5, then merge BLAST and RDP taxonomies using concat_blast_rdp_tax.pbs
Clustering of ASVs, chimera removal, taxonomic assignment (optional)	FROGS v.2.0.0	frogs.pbs using clustering.py with d=3 for 18S V1-V2 and d=6 for COI, then remove_chimera.py, and affiliation_OTU_identite_couverture.py
Blank correction	Rscript	Data_refining.Rmd using packages decontam v.1.2.1 and phyloseq v.1.26.0
Removal of unassigned and non-target clusters		
Deletion of defective samples (< 10,000 target reads)		
LULU curation	LULU v.0.1	lulu.R using minimum_ratio_type = "min", minimum_ratio = 1, minimum_match = 84 for COI and 90 for 18S V1-V2, minimum_relative_cooccurence = 0.90

Table S4. Number of reads and clusters (ASVs for 18S V4 and 16S, OTUs for 18S V1-V2 and COI) obtained at different analysis steps, depending on sample processing category. Data refining was performed in R, based on BLAST assignments obtained using the Silva v132 database for 18S V1-V2 and 16S loci, on the PR2 database for 18S V4, and on the MIDORI marine-only database for COI. Final number of target reads represent the number of target-taxa reads after data refining (decontamination, removal of unassigned and unknown clusters), additional LULU curation for metazoans, and removal of all clusters with less than 86% BLAST hit identity for rDNA loci and 80% for COI. Final number of target clusters are the corresponding ASVs for 18S V4 and 16S, and the corresponding OTUs for 18S V1-V2 and COI.

Sample type	Number of samples	Raw reads	Quality-filtered reads	Merged reads	Length-filtered reads	Non chimeric reads	% reads retained	Number of raw clusters	Total raw clusters	Final number of target reads	Final number of target clusters
LOCUS											
COI											
Sampling box	4	5,633,675	4,972,882	4,924,011	4,882,680	4,837,720	86	1,103	10,350	5,109,839	507
in situ pump	6	5,904,107	5,560,956	5,457,180	5,038,842	4,770,288	81	2,709			
Not sieved sediment	6	5,139,091	4,395,309	4,327,685	3,877,338	3,855,547	75	6,114			
Sieved sediment	6	5,210,192	4,384,663	4,332,251	3,702,371	3,671,540	70	4,888			
PCR Control Sample	6	465,237	262,309	255,630	254,827	254,817	55	78		na	na
Extraction Control Sample	2	3,233,904	2,475,388	2,431,972	2,426,469	2,417,278	75	29		na	na
18S V1-V2											
Sampling box	2	2,079,109	1,850,870	1,817,310	1,759,578	1,692,874	81	853	17,009	5,821,448.00	405
in situ pump	6	5,357,877	4,709,302	4,367,841	4,238,624	3,976,354	74	4,550			
Not sieved sediment	6	3,890,833	2,665,878	2,463,769	2,050,489	2,006,796	52	6,275			
Sieved sediment	5	4,719,584	3,621,503	3,329,633	3,077,489	3,031,583	64	6,497			
PCR Control Sample	4	1,976,483	1,123,546	1,080,187	914,275	545,311	28	194		na	na
Extraction Control Sample	2	1,434,680	1,143,161	1,082,658	729,709	710,436	50	145		na	na
18S V4											
Sampling box	4	4,142,505	3,691,442	3,647,381	3,645,131	3,597,953	87	5,176	35,538	4,549,661	7,081
in situ pump	6	6,055,792	5,355,958	5,204,132	5,203,747	4,883,139	81	6,905			
Not sieved sediment	6	1,190,150	855,544	742,230	731,661	728,993	61	5,572			
Sieved sediment	5	1,655,300	1,219,956	1,030,018	1,020,309	1,016,622	61	5,380			
PCR Control Sample	4	5,153	1,014	867	867	867	17	21		na	na
Extraction Control Sample	2	1,097,221	848,500	838,436	837,909	825,498	75	174		na	na
16S V4-V5											
Sampling box	4	3,434,755	2,952,132	2,623,265	2,606,614	2,517,523	73	7,363	62,646	6,982,401	38,816
in situ pump	6	3,987,382	3,148,002	1,219,715	1,169,737	1,154,763	29	16,567			
Not sieved sediment	6	3,544,156	2,431,818	2,095,989	2,083,590	1,995,875	56	39,942			
Sieved sediment	6	2,829,692	1,952,664	1,612,822	1,608,064	1,540,155	54	38,280			
PCR Control Sample	4	1,516,645	1,327,867	1,324,078	1,320,742	1,320,640	87	125		na	na
Extraction Control Sample	2	2,095,233	1,677,554	1,428,751	1,398,492	1,397,637	67	268		na	na

Supplemental figures

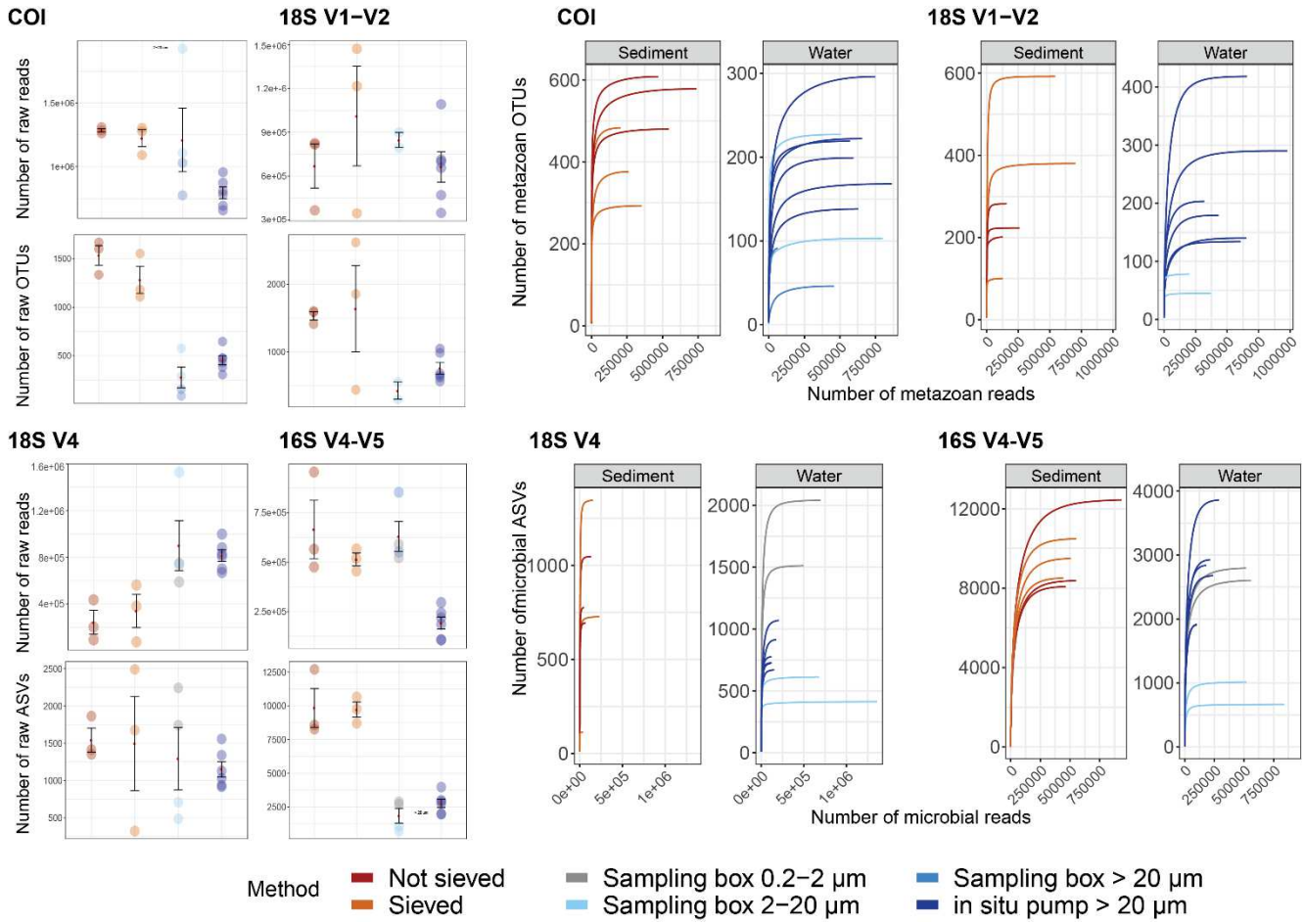


Figure S1. Raw read and cluster numbers in deep-sea sediment (brown) and aboveground water (blue) with different sampling methods in metabarcoding inventories of metazoans (COI, 18S V1-V2), micro-eukaryotes (18S V4), and prokaryotes (16S V4-V5). Sediment was either sieved through 5 mesh sizes to size-sort organisms prior DNA extraction, or DNA was extracted directly from crude sediment samples. Water was collected with a 7.5 L sampling box, allowing recovery of two size classes, or sampled in large volumes with an *in situ* pump. Rarefaction curves were performed on refined datasets and show a plateau is reached in most samples, except sediment samples with 18S V4.

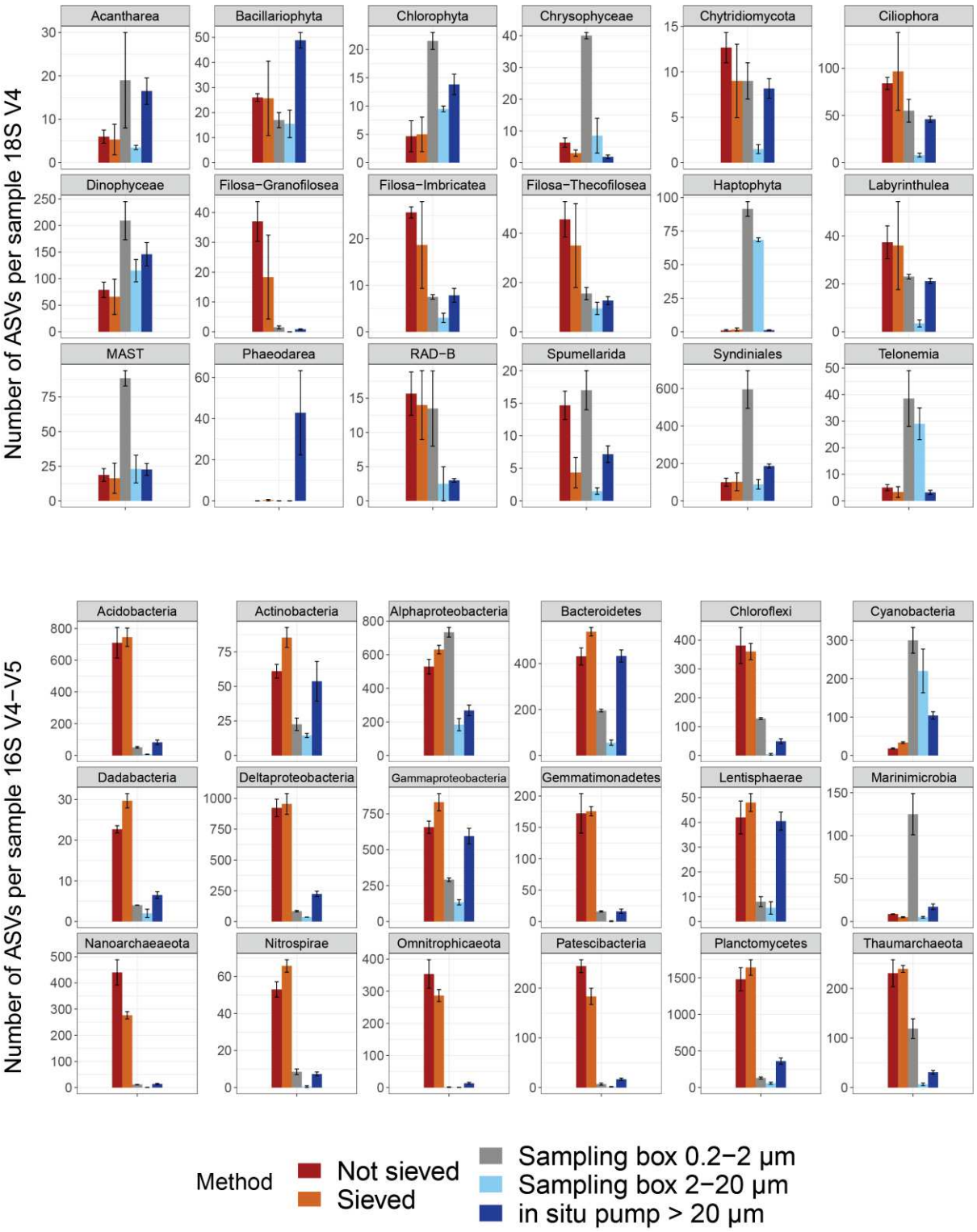


Figure S2. Mean numbers (±SE) of protist (18S V4) and prokaryote (16S V4-V5) Amplicon Sequence Variants (ASVs) in major taxonomic lineages, recovered by deep-sea sediment (brown) and aboveground water (blue), using two sampling methods for both sample types. Sediment was either sieved to size-sort organisms prior DNA extraction, or DNA was extracted directly from crude sediment samples. Water was collected with a 7.5 L sampling box, allowing recovery of two size classes, or sampled in large volumes with an *in situ* pump.

ASV numbers were calculated on the rarefied datasets.

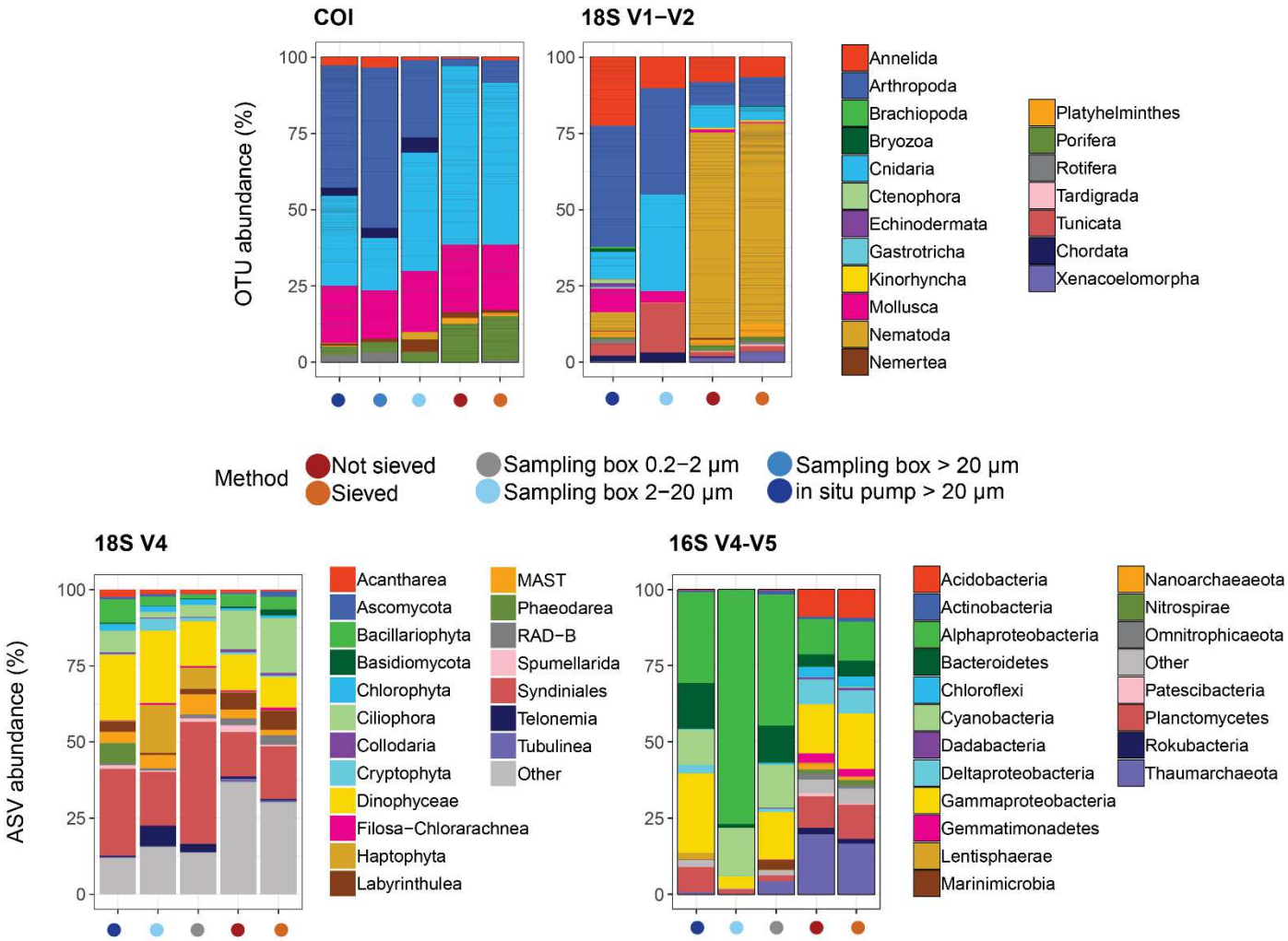


Figure S3. Patterns of relative cluster abundance resolved by eDNA metabarcoding of deep-sea sediment (brown) and aboveground water (blue), using two sampling methods for both sample types, and using four barcode markers targeting metazoans (COI, 18S V1-V2), micro-eukaryotes (18S V4), and prokaryotes (16S V4-V5). Sediment was either sieved to size-sort organisms prior DNA extraction, or DNA was extracted directly from crude sediment samples. Water was collected with a 7.5 L sampling box, allowing recovery of up to two size classes per taxonomic compartment, or sampled in large volumes with an *in situ* pump. Top 20 most abundant taxa are displayed for microbial groups.

Supplementary material Chapter V.

Supplemental tables

Table S1. Sampling sites, their GPS locations and associated habitats

Station	Cruise	Depth (m)	Latitude	Longitude	Habitat	**Location**	Region
PCT-TYRR	PEACETIME	3,400	39.3402	12.5834	Abyssal plain	Tyrrhenian Sea	Western Mediterranean
ESN-2400m	EssNaut	2,400	42.9423	6.7423	Abyssal plain	Gulf of Lyon	Western Mediterranean
ESN-330m	EssNaut	334	43.0867	6.4512	Continental slope	Gulf of Lyon	Western Mediterranean
CHR	CANHROV	2,490	42.7167	6.1333	Marine canyon	Gulf of Lyon	Western Mediterranean
PCT-FA	PEACETIME	2,800	37.9467	2.9167	Abyssal plain	South Balearic islands	Western Mediterranean
MDW-ST179	MEDWAVES	729	36.4808	-2.8945	Seamount	Alboran sea	Gibraltar Strait East
MDW-ST201	MEDWAVES	381	36.546	-2.8135	Seamount	Alboran sea	Gibraltar Strait East
MDW-ST215	MEDWAVES	554	36.5157	-2.7942	Seamount	Alboran sea	Gibraltar Strait East
MDW-ST22	MEDWAVES	470	36.5598	-6.9492	Mud volcano	Gulf of Cadiz	Gibraltar Strait West
MDW-ST23	MEDWAVES	470	36.5605	-6.9498	Mud volcano	Gulf of Cadiz	Gibraltar Strait West
MDW-ST38	MEDWAVES	1,920	36.8442	-11.3025	Seamount	Soutwest Portugal	North Atlantic
MDW-ST68	MEDWAVES	1,245	37.2837	-24.7873	Seamount	Mid-Atlantic ridge	North Atlantic
MDW-ST117	MEDWAVES	1,325	37.34	-24.7552	Seamount	Mid-Atlantic ridge	North Atlantic

Table S2. Taxonomic and relative composition (% DNA input) of the deep-sea metazoan mock communities used in this study.

Taxonomic group	Species	Mock 3 (%)	Mock 5 (%)
Polychaeta; Eunicida	*Eunice norvegica*	40	80
Crustacea; Malacostraca	*Chorocaris* sp. (now *Rimicaris* sp. or *M. fortunata*)	3	0.7
Crustacea; Malacostraca	*Alvinocaris muricola*	3	0.7
Crustacea; Malacostraca	*Munidopsis* sp.	3	0.7
Anthozoa;Alcyonacea	*Acanella arbuscula*	20	10
Anthozoa; Scleractinia;Caryophylliidae	*Desmophyllum dianthus*	3	0.7
Bivalvia;Veneroida; Vesicomyidae	*Calyptogena pacifica*	3	0.7
Bivalvia;Veneroida; Vesicomyidae	*Christineconcha regab* (formerly *Calyptogena* sp.)	3	0.7
Bivalvia;Veneroida; Vesicomyidae	*Vesicomya gigas*	3	0.7
Gastropoda; Patellogastropoda	*Paralepetopsis* sp.	20	5

Table S 3. ABYSS metabarcoding pipeline.

Process	**Software**	**Script(s) and command(s)**
Raw reads preprocessing for ligation data	Abyss-preprocessing: separate forward and reverse reads in each run, and re-pair reads	extractR1R2.pbs using cutadapt v1.18 (-e 0.17 for 18S/16S and 0.27 for COI , -O length of primer -1) and BBMAP Repair v38.22
Read quality-filtering	Dada2 v.1.10	filterAndTrim() in dada2main.R maxEE=2, maxN=0, truncQ=11, truncLen=220 (18S, 16S) or 200 (COI)
Read error learning	Dada2 v.1.10	learnErrors() in dada2main.R nbases=1e8, multithread=TRUE, randomize=TRUE
Read dereplicating	Dada2 v.1.10	derepFastq() in dada2main.R
Read correction	Dada2 v.1.10	dada() in dada2main.R
Read merging	Dada2 v.1.10	mergePairs() in dada2main.R minOverlap=12, maxMismatch=0
Make sequence table and filter by length	Dada2 v.1.10	makeSequenceTable() in dada2main.R seqtab[,nchar(colnames(seqtab)) %in% seq(lengthMin,lengthMax)] lengthMin= 330 (18S-V1), 300 (COI), 250 (18S-V4), 350 (16S) lengthMax= 390 (18S-V1), 326 (COI), 450 (18S-V4), 390 (16S)
Chimera removal	Dada2 v.1.10	removeBimeraDenovo() in dada2main.R
Taxonomic assignment with RDP Classifier	Dada2 v.1.10	assignTaxonomy () in dada2outputfiles.R minBoot=50, outputBootstraps=TRUE
Taxonomic assignment with BLAST+	blastn (megablast) v.2.6.0	blast.pbs -outfmt 11 -qcov_hsp_perc 80 -perc_identity 70 -max_hsps 1, -evalue 1e-5, then merge BLAST and RDP taxonomies using concat_blast_rdp_tax.pbs
Clustering (optional)	FROGS v.2.0.0	clustering.py with d=3 for 18S V1-V2 and d=6 for COI, remova_chimera.py, affiliation_OTU_identite_couverture.py
Blank correction	Rscript	Data_refining.Rmd using packages decontam v.1.2.1 and phyloseq v.1.26.0
Removal of unassigned and non-target clusters		
Deletion of defective samples (< 10,000 target reads)		
Tag-switching renormalisation	Rscript	owi_renormalize.R
LULU curation	LULU v.0.1	lulu() in lulu_final.R minimum_ratio_type = "min", minimum_ratio = 1, minimum_match =84 (COI) or 90 (18S-V1), minimum_relative_cooccurence =0.90

Table S4. Read-track table. Number of reads obtained in samples after each processing step. Data refining was performed in R, based on BLAST assignments obtained using the Silva v132 database for 18S, and on the MIDORI-UNIQUE database subseted to marine taxa only for COI. Numbers in parentheses indicate final numbers after reducing to target taxa.

Sample type	Number of samples	Raw clusters	Raw reads	Quality-filtered reads	Merged reads	Length-filtered reads	Non chimeric reads	% reads retained	Number of samples after refining	Target reads after all refining steps	Target clusters after all refining steps
LOCUS											
18S V1-V2											
Field/Extraction Control Sample	12	58,912	4,353,270	2,783,839	2,751,809	2,555,319	2,472,848	57	0	31,014,318 (8,387,537)	3,992 (3,008)
Mock Sample (Metazoa only)	2		2,096,631	1,607,219	1,437,248	1,431,286	1,290,982	62	2		
PCR Control Sample	18		8,757,880	1,178,326	1,128,169	941,596	574,595	7	0		
True Sample	133		147,945,682	104,073,185	95,115,017	81,052,159	78,648,448	53	112		
COI											
Field/Extraction Control Sample	12	65,544	3,219,092	2,267,010	2,244,694	2,238,420	2,231,017	69	0	21,383,401 (21,261,985)	11,938 (11,808)
Mock Sample (Metazoa only)	2		1,482,785	1,261,045	1,252,953	1,252,039	1,224,751	83	2		
PCR Control Sample	19		1,694,231	551,913	500,162	491,479	491,396	29	0		
True Sample	82		83,062,120	69,457,315	67,038,612	57,589,525	57,219,446	69	81		

Supplemental figures

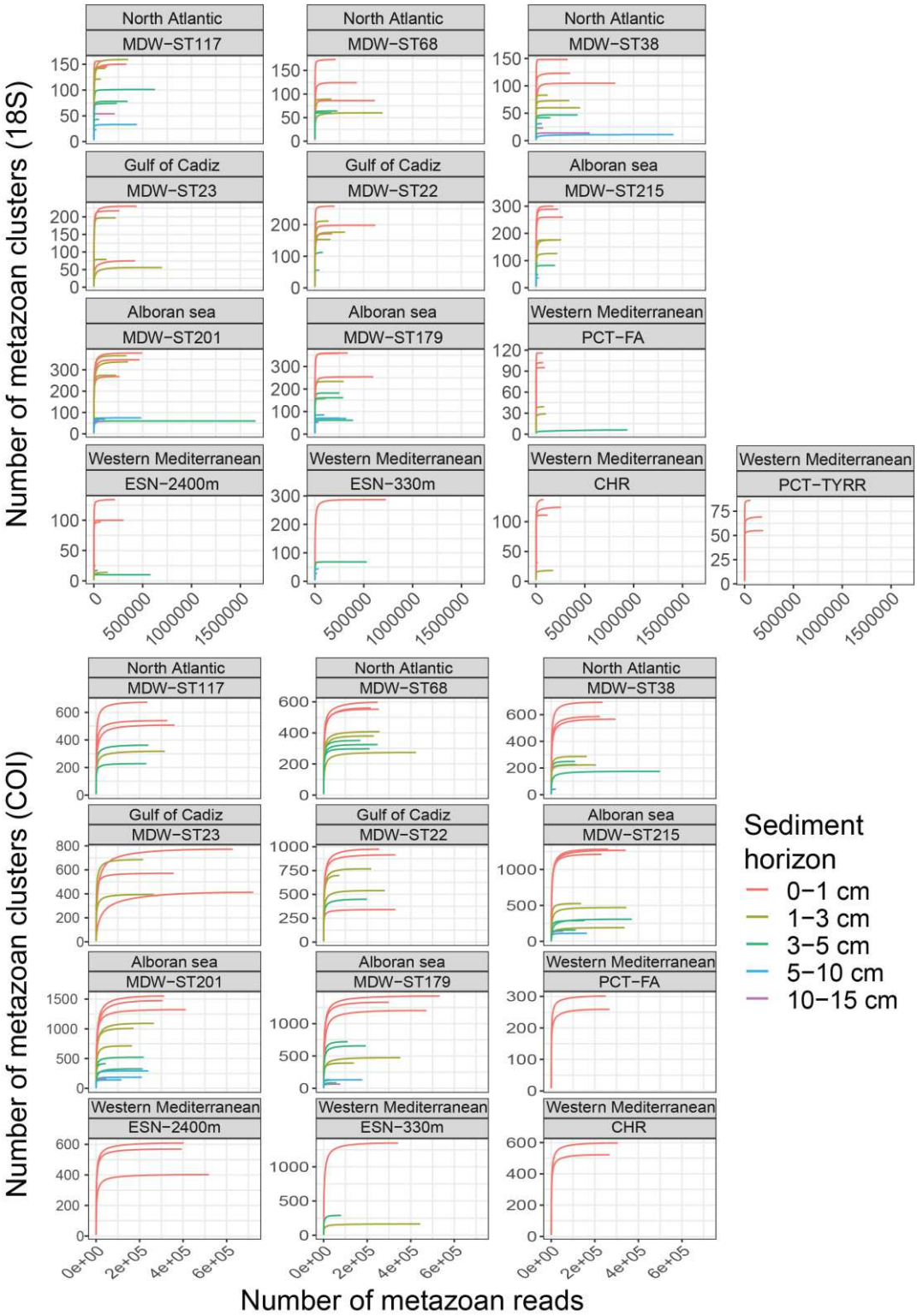


Figure S1. Rarefaction curves of metazoan biodiversity inventories in sediment horizons from 13 deep-sea sites covering the Atlantic-Mediterranean transition zone. Inventories were produced by metabarcoding with the 18S V1-V2 and COI barcode markers. A plateau is reached in most samples, except in some 0-1 cm horizons with 18S, suggesting not all diversity was revealed in these samples.

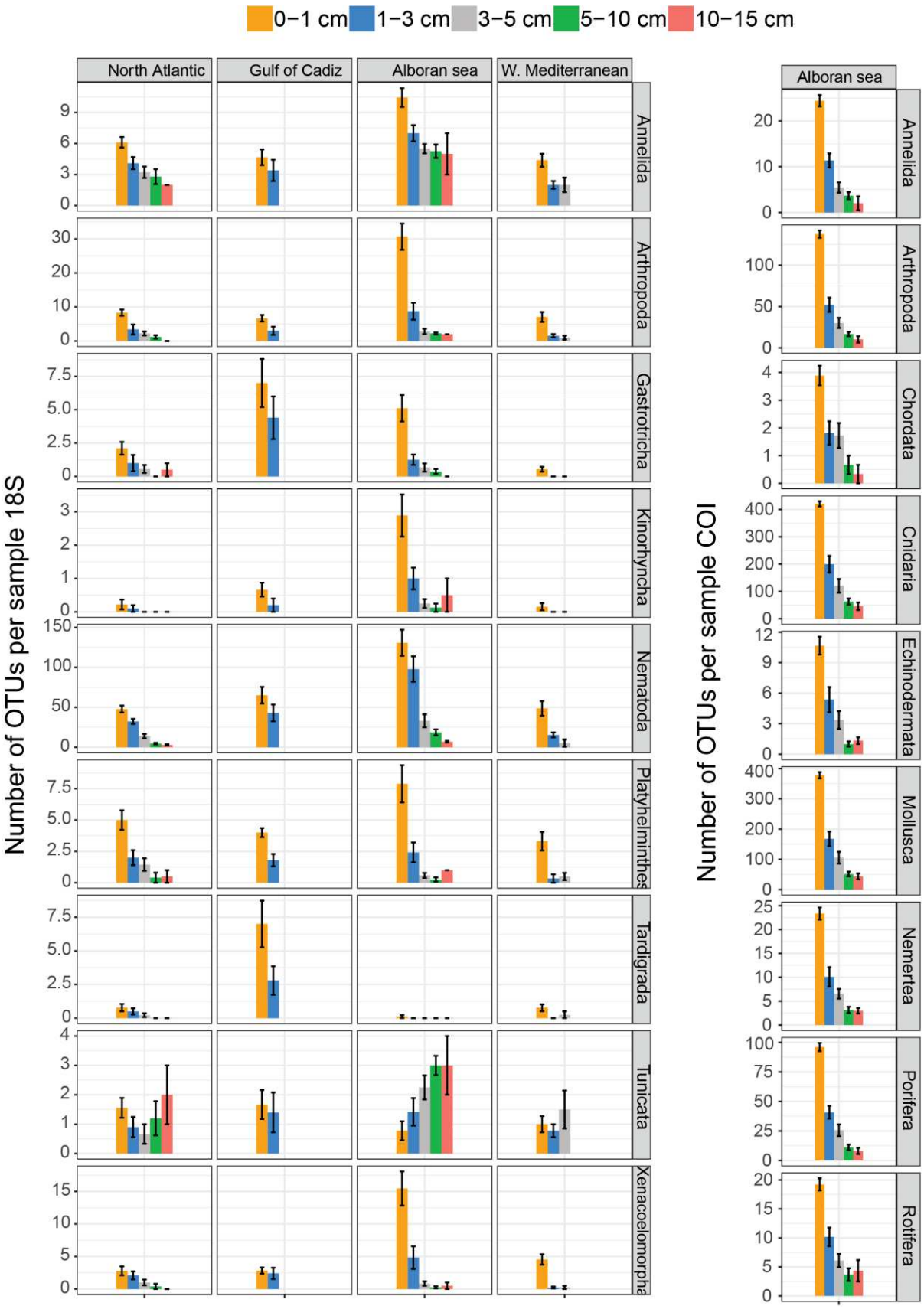


Figure S2. Mean number of metazoan OTUs for selected phyla detected by 18S V1-V2 and COI, in sediment horizons of 13 deep-sea sites from four regions covering the Atlantic-Mediterranean transition zone. Only one region is shown for COI as the others had low success in deeper horizons. Cluster numbers were calculated on rarefied datasets. Error bars represent standard errors.

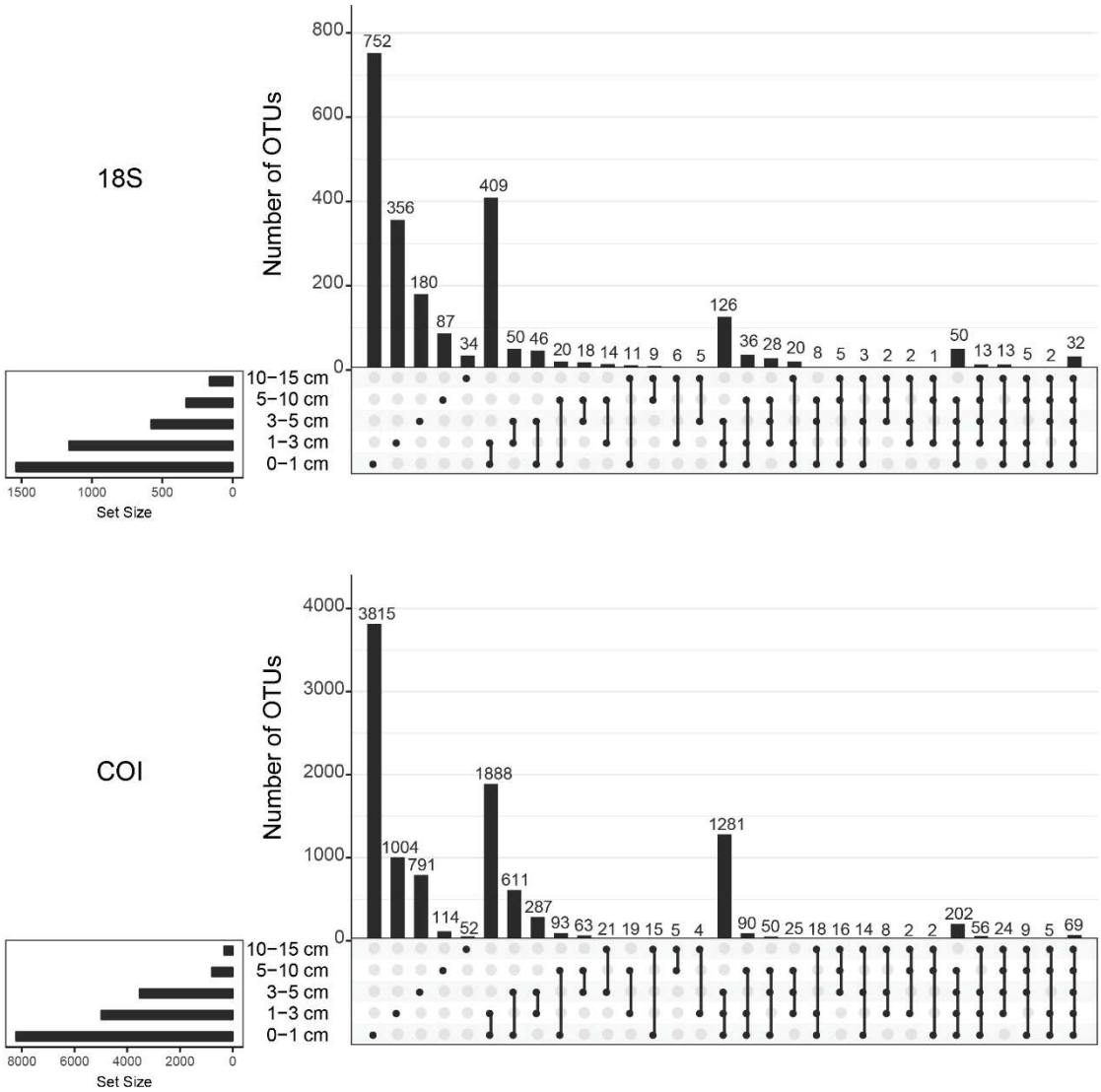


Figure S3.Shared and unique metazoan OTUs (18S V1-V2, COI) among sediment horizons of 13 deep-sea sites across the Atlantic-Mediterranean transition zone. Numbers were calculated on rarefied datasets.

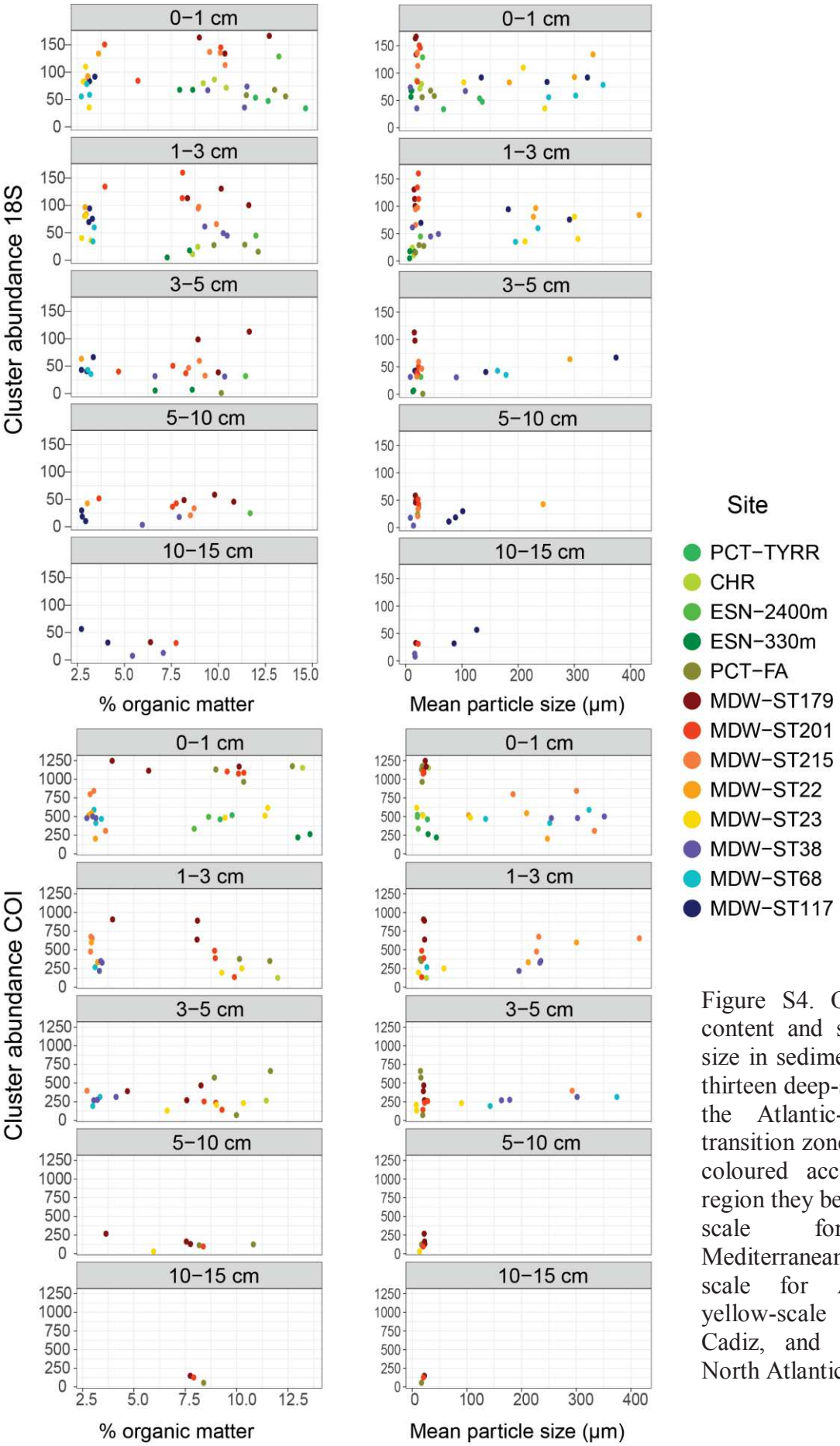
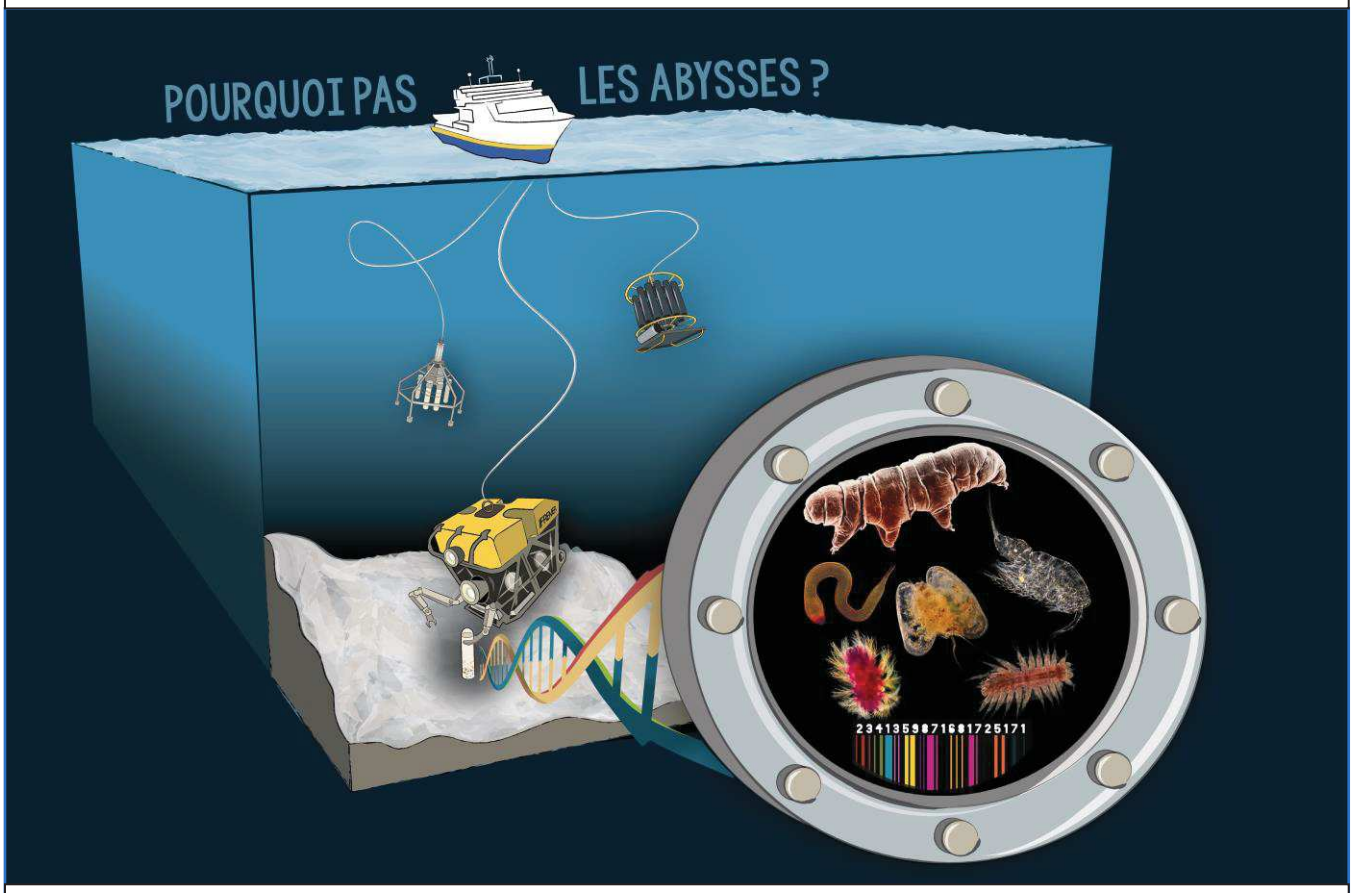


Figure S4. Organic matter content and sediment grain size in sediment horizons of thirteen deep-sea sites across the Atlantic-Mediterranean transition zone. The sites are coloured according to the region they belong to: green-scale for Western Mediterranean sites, red-scale for Alboran Sea, yellow-scale for Gulf of Cadiz, and blue-scale for North Atlantic.

The abyssal seafloor covers more than half of planet Earth. It can host a large number of, mostly small and still undescribed, organisms (~50,000-5 million individuals per square meter), contributing to key ecosystem functions such as nutrient cycling, sediment stabilisation and transport, or secondary production. Technological developments in the past 30 years have allowed remarkable advances, yet due to the vastness and remoteness of deep-sea habitats, ecological studies have been limited to local and regional scales. Indeed, we have so far explored less than 1% of the deep seafloor, and this contrasts with the fact that deep-sea ecosystems form one of the largest biomes on Earth, and are under increased threat from a variety of direct and indirect anthropogenic pressures. This PhD aims at bringing new perspectives for the study of biodiversity and biogeography in the deep-sea, to bridge this large knowledge gap, and advance toward the development of efficient biomonitoring protocols to preserve this vast and elusive backyard. We investigated the potential of multi-marker environmental DNA (eDNA) metabarcoding to assess the extent and distribution patterns of biodiversity in this remote ecosystem. Using mitochondrial and nuclear marker genes, this PhD aimed at producing and testing an optimized eDNA metabarcoding workflow for deep-sea sediments, on a bioinformatic, molecular, and sample processing level, applicable to multiple life compartments including microbiota and metazoans.



Biodiversity assessment with eDNA is confronted with the difficulty in defining accurate “species-level” Operational Taxonomic Units (OTUs), as numerous sources of error induce frequent overestimations. The first part of this thesis describes how newly developed bioinformatic tools can be combined in order to get more conservative and reliable biodiversity inventories, approaching a 1:1 species-OTU correspondence, and underline the advantages of clustering and LULU-curation for producing more reliable metazoan biodiversity inventories. Moreover, the accuracy of protocols based on eDNA in deep sea sediments still needs to be assessed, as results may be biased by ancient DNA, resulting in biodiversity assessments not targeting *live* organisms.This thesis assessed the potential bias of ancient DNA by 1) evaluating of the effect of removing short DNA fragments, and 2) comparing communities revealed by co-extracted DNA and RNA in five deep-sea sites. Results indicated that short extracellular DNA fragments do not affect alpha and beta diversity, but that DNA obtained from 10g of sediment should be favoured over RNA for logistically realistic, repeatable, and reliable surveys. Results also confirm show that increasing the number of biological rather than technical replicates is important to infer robust ecological patterns. Sieving sediment to separate benthic size classes increased the number of detected metazoan OTUs, but was not essential for achieving comprehensive and accurate biodiversity estimates, and should be avoided if unicellular taxonomic compartments are also of interest. Finally, this thesis applied the optimized eDNA metabarcoding protocols to investigate the influence of biotic and abiotic factors on the extent and distribution of deep-sea metazoan biodiversity on an East-West transect ranging from the Central Mediterranean to the Mid-Atlantic Ridge. Results, consistent to morphology-based studies, confirm that small-scale biotic and abiotic factors lead to significant vertical changes in metazoan richness and community structure within the sediment, and highlight that regional beta-diversity patterns result from a combined influence of past biogeography and present day processes. This thesis opens the way to large-scale eDNA-based studies in the deep-sea realm, thus contributing to a better understanding of biodiversity, biogeography, and ecosystem function in this vast and still poorly known biome.