Marine biodiversity from past to recent: insights from sedimentary DNA

Barrenechea Angeles, Inès

UNIVERSITÉ DE GENÈVE
Département des Sciences de la Terre
Polish Academy of Sciences

FACULTÉ DES SCIENCES
Directeur Prof. Daniel Ariztegui
Co-director Prof. Jan Pawlowski

Marine biodiversity from past to recent: insights from sedimentary DNA

THÈSE

Présentée à la Faculté des sciences de l'Université de Genève

Pour obtenir le grade de Docteur ès Sciences, mention sciences de la Terre

par

Inés Barrenechea Angeles

de

Genève (GE)

Thèse N° 5710

GENÈVE

Atelier de reprographie ReproMail

202

UNIVERSITÉ
DE GENÈVE

FACULTÉ DES SCIENCES

DOCTORAT ÈS SCIENCES, MENTION SCIENCES DE LA TERRE

Thèse de Madame Inès BARRENECHEA ANGELES

intitulée :

«Marine Biodiversity from Past to Recent : Insights from Sedimentary DNA»

La Faculté des sciences, sur le préavis de Monsieur D. ARIZTEGUI, professeur honoraire et directeur de thèse (Département des sciences de la Terre), Monsieur J. PAWLOWSKI, professeur associé et codirecteur de thèse (Department of Marine Ecology, Institute of Oceanology Polish Academy of Sciences, Sopot, Poland), Madame S. ABRAMOVICH, professeure (Department of Earth Sciences and Environment, Ben Gurion University of the Negev, Beer-Sheva, Israel), Madame S. ARNAUD-HAOND, docteure (Laboratoire Evolution et Génétique des Populations, Stations de Sète, Sète, France) et Madame I. DOMAIZON, docteure (UMR CARRTEL - INRAE & USMB, Thonons-les-Bains, France), autorise l'impression de la présente thèse, sans exprimer d'opinion sur les propositions qui y sont énoncées.

Genève, le 20 décembre 2022

Thèse - 5710 -

Le Doyen

N.B. - La thèse doit porter la déclaration précédente et remplir les conditions énumérées dans les "Informations relatives aux thèses de doctorat à l'Université de Genève".

Acknowledgments/Remerciements

Dans ces pages, je voudrais remercier ceux qui ont été présent de près ou de loin durant ces années de thèse, mais aussi à toutes les personnes qui m'ont fait évoluer personnellement et scientifiquement.

Tout d'abord, un immense merci à mes superviseurs Jan Pawlowski et Daniel Ariztegui, qui malgré leur passage à la retraite m'ont soutenue et été toujours disponibles.

Merci Jan, pour m'avoir donné l'opportunité d'abord en 2013 de faire un stage dans ton laboratoire alors que j'étais en 2è année de géologie et que mes connaissances en foraminifères provenaient essentiellement des cours de micropaléontologie. Merci d'avoir intégré dans un projet de master touchant aussi bien des sédiments récents qu'anciens et les foraminifères, de faire le possible pour que je commence et continue ma thèse après ta retraite. Durant ces années de thèse, merci de la confiance accordée et laissée encadrer toutes les personnes qui ont fait un séjour de recherche dans le laboratoire et ainsi avoir pu participer dans divers projets. Merci de m'avoir donné l'opportunité de partir deux mois sur l'expédition au Pacifique, Abyss2020, et de cette manière avoir pu voir ce qu'est le travail de terrain : la récolte d'échantillons et le tri des foraminifères sous un microscope avec le tangage du bateau.

Merci Daniel de m'avoir soutenue déjà pendant le master et de m'avoir guidée malgré mes sujets de recherche loin de tiennes auxquelles tu trouvais quand même des similitudes : la mer/l'océan est comme un énorme lac, que les processus sédimentaires sont presque les mêmes. Merci pour avoir gardé un intérêt particulier pour mes divers projets marins.

I am very grateful to Dr. Sigal Abramovich, Dr. Sophie Arnaud-Haond and Dr. Isabelle Domaizon, for accepting to review my thesis.

Merci Maria Holzmann, pour l'initiation au barcoding et séquençage des foraminifères. Merci d'avoir toujours ton bureau ouvert pour moi et de partager ton parcours, tes connaissances et surtout ta passion pour les foraminifères.

Merci à toutes les personnes du laboratoire que j'ai côtoyées. Je pense particulièrement à Franck Lejzerowicz pour m'avoir guidée pendant le master et aidée à combler mes lacunes en biologie moléculaire et bioinformatique. Tristan Cordier, pour toutes nos discussions sur mes papiers, R et stats. Laure Apothéloz Perret Gentil pour tes conseils au laboratoire et les après-midis passées aux TPs. Kristina Cermakova, merci pour ta compagnie au labo et les discussions sur les manips, PCRs ou des montagnes. Manu Reo pour avoir lancé mes runs Illumina et fait en sorte que je puisse avoir toujours tout le matériel nécessaire. Yoann Dufresne, Florian

Mauffrey et Slim Chraiti et aux étudiants et chercheurs venus au laboratoire Léo, Jade, Kristel Panksep, Lorena, Piotr et Océanne.
Raphael Morad pour m'avoir accueillie au MARUM et initiée aux foraminifères planctoniques et bien sûr pour m'avoir aidée avec mon premier papier.
Du côté des Sciences de la Terre, merci à Rossana Martini, à Elias Samankassou, à Camille Thomas, à Sandrine Le Houedec, Tamara, Marine, Aurélia, Jeanne. D'autres doctorants rencontrés lors des workshops et terrain de la CUSO, B. Lauper, A. Lehmann, C. Ferrante, J. Sturny.
Merci Eric Capo et Kevin Nota de m'avoir accueillie bras ouverts à Université d'Uppsala et de m'avoir donnée l'opportunité d'apprendre un peu plus sur la métagénomique de sédiments anciens.
Je remercie également la Fondation Augustin Lombard pour le financement de mon séjour de à Université d'Uppsala.
Merci à Koh Siang et aux membres de son laboratoire de l'institut Tropical Marine Science Institute (TMSI) de l'Université de Singapore qui m'ont « recrutée » pour aller échantillonner dans l'océan Pacifique.
Finalement, je tiens à remercier à mon entourage proche, mes amis. Et surtout à ma famille, à ma maman Irma, à Hung, à mes grands-parents, à Vishal et à la famille Joneja qui m'ont également encouragée et soutenue.

Abstract

Seafloor sediments are inhabited by benthic organisms and accumulate the remains of pelagic organisms that falls down from the water column. The sediments are therefore a repository of the whole marine biodiversity comprising benthic and pelagic components. All these organisms contain DNA that will be incorporated into the sediments thus becoming sedimentary DNA (sedDNA). This DNA can persist over time as ancient sedimentary DNA (sedaDNA). By sequencing this sedimentary DNA, it is not only possible to access the current biodiversity but also to infer biodiversity of the past.

In this thesis I present several case studies that use sedimentary DNA to identify marine biodiversity and to analyse its variation in time and space. Primary focus of my studies are foraminifera, single-celled marine organisms, present both on the ocean surface and on the seabed. Being sensitive to changes in their environment, they are important ecological markers.

This thesis is divided into three parts, each containing two case studies.

The first part focuses on the preservation of ancient sedimentary DNA in the marine environment. In the first study (Chapter 3), the DNA preservation of benthic and planktonic foraminifera is compared across subsurface sediments. We found that benthic DNA dominates over partially degraded planktonic DNA in the first ten cm. Below ten cm, the genetic signal of both groups stabilizes and planktonic foraminifera DNA exceeds benthic one. Moreover, this signal reflects the local biogeography of planktonic species, making possible to trace the cold and warm marine currents and their confluence point.

In the second study, a sediment core spanning the period from Recent to 384 Ka was examined (Chapter 4). This core collected from the depth of 1893 m at the tropical Bismarck Sea was stored for fourteen years at 4°C. The poor storage conditions strongly impacted DNA preservation and caused contamination by modern amoebae and fungi. Nevertheless, by using specific markers we were able to recover a reliable signal of foraminifera, radiolarian and terrestrial plants DNA down to the 200 Ka. Our preliminary results comparing protist and floristic diversity across glacial and interglacial intervals show a multitaxon evidence of a depletion of terrestrial and marine diversity in the tropics since the last glacial interval (MIS 2).

The second part shows how sedimentary DNA is used to assess deep-sea benthic biodiversity at a large scale with samples from all oceans (Chapter 5) and at a local scale, focused on the Clarion-Clipperton zone in the Tropical East Pacific (Chapter 6).

The objective of the first study was to discriminate autochthonous benthic DNA from pelagic DNA deposited at the deep-sea floor. This was done by comparing the 18S V9 sequences from the bathyal and abyssal areas to the sequences obtained from the photic and aphotic areas. Thus, it was determined that about 21% of the sequences found in the deep-sea sediments are of pelagic origin. The identity of two thirds of the remaining sequences is still unknown, underlining the great diversity of the deep-sea benthos and gaps in its knowledge. Moreover, it was also possible to evaluate the flux of particulate organic carbon and to identify previously unknown taxa that could be at the origin of the biological carbon pump.

The second study in this part deals with unknown sequences, in this case those of deep-sea foraminifera. In general, the unknown sequences are either removed from ecological analyses or grouped as "unassigned". In this study, we propose a method to annotate these unassigned sequences based on their genetic signature. As a result, 61 new foraminiferal lineages grouped into 27 clades were found in Clarion-Clipperton sediment samples. Many of these lineages were present in other deep-sea areas, but only a few of them appeared in the coastal datasets, suggesting that deep-sea benthic fauna is highly endemic.

The last part of my thesis describes the use of sedimentary DNA to characterize the ecological quality status of a highly polluted coastal marine site, based on surface sediments (chapter 7) and core sediments covering the last 180 years (chapter 8). The site is located at the Bay of Bagnoli (near Naples), where an industrial complex, now dismantled, was established during the 20th century. We first calculated the current ecological status of the site along a distal to proximal gradient using morphological and sedimentary DNA data from the foraminiferal communities. Both the morphology and molecular approaches delineated low polluted areas from high polluted areas. However, the molecular data were found to be more consistent with the geochemical measurements of the pollution. Reconstruction of the reference conditions prior to the industrialization was the objective of the second study. The analysis of sedimentary DNA shows variations in prokaryotic and eukaryotic community composition resulting from habitat degradation. Prior to the industrialization, the Bay of Bagnoli was covered by a seagrass meadow with highly diversified fauna. As a result of anthropogenic activities, the Posidonia seagrass progressively disappeared and eukaryotic community was replaced by dinoflagellates

and infaunal metazoan species. Sedimentary ancient DNA allows us to follow the past changes linked to human activities and to recover the reference conditions in pre-industrial times.

Finally, I discuss different challenges encountered when analysing ancient and recent sedimentary DNA sequences and I conclude by presenting new sequencing technologies and computational methods that could be used in the future sedimentary DNA research.

Résumé

Les sédiments des fonds marins sont habités par des organismes benthiques et accumulent les restes d'organismes pélagiques qui tombent de la colonne d'eau. Les sédiments sont donc un réservoir de l'ensemble de la biodiversité marine comprenant des composants benthiques et pélagiques. Tous ces organismes contiennent de l'ADN qui sera incorporé dans les sédiments, devenant ainsi de l'ADN sédimentaire (sedDNA). Cet ADN peut persister dans le temps sous forme d'ADN sédimentaire ancien (sedaDNA). En séquençant cet ADN sédimentaire, il est non seulement possible d'accéder à la biodiversité actuelle mais aussi de déduire la biodiversité du passé.

Dans cette thèse, je présente plusieurs études de cas qui utilisent l'ADN sédimentaire pour identifier la biodiversité marine et pour analyser sa variation dans le temps et l'espace. Mes études portent principalement sur les foraminifères, des organismes marins unicellulaires, présents à la fois à la surface des océans et dans les fonds marins. Sensibles aux changements de leur environnement, ils sont des marqueurs écologiques importants.

Cette thèse est divisée en trois parties, chacune contenant deux études de cas.

La première partie se concentre sur la préservation de l'ADN sédimentaire ancien dans l'environnement marin. Dans la première étude (Chapitre 3), la préservation de l'ADN des foraminifères benthiques et planctoniques est comparée à travers les sédiments de subsurface. Nous avons constaté que l'ADN benthique domine sur l'ADN planctonique partiellement dégradé dans les dix premiers centimètres. En dessous de dix centimètres, le signal génétique de deux groupes se stabilise et l'ADN des foraminifères planctoniques dépasse celui des benthiques. De plus, ce signal reflète la biogéographie locale des espèces planctoniques, permettant de retracer les courants marins froids et chauds et leur point de confluence.

Dans la deuxième étude, une carotte de sédiments couvrant la période allant du récent aux derniers 384 Ka a été examinée (Chapitre 4). Cette carotte collectée à une profondeur de 1893 m dans la mer tropicale de Bismarck a été stockée pendant quatorze ans à 4°C. Les mauvaises conditions de stockage ont fortement affecté l'ADN et ont entraîné une contamination par des amibes et des champignons modernes. Néanmoins, en utilisant des marqueurs spécifiques, nous avons pu récupérer un signal fiable d'ADN de foraminifères, de radiolaires et de plantes terrestres jusqu'à 200 Ka. Nos résultats préliminaires comparant la diversité des protistes et de la flore à travers les intervalles glaciaires et interglaciaires montrent une preuve

multitaxonomique d'un appauvrissement de la diversité terrestre et marine dans les tropiques depuis le dernier intervalle glaciaire (MIS 2).

La deuxième partie montre comment l'ADN sédimentaire est utilisé pour évaluer la biodiversité benthique en eaux profondes à grande échelle avec des échantillons provenant de tous les océans (Chapitre 5) et à une échelle locale, centrée sur la zone de Clarion-Clipperton dans le Pacifique Est tropical (Chapitre 6).

L'objectif de la première étude était de distinguer l'ADN benthique autochtone de l'ADN pélagique déposé au fond des océans. Pour ce faire, les séquences 18S V9 des zones bathyales et abyssales ont été comparées aux séquences obtenues dans les zones photiques et aphotiques. Ainsi, il a été déterminé qu'environ 21% des séquences trouvées dans les sédiments des grands fonds sont d'origine pélagique. L'identité de deux tiers de séquences restantes est encore inconnue, soulignant la grande diversité du benthos des grands fonds et les lacunes dans sa connaissance. Par ailleurs, il a été possible d'évaluer le flux de carbone organique particulaire et d'identifier des taxons jusqu'alors inconnus qui pourraient être à l'origine de la pompe à carbone biologique.

La deuxième étude de cette partie traite des séquences inconnues, en l'occurrence celles des foraminifères d'eau profonde. En général, les séquences inconnues sont soit retirées des analyses écologiques, soit regroupées comme "non attribuées". Dans cette étude, nous proposons une méthode pour annoter ces séquences non attribuées sur la base de leur signature génétique. En conséquence, 61 nouvelles lignées de foraminifères regroupées en 27 clades ont été trouvées dans les échantillons de sédiments de Clarion-Clipperton. Beaucoup de ces lignées étaient présentes dans d'autres zones d'eaux profondes, mais seules quelques-unes d'entre elles sont apparues dans les ensembles de données côtières, ce qui suggère que la faune benthique des eaux profondes est hautement endémique.

La dernière partie de ma thèse décrit l'utilisation de l'ADN sédimentaire pour caractériser l'état de qualité écologique d'un site marin côtier fortement pollué, à partir de sédiments de surface (Chapitre 7) et de sédiments d’une carotte couvrant les 180 dernières années (Chapitre 8). Le site est situé dans la baie de Bagnoli (près de Naples), où un complexe industriel, aujourd'hui démantelé, a été établi au cours du 20ème siècle. Nous avons d'abord calculé le statut écologique actuel du site le long d'un gradient distal à proximal en utilisant les données morphologiques et d'ADN sédimentaire des communautés de foraminifères. Les approches morphologiques et moléculaires ont permis de délimiter les zones faiblement polluées des

zones fortement polluées. Cependant, les données moléculaires se sont avérées plus cohérentes avec les mesures géochimiques de la pollution. La reconstruction des conditions de référence avant l'industrialisation était l'objectif de la deuxième étude. L'analyse de l'ADN sédimentaire montre des variations dans la composition des communautés procaryotiques et eucaryotiques résultant de la dégradation de l'habitat. Avant l'industrialisation, la baie de Bagnoli était couverte d'une prairie marine avec une faune très diversifiée. En raison des activités anthropogéniques, l'herbier de posidonies a progressivement disparu et la communauté eucaryote a été remplacée par des dinoflagellés et des espèces métazoaires infauniques. L'ADN ancien sédimentaire nous permet de suivre les changements passés liés aux activités humaines et de retrouver les conditions de référence de l'époque préindustrielle.

Finalement, je discute des différents défis rencontrés lors de l'analyse des séquences d'ADN sédimentaire ancien et récent et je conclus en présentant les nouvelles technologies de séquençage et les méthodes de calcul qui pourraient être utilisées dans les futures recherches sur l'ADN sédimentaire.

Table of Contents

Chapter 1: Introduction

1.1. General background

The understanding of a natural ecosystem implies the study of the interactions between living organisms and their environment. The abundance and diversity of organisms are linked to variations in physical and biochemical factors. Changes in temperature, salinity, oxygen levels and major elements such as carbon influence the biodiversity.

To study biodiversity in a given environment, taxonomists first collect, identify and classify species, thus creating inventories (biological databases). Then, ecologists can relate the richness and diversity of species in that given habitat. This leads to the elaboration of species distribution maps according to the characteristics of the environment (biogeography). Similarly, (micro)palaeontologists study living remains, (micro)fossils, and establish relationship with the environment adding the temporal dimension.

Determining and compiling the biodiversity of any environment has become increasingly accessible with advances in molecular technologies, in particular with the advent massive high-throughput DNA sequencing (HTS). By using HTS to the amplified barcodes it was possible to simultaneously identify organisms in an environmental sample. This method, called metabarcoding, is based on PCR amplification, using either universal or specific-taxon markers (barcodes) and HTS sequencing. It provides a quick and easy overview of the different organisms (eukaryotes and prokaryotes) in the environment (Taberlet et al. 2018). The same approach can also be applied to ancient sediments that contain degraded and low-concentration DNA from diverse organisms that lived in a given environment at a given time. This allows access to past diversity and inference of past environments from sedimentary DNA (Pedersen et al. 2015; Capo et al. 2021).

In the following sections I will give some definitions and explanation for better understanding the sedimentary DNA preservation and applications.

1.2. Sedimentary DNA (sedDNA) and sedimentary ancient DNA (sedaDNA)

Environmental DNA (eDNA) refers to the total DNA obtained from a given environment, such as water, air, ice, soil, and sediment (Taberlet et al. 2012). In the latter, in addition to sedimentary particles (clays, sand), there are also remains of organisms (cells, shells, leaf

fragments) and DNA released by organisms during their life and decomposition after death. Although there is some controversy about whether environmental DNA should be distinguished into organismal and extra-organismal DNA, in the work presented here, sedDNA refers to all DNA extracted from sediments regardless of size or group of organisms (Pawlowski et al. 2020, 2021b; Rodriguez-Ezpeleta et al. 2021)

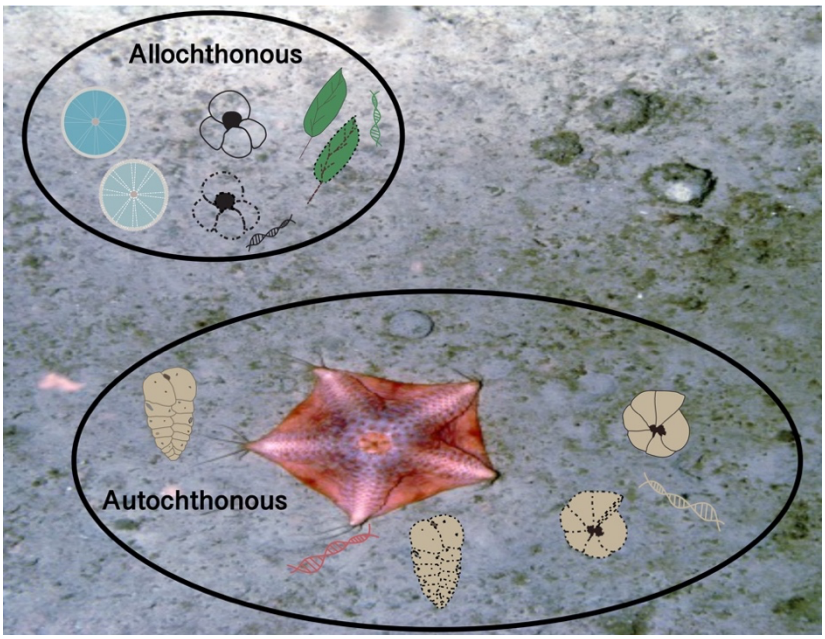

Figure 1.1: Seafloor with DNA inputs from allochthonous and autochthonous elements. Allochthonous elements are mainly remains of pelagic or terrestrial organisms. The autochthonous ones are benthic organisms, here deep-sea starfish, and foraminifera. The dotted organisms are in the decomposition phase. Free DNA represents extracellular DNA. Photo of the seafloor taken by the University of Plymouth's Autonomous Underwater Vehicles (AUV).

Marine sedimentary DNA can be differentiated according to its origin (Fig. 1.1). DNA from allochthonous organisms, i.e., pelagic organisms that have fallen to the bottom of the sea. These can decompose during their fall and arrive already degraded. Depending on the depth, they may also arrive intact, e.g., still with their shells, but will subsequently be degraded on the sea floor. Near a catchment, it is also possible to have terrestrial components, for example plant residues, as in the case of lakes that become a collection point for the remains of the flora and fauna surrounding them (Giguet-Covex et al. 2019; Capo et al. 2022). Finally, the DNA of benthic organisms, living on or within the sediments, is referred to as autochthonous DNA (Fig. 1.1).

DNA accumulated over geological time becomes ancient sedimentary DNA (sedaDNA). Thus, sediments become archives of past marine environments (Boere et al. 2011; More et al. 2018; Armbrecht et al. 2019). Although this DNA is subject to physical-chemical (Ultraviolet

radiation) and biological (bacterial activity) damage, it is often possible to extract and sequence it. Coolen and Overmann (2007) obtained 217 ka old prokaryotic DNA from Mediterranean sapropel, for the first time in the marine environment.

Since the beginning of marine sedimentary DNA research, many sediment samples of all ages have been analysed with a focus on either prokaryotes or eukaryotes. From the microbiomes of the Arabian Sea (52 ka) (Orsi et al. 2017), to the planktonic communities of the Black Sea (11 ka) (Coolen et al. 2013), the Mediterranean Sea (125 ka) (Boere et al. 2011). Sometimes some groups have been such as diatoms from the Arabian Sea (43 ka) (More et al. 2018) or from offshore California (11 ka) (Armbrecht et al. 2021b) and foraminifera from deep water sediments (32 ka) and shallow water in Svalbard (1 ka) (Pawlowska et al. 2014, 2016).

In terms of age, the 1.4 Ma chloroplast DNA is the oldest to date (Kirkpatrick et al. 2016).

1.3. Sedimentary DNA preservation

Under favourable environmental conditions such as low temperature and high pH, the DNA persists for a long time. Such places are found in cold sites and high latitudes. However, some studies have shown that DNA can be well preserved in oxic environments (Coolen et al., 2013, Lejzerowicz et al., 2013) and in tropical sites (lake sediments, Moguel et al. 2021; marine sediments, More et al. 2018). In Chapter 3.3, sedimentary DNA was recovered from a tropical warm pool (TWP) where seawater temperature remains high throughout the year (> 28°C, de Deckker 2016).

In general, DNA will be damaged over time leading to its fragmentation. This is caused by depurination (Dabney et al. 2013). Thus, sedimentary aDNA is found in small fragments, which ranges from ~500 bp in 18S rDNA (Boere et al. 2011) to 76 bp in rcbl DNA (Armbrecht et al. 2021a). SedaDNA is also subject to base modification or deamination over time. Cytosine (C) becomes uracil (U) (Dabney et al. 2013). The rate of deamination and the size of fragments can discriminate ancient DNA from modern DNA.

In addition to this physico-chemical degradation, there is biological damage due to microbial activity (Blum et al. 1997). For example, anoxic conditions can limit microbial activity and favour better preservation. The presence of clays and humic or organic matter (OM) in the sediment also contributes to DNA preservation. Released DNA rapidly binds to these particles in marine soils (Blum et al. 1997; Dell'Anno et al. 2002; Pietramellara et al. 2009; Torti et al.

2015). Thus, sediment mineralogy, and external factors (temperature, bacteria) favour DNA preservation.

1.4. Sedimentary DNA applications

The main applications of metabarcoding of ancient and recent samples are the monitoring of polluted sites, biodiversity survey and palaeoecological reconstructions. The monitoring of sites helps to identify the perturbed areas (chapter 7) by chemical pollutants (heavy metals). The assessment of resource exploitation impacts, such as oil and gas (Laroche et al. 2016; Cordier et al. 2019b; Frontalini et al. 2020a) and deep-sea mining (Ramirez-Llodra et al. 2010; Lins et al. 2021). Industrial exploitation affects deep-sea biodiversity and ecosystem functioning by destructing the habitats or realising chemical substances (Zeppilli et al. 2016; Frontalini et al. 2018a, 2020a).

Metabarcoding facilitates the inventory of present environments with limited access, such as deep-sea (chapters 5 and 6) and past environments having access to temporal changes (chapters 3, 4 and 8).

Past environments reconstructions: In the past geological periods the changes had been driven by climate and in recent past years, since the Anthropocene, ecosystem's perturbation has been accentuated by human activities (anthropogenic impact). Metabarcoding enables the possibility to characterize the biodiversity before the pollution and to make a baseline of pristine environments accompanying restoration measures (Chapter 8).

SedaDNA, together with other paleo proxies, allows the reconstruction of past oceanographic conditions, Sea Surface temperatures (SST) (Zimmermann et al., 2021), salinities and oxic/anoxic variations (More et al. 2019), sea-ice presence (De Schepper et al., 2019; Pawłowska et al., 2020a; Zimmermann et al., 2021), and even species evolution related to climate changes (Pawłowska et al. 2020; Armbrecht et al. 2022).

Compared to conventional methods, the main advantage of sedDNA metabarcoding of recent or past samples is its capacity to retrieve all organisms regardless of size fraction, from microbes to megafauna, and regardless of their ability to preservation, including soft and hard bodies organisms. This global character of metabarcoding data gives more complete overview of present or past ecosystems.

1.5. Deep-sea Habitats

The samples used in this thesis covered mainly the deep-sea benthic habitats. The deep-sea can be divided in bathyal, abyssal and hadal zones (Fig. 1.2). Their limits are variable, but in general the bathypelagic zone extends from 1000 m to 4000 m, the abyssal zone from 4000 m to 6000 m and finally the hadal zone from 6000 m to -11 000 m. The deep-sea is the largest biome on Earth but is less known. It represents approximately 60% of the lithosphere (Glover and Smith 2003). The deep-sea cartographies (Wright 1999; Wölfl et al. 2019) reveal a varied topography with seamounts, hydrothermal vents, and trenches. In such heterogenic environments, it is predicted to encounter a high number of coexisting species in each area (Danovaro et al. 2010; Costa et al. 2020; Paulus 2021). Many eDNA studies have demonstrated the existence of biodiversity hotspots in seamounts and high richness around hydrothermal vents (Ramirez-Llodra et al. 2010; Consalvey et al. 2010; Ebbe et al. 2010). However, the knowledge of whole deep-sea biodiversity remains still unknown, and many species new to science are continuously discovered during scientific expeditions (Danovaro et al. 2010; Ramirez-Llodra et al. 2010; Goineau and Gooday 2017; Bribiesca-Contreras et al. 2022).

Among the most diverse abyssal zones is the Clarion Clipperton Fracture Zone (CCFZ) which lies between the tropics in the Pacific Ocean. In 2020, I had the opportunity to go there to collect surface sediments for eDNA metabarcoding and to sort foraminifera.

This vast zone hosts the greatest concentration of polymetallic nodules, but also a great biodiversity on all scales (megafauna to microbial). Besides morphological descriptions, eDNA metabarcoding surveys were also applied to recover biodiversity, confirming the high biodiversity in CCFZ. Moreover, many sequences could not be assigned to known taxa, highlighting a gap in knowledge of the biodiversity of the area (Lejzerowicz et al. 2021; Jones et al. 2021).

Despite recent and abundant investigations, the impacts of deep-sea mining on biodiversity by disturbing benthic habitats remain poorly understood (Macheriotou et al. 2020; Lins et al. 2021; Pape et al. 2021).

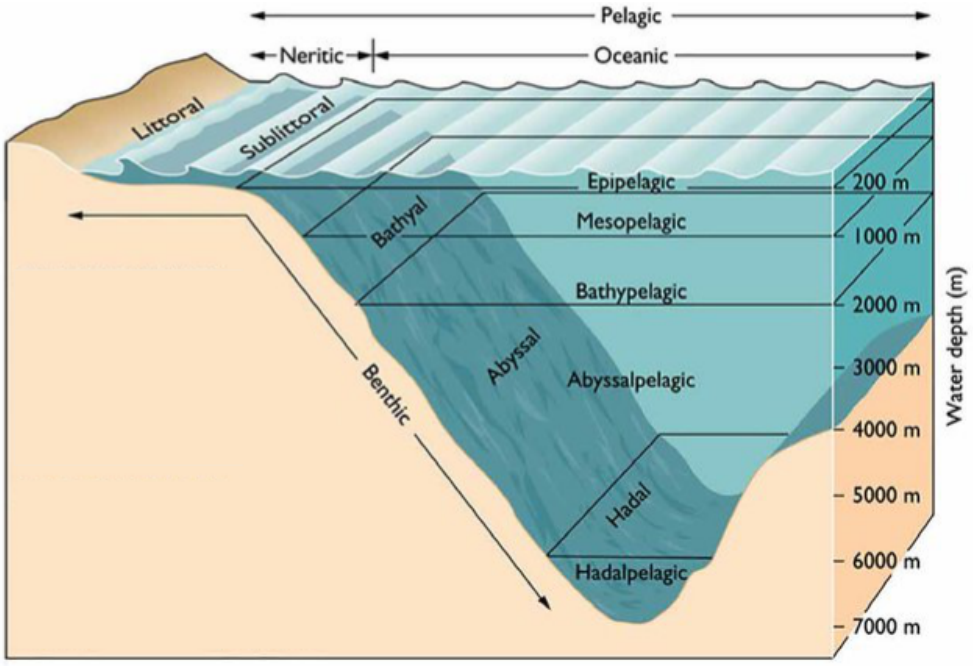

Figure 1.2: Vertical zonation of the ocean, in the pelagic and benthic environments. Modified from (Sayre et al. 2017).

1.6. Foraminifera

Although in my thesis I have analysed various components of marine diversity, my main interest was in analysing unicellular eukaryotes (protists) with special emphasis on foraminifera.

Foraminifera are rhizarian protists characterized by a granuloreticulate pseudopodia and sometimes by a test, a shell, englobing the cytoplasm. The test can be either made of calcium carbonate (calcareous test) or of external particles cemented to the organic wall (agglutinated test). Some foraminifera possess soft-walled theca or are naked.

Foraminifera have been classified into three classes according to their test morphology and molecular phylogeny: Monothalamea, Globothalamea and Tubothalamea (Pawlowski et al. 2013). There are also other taxonomic groups, for which phylogenetic data are lacking such as Nodosaria (Lagenida).

Monothalamea are single-chambered and constitute a paraphyletic group, which means that they are at the basis of two other groups. Their test is either agglutinated or organic. Monothalamea are divided into 28 clades including 8 marine environmental clades ENFOR (1-8) (Pawlowski et al. 2011), 4 others from freshwater and soil samples FWG (1-4) (Lejzerowicz et al. 2010), and undetermined monothalamids.

The multichambered foraminifera are Globothalamea with a globular chamber and Tubothalamea with tubular chambers. Their test is hard-walled, calcareous or agglutinated. Globothalamea are subdivided into Robertinids, Rotallids and Textularids while Tubothalamea are classified into Miliolids and Spirillinids.

The latter two classes are well preserved in microfossil register and widely studied by micropalaeontologists. However, molecular studies have shown a high diversity and richness of monothalamids, especially in deep-sea samples (Lejzerowicz et al. 2015a; Gooday et al. 2017; Goineau and Gooday 2017; Cordier et al. 2019a).

Foraminifera range size varied from under 63 µm for tiny foraminifera (*Rotaliella elatiana*, (Pawlowski and Lee 1991) to several cm for giant ones. Among these, those reaching about 15 cm include the *Xenophyophores* in the deep-sea (Gooday et al. 2017) and *Jullienella foetida* in shallow waters (Langer et al. 2022).

Foraminifera are present in all habitats, but especially in marine environments. They are present at all depths, from the surface to the bottom of the ocean. They can be divided according to their vertical distribution on the ocean, into planktonic and benthic (Fig. 1.3).

1.6.1. Planktonic foraminifera

Planktonics float in the upper part of water column (photic zone) in open ocean. They include nearly 50 extant living morphospecies having several genotypes (Kucera and Darling 2002; Darling and Wade 2008; André et al. 2014). They belong to the Globothalamea class and Rotaliid order (Pawlowski et al. 2013). Molecular evidence has shown that planktonics have evolved from benthics (Schiebel and Hemleben 2017). Extant planktonic foraminifera are geographically distributed into five provinces according to their sensitivity to sea surface water temperature: tropical, subtropical, temperate, subpolar, and polar (Kucera 2007; Schiebel et al. 2018). Because of their exceptional preservation in oceanic sediments, they have become palaeoceanographic, paleoclimate and biostratigraphic proxies. Planktonics have been widely used for the reconstruction of the ecology, oceanic circulation, and climate due to their water temperature sensitivity (Schmiedl 2019). In addition, their tests record the chemical properties of the seawater.

1.6.2. Benthic foraminifera

Benthics include about 4000 to 10,000 living species (Murray 2007; Hayward et al. 2022), counting only hard species. Many more if we consider all the monothalamids and other deep-

sea species. Benthics count for 50% of deep-sea meiofauna and are an important component of macrofauna. Some also constitute the deep-sea megafauna (Xenophyophores)(Gooday 2019). They are sediment-dwelling or epifaunal species, i.e., they may live on the seabed or be attached to the substrate. This substrate can be rocks, polymetallic nodules, as in the case of xenophyophores, coral reefs and algae/seagrass.

Benthic foraminifera occupy all bathymetric levels, from reefs to hadal zones to abyssal plains. Environmental factors such as food supply (organic matter), salinity, oxygen, nitrogen, and currents influence their temporal and spatial distribution (Gooday 2014; Murray 2014).

Species density and composition (assemblages) change in response to chemical and physical factors (Alve 1999). Thus, benthic species are commonly used as bioindicators for monitoring of marine environment (Schönfeld et al. 2012).

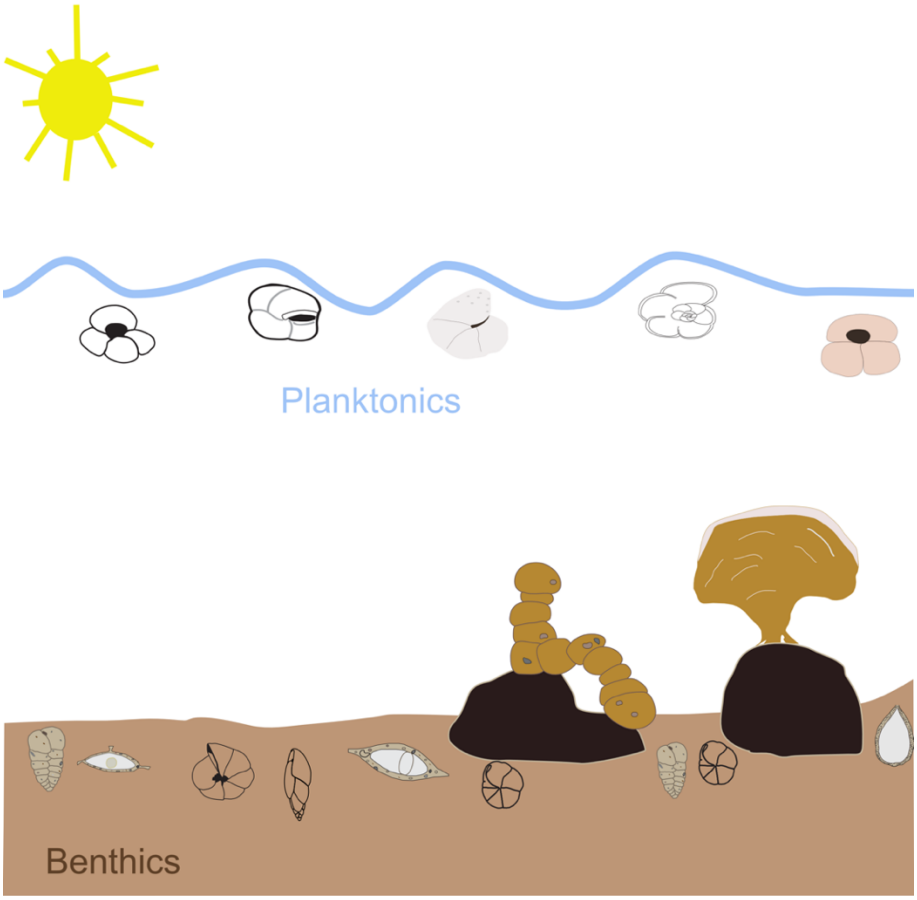

Figure 1.3: Foraminifera mode of life: Planktonics occurring preferentially in the photic zone and benthics living either inside the sediment or at its substrate, sometimes attached to hard substrate (such as xenophyophores on nodules, in dark brown here). Drawings of foraminifera not represented to scale.

1.7. Objectives

In this thesis, I present sedimentary DNA (*sed*DNA) and sedaDNA studies dealing with ecological, biodiversity and methodological questions. My thesis is divided in three areas: Reconstruction of paleoenvironments, diversity survey and biomonitoring (Fig. 1.4)

The main research goals are:

1) to understand how DNA is deposited and preserved in marine sediments.
    a. comparison of allochthonous and autochthonous sedimentary DNA
    b. preservation of sedaDNA in low latitudes (tropics)
2) to identify and document deep-sea biodiversity:
    a. assessing deep-sea benthic diversity at local or global and temporal scale
    b. taxonomic assignment of deep-sea lineages
3) to use sedimentary DNA for biomonitoring
    a. defining current ecological status of a coastal site
    b. inferring preindustrial ecosystem conditions

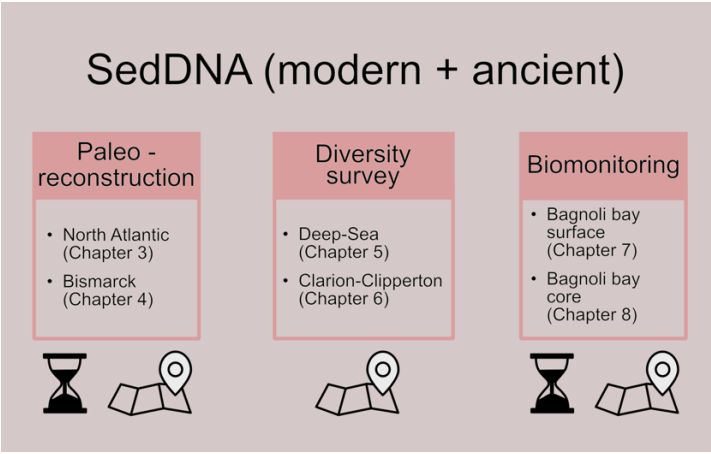

Figure 1.4:Schematic representation of this thesis. The work is subdivided in three main topics: Paleoreconstruction, diversity survey and biomonitoring. The time and space icons showed which of these two factors were involved.

Chapter 2: Methodology

This chapter summarises the workflow followed to process sedimentary samples for eDNA or aDNA metabarcoding. Figure 2.3 illustrates all the steps from the recovery of the material to the analysis of sequenced metabarcodes. Some descriptions will focus mainly on foraminifera sedimentary DNA.

2.1. Sedimentary material for eDNA and aDNA (Fig. 2.3 A.)

Various types of cores are used to retrieve sediments from the seabed, depending on the purpose of the study and the water depth. Multicores or box corers are used for recent material and only the first few centimetres are sub-sampled (Chapters 4.1, 4.2 and 5.1) (fig 2.1). Multicores or gravity/piston cores (Chapters 3.1, 3.2 and 5.2) are used to retrieve older sediments (1832-340,000 years).

Subsequently, subsampling is completed to the desired resolution, taking sediment every 1-2 cm and avoiding the core edges and exposed surface. In the case of Chapter 3.2, subsampling was carried out on a half-core cut lengthwise by selecting specific layers according to the ages of the core. Certain precautions were taken when collecting sediments for eDNA (Taberlet et al. 2018; Capo et al. 2021) or sedaDNA (Epp et al. 2019; Armbrecht et al. 2019) studies to avoid contamination during the whole sampling and sub-sampling process. Single-use and sterile equipment (gloves, tubes, spoons, syringe, etc.) or equipment previously decontaminated with bleach (metal spatula, etc.) between two samples are used. After subsampling, the tubes containing the sediment are either filled with LifeGuard solution or frozen immediately at 20°C. This is to limit further degradation of the DNA and/or RNA. All collected sediment was frozen and shipped to the laboratory of the University of Geneva.

Sometimes subsampling could not be completed in situ, so the core was refrigerated or frozen. This is the case in Chapter 3.2, where subsampling was carried out 14 years after coring retrival. Half of the core was stored at 4°C in a core repository at CEREGE (Aix-en-Provence, France). One of the gravity cores (unpublished), it was frozen during the cruise, so subsampling had to be done in a room at -20°C to avoid freeze-thaw cycles.

The studied sites covered large marine areas from the tropics to the sub-polar regions such as the Pacific Ocean, the Atlantic Ocean, the Bismarck Sea, and the Tyrrhenian Sea. Sedimentary cores were retrieved from deep and shallow waters ranging from -9,000 m (Kuril Kamchatka Trough) to -60 m (Bagnoli Bay) (Fig. 2.2).

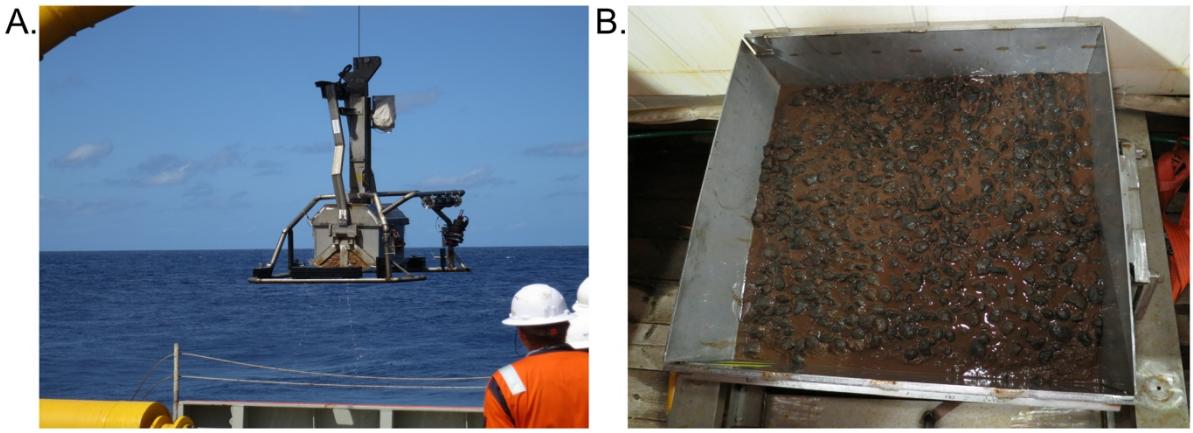

*Figure 2.1: **A.** Box corer being recovered from CCZ area, **B**. A top view of a box core of 1m x 1m x 1m. Mud and polymetallic nodules can be seen on the surface.*

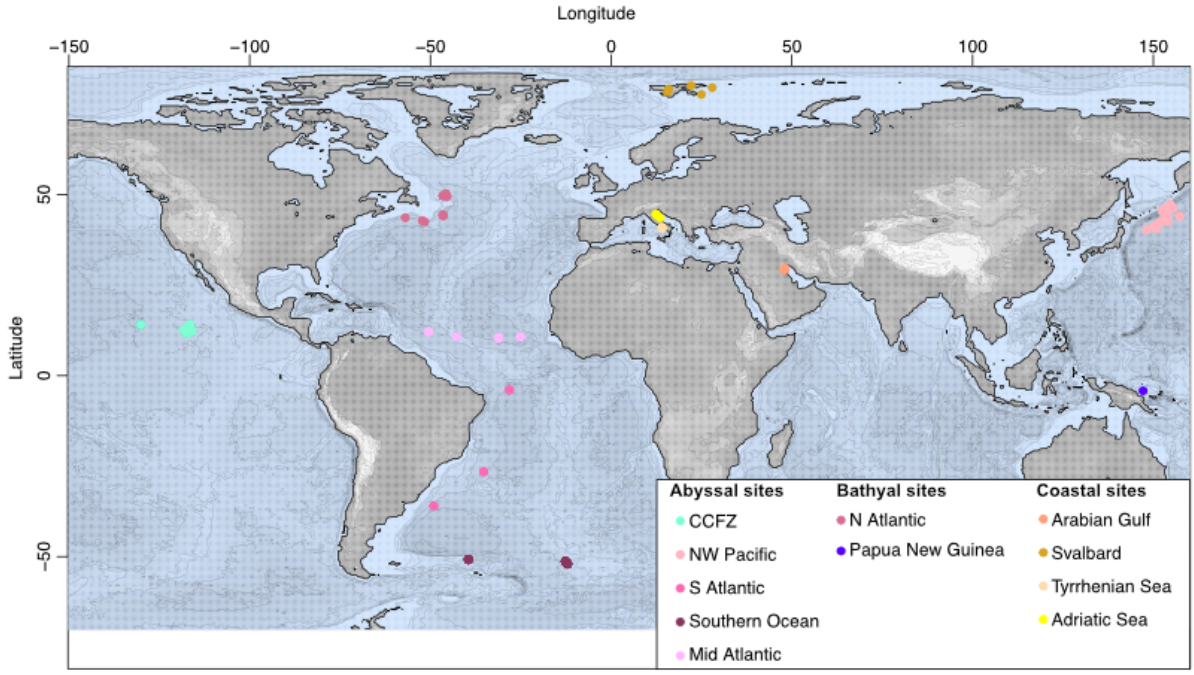

Figure 2.2: Location of sampling sites. They are classified according to their bathymetry: Abyssal sites corresponding to -9000 m to -4000 m, bathyal sites to -4000 m to -1800 m and coastal sites to less than -100m.

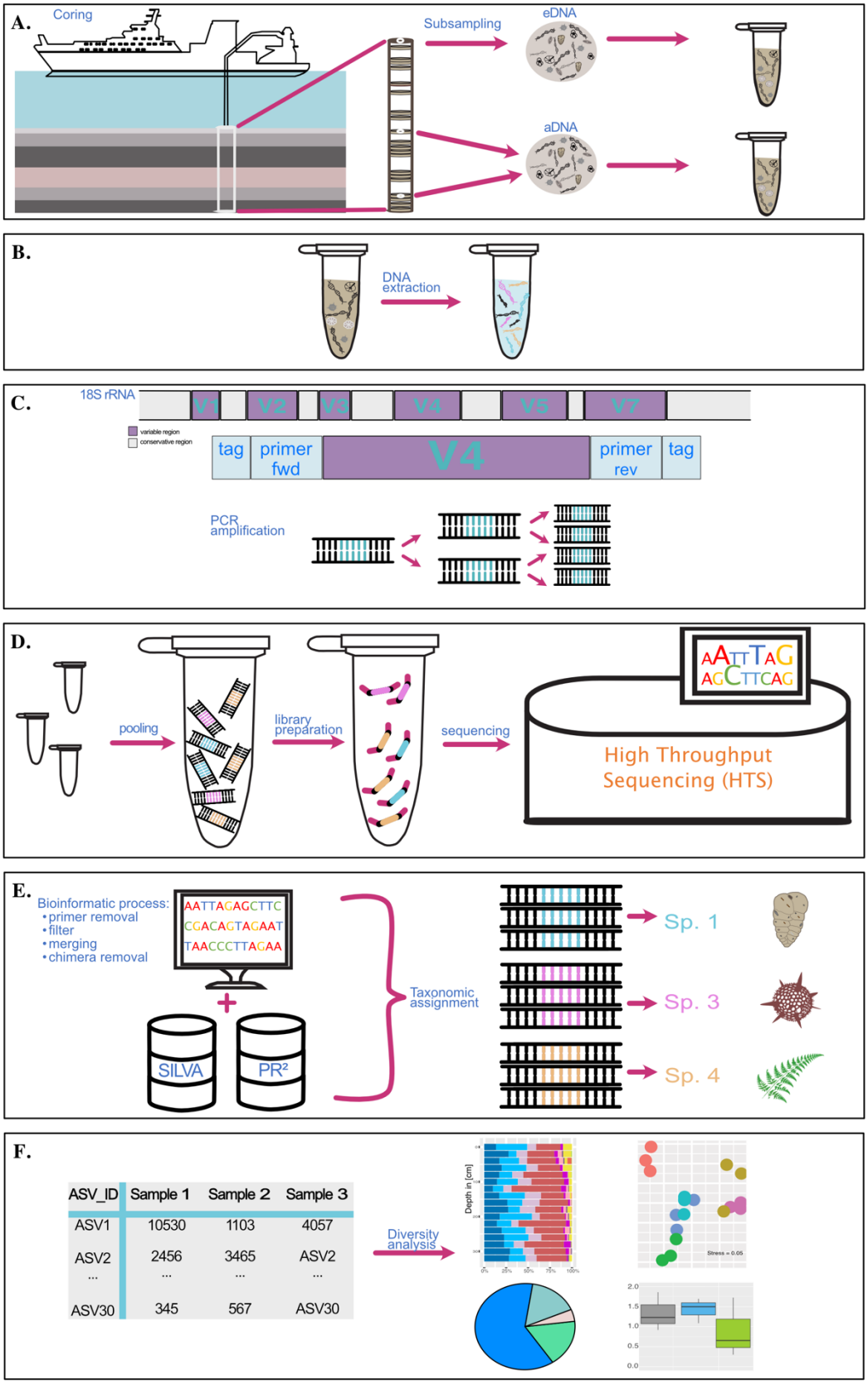

*Figure 2.3: Schematic metabarcoding workflow. **A**. Sampling and subsampling of a core: first centimeters (recent sediment) for eDNA metabarcoding, deep layers (older sediments) for ancient eDNA; **B**. Total DNA extraction; **C**. PCR amplification using tagged primers which target a variable*

region in the 18rDNA gene; ***D****. Pooling amplicons, library preparation where adaptors and indexes are added to each pool and finally sequencing (HTS);* ***E****. Bioinformatics steps comprising primer removal, filtering, trimming reads and removal of chimaeras. Comparison of obtained sequences to reference database allows their assignment, and;* ***F****. Sequences list, species distribution, and ecological analysis alpha and beta diversity.*

2.2. DNA extraction (Fig. 2.3 B.)

DNA was extracted using commercial kits: DNeasy PowerSoil for a small amount of sediment (0.5g – 1g) and DNeasy PowerMax Soil for samples up to 10g (5-7 ml). In the case of PowerMax, after the final elution, an additional precipitation step was required to concentrate the DNA. We add 0.2 m of 5M NaCl and 10.4 ml of cold EtOH (99%) and left overnight at -20°C. After the centrifugation of the mixture for 30 min, the supernatant was discarded and resuspended by adding 400 µl of C6 solution when the samples were recent, 300µl when the samples were older. At least one extraction control was added to each extraction batch per session. These controls are necessary to ensure the cleanliness of the room and the reagents and to control contamination.

2.3. Markers gene and primers (Fig. 2.3 C.)

Ribosomal, mitochondrial and chloroplast genetic markers are used depending on the targeted group of organisms and the purpose of the studies. The small subunit of the ribosomal gene SSU rRNA or 18S rRNA contains conservative and variable regions (V1-V9) (Hadziavdic et al. 2014). Genetic differences in those hypervariable regions (e.g., V4 and V7) enable differentiation of taxa at higher resolution, even at the species level. In the V4 and V9 regions, universal primers are used to obtain eukaryotic amplicons. The V9 region is smaller than V4 and, therefore, more suitable for amplifying degraded or ancient samples. In addition to the universal primers, there are specific primers for phyla, such as foraminifera (Pawlowski and Lecroq 2010), or more specific taxa (e.g., Cephalopoda; de Jonge et al. 2021, nematodes; Floyd et al. 2005).

Other regions in the mitochondrial cytochrome c oxidase I (COI) gene are used to amplify metazoans (Leray and Knowlton 2015). Finally, in the chloroplast trnL gene, primers targeting the P6 variable region allow amplifying plants (Taberlet et al. 2007).

The choice between numerous primers and targeted genes is based on the purpose of the study. If the objective is to obtain multiple phyla to assess the whole biodiversity or certain functional

groups (e.g., planktonic), it is preferable to use universal primers. If the study is focused on a particular group, or on cryptic species, specific primers will be more appropriate and will avoid amplification of unwanted taxa.

Taxonomic group	Gene - Locus	fwd name	Forward (5'-3') Reverse (5'-3')	rev name	Amplicon length (bp)	Reference
Eukaryotes	18S rDNA - V9	1389F	TTGTACACACCGCCC CCTTCYGCAGGTTCACCTAC	1510R	~100 -130	(Amaral-Zettler et al. 2009)
Foraminifera	18S rDNA - 37f-41f	s14F1	AAGGGCACCACAAGAACGC CGGTCACGTTCGTTGC	s17	~280 - 300	(Pawlowski and Lecroq 2010)
Foraminifera	18S rDNA - 37f	s14F1	AAGGGCACCACAAGAACGC CCACCTATCACAYAATCATG	s15.3	~190 - 230	(Pawlowski and Lecroq 2010)
Foraminifera	18S rDNA - 37f	s14F1	AAGGGCACCACAAGAACGC GAAAGGACTAGCATATTTAAC	s15rotex	~120 - 140	(Lejzerowicz et al. 2013a)
Radiolarian	18S rDNA - V4rad radiolaria	S879	CCAGCTCCAATAGCGTATAC CCACCTATCACAYAATCATG	s32Jmod	~250 - 260	(Decelle et al., 2012)
Eukaryotes	mt COI	mlCOIintF	GGWACWGGWTGAACWGTWTAYC CYCC TAAACTTCAGGGTGACCAAARAAY CA	dgHCO2198	~310 - 320	(Leray and Knowlton, 2015)
Plants	trnL - P6 loop	g	GGGCAATCCTGAGCCAA CCATTGAGTCTCTGCACCTATC	h	~10 - 143	(Taberlet et al. 2007)

Table 1: The taxonomic groups targeted in this thesis with their respective markers (gene-locus). Forward (fwd) and reverse (rev) primers with their respective sequence from 5'-3. Markers with their approximative size.

2.4. Amplification Pooling, library preparation and sequencing (Fig. 2.3 C. –D.)

A tag of eight nucleotides is attached to each 5' extremities of forward and reverse primers (Esling et al. 2015a) allowing multiplexing samples (Fig. 2.4). The targeted regions are amplified by Polymerase Chain Reaction (PCR). The PCR amplification are performed in a reaction volume of 25µl using Taq polymerase, reaction buffer containing $MgCl_2$, Bovine Serum Albumin (BSA), and DNA free water. PCR conditions (temperature, time, and number of cycles) change according to primers characteristics. For modern DNA sample, three PCR replicates are obtained and for ancient DNA five to eight replicates, pooled and checked on agarose gel (1.5%). In addition, PCR blanks are performed for each sample. The amplicons are quantified with a high-resolution capillary electrophoresis with QIAxcel System (QIAGEN). Based on the quantification, the pools are prepared to get samples with the same DNA concentration. To remove primer-dimers formed during the PCR amplification, the pools are purified using the High Pure PCR Product Purification kit (Roche) which removes fragments under 100bp. In the case of shorter amplicons (<70bp) such TrnL, the purification step is performed using MinElute PCR purification kit (QIAGEN). The pools are then prepared to be

sequenced with a using TruSeq® DNA PCR-Free Library Preparation Kit (Illumina) kit. In this step, adapters and indexes are appended to each amplicon in the library (pool). This ensures that each library is identified by a label and that multiple libraries can be sequenced per run. Each library is measured by qPCR using the Kapa Library Quantification kit and loaded equimolarly onto the Illumina Miseq sequencer. The V2 reagent kit for paired-end sequencing with 300 cycles sequencing small amplicons (e.g., V9, TrnL, 37F foraminifera) and 500 cycles for longer amplicons (e.g., COI, diatoms).

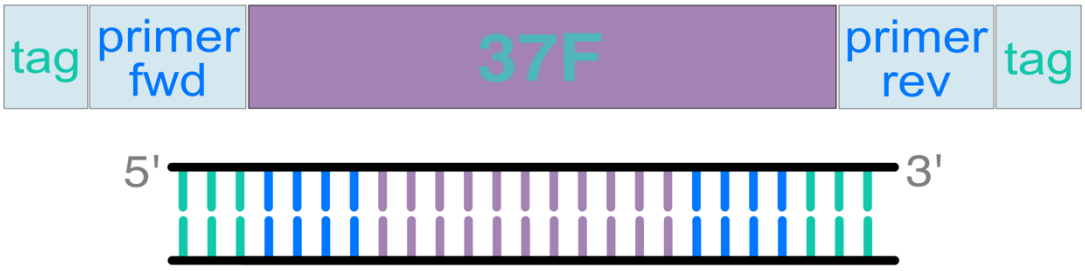

Figure 2.4: The 37F amplicon flanked by forward and reverse primers and their own tags at each extremity.

2.5. Bioinformatic analysis (Fig. 2.3 E. – F.)

2.5.1. Pre-processing of dataset (Fig. 2.5)

The obtained raw reads were analysed using SLIM platform (Dufresne et al. 2019) which integrates modules with algorithms such VSEARCH (Rognes et al. 2016), DADA2 (Callahan et al. 2016). The single pairs of FASTQ files (raw files) were demultiplexed using the sample tags with the module *demultiplexer* allowing 0 mismatches. The sequences (reads) not having the primers and the tags at the extremities were discarded.

After demultiplexing, the sequences were processed using several independent tools (modules). The aim is to pass quality filters and to keep only good sequences to be identified taxonomically. The steps to achieve this were the removal of primers and low-quality sequences, then merging of reads, and the removal of chimeras. After this step, the sequences were either be grouped into operational taxonomic units (OTUs) or processed with *DADA2* to obtain amplicon sequence variants (ASVs). The final fasta files contained all the OTU or ASV sequences and their distribution in samples.

Since the introduction of the DADA2 package (Callahan et al. 2016), it has been widely used in eDNA studies (ref) as it eliminates sequencing errors by generating an error model with the dataset itself and infers the exact variants of the amplicon. One of the advantages is the assessment of biodiversity at a finer level (nucleotide-level), a difference of one nucleotide results in another ASV (Callahan et al. 2017). This can help to identify not only species, but also subspecies, cryptic species, and polymorphisms. In Pawłowska et al. 2020, the ASV scale was used to track genetic variation of *N.pachyderma* over time and in relation to palaeoceanographic changes. Other advantages are reproducibility as the ASV is the smallest single unit and therefore facilitate the comparison between samples (Callahan et al. 2017)

In studies using OTUs, sequences are generally clustered. Sequences can be regrouped by an arbitrary similarity threshold typically between 97 – 99 % (Edgar 2018; Xiong and Zhan 2018) with one of the algorithms such as VSEARCH (Rognes et al. 2016) and USEARCH (Edgar 2010). Another possibility of clustering sequences is by nucleotide distance *d* (SWARM; Mahé et al. 2015), i.e., by iteratively aggregating sequences according to a network approach. The clustering aims in general to reduce PCR and sequencing and the number of sequences errors (Mahé et al. 2015; Rognes et al. 2016; Edgar 2018). Other algorithms such as LULU (Frøslev et al. 2017) help to filter erroneous clusters, created during the PCR amplification and sequencing, or from intra-individual variability. LULU identifies and merges co-occurring ASV/OTUs assuming that they are artefacts and considers their abundance in a dataset.

The choice between ASV or OTU depends on the research objective and the desired level of assignment. There is still much debate about the use of one of the two approaches, with some authors suggesting combining them (Joos et al. 2020; Brandt et al. 2021) . On the one hand, some studies show that the ESV/ASV approach should be used rather than the OTU approach (Callahan et al. 2017; Porter and Hajibabaei 2020) as it allows to detect small variants giving a full diversity and advantages as previously mentioned. Other studies also point out that both approaches have similarities in microbial community composition patterns at taxonomic level of phyla to family (García-López et al. 2021; Kerrigan and D'Hondt 2022). On the other hand, some authors suggest that neither approach should be chosen, but both ASV and OTU are complementary and should be combined. Hence, Brandt et al. 2021 propose to cluster ASVs and then use a LULU curation for metazoans. Similarly, Antich et al. 2021suggest that denoising (ASV) and clustering (OTU) should be combined in COI metabarcoding. In this way ASV is equated as haplotype-level proxy and OTU as species-level proxy.

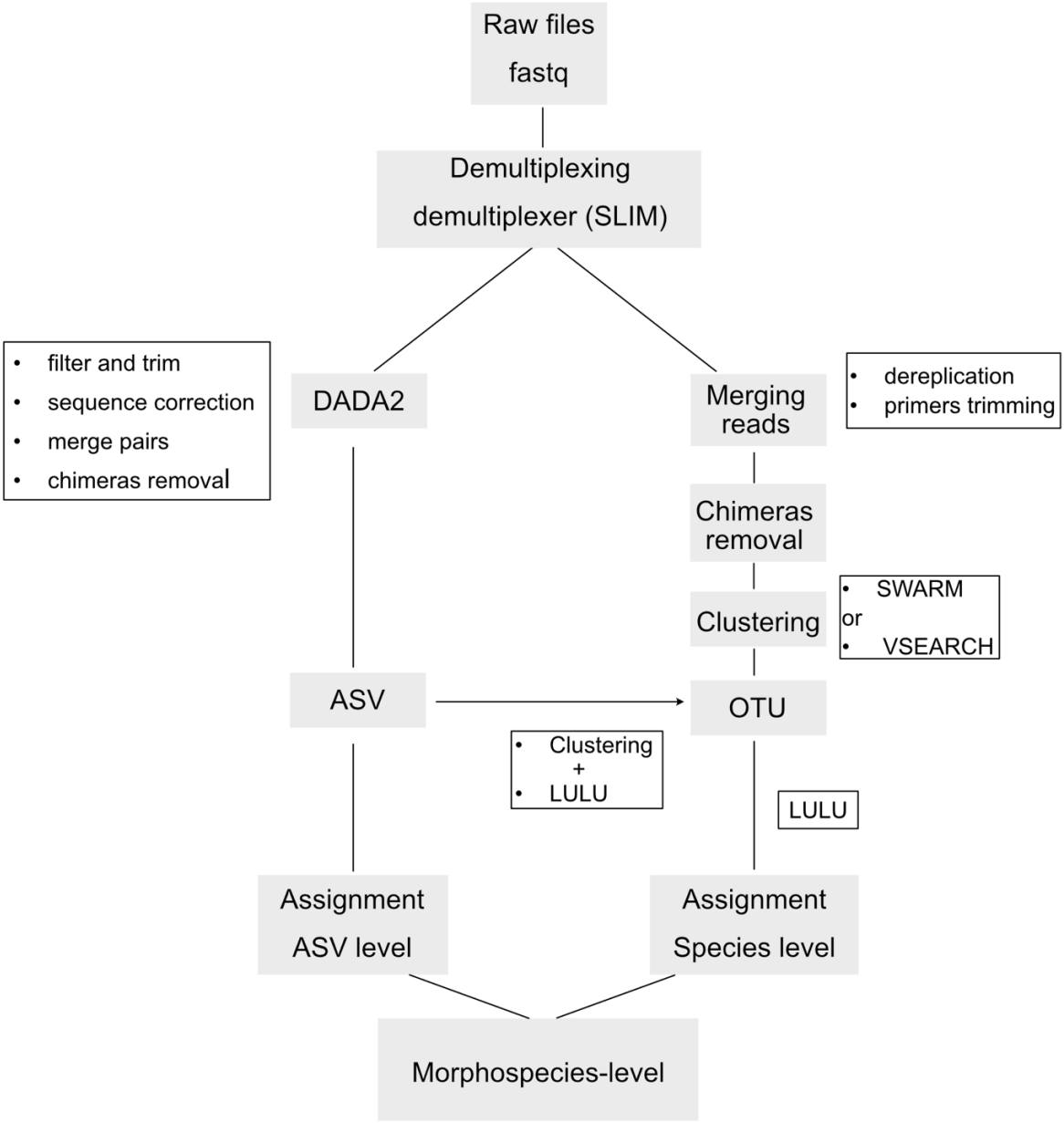

Figure 2.5: Bioinformatic workflow. From fastq raw sequences to ASV or OTU sequences and finally to morphospecies annotations.

2.5.2. Taxonomic assignments

Multiple tools are available to assign sequences; some are probabilistic and alignment-based approach such as VSEARCH (Rognes et al. 2016), Blast (Camacho et al. 2009) that compare the final sequences to reference sequences from a database. Others include phylogenetic, machine learning and alignment-based approach such as IDTAXA (Murali et al. 2018). However, none of these methods assigns all the sequences. Depending on the quality and the completeness of the reference databases, a part of sequence remains unidentifiable. In chapter

4.2 we present a method to deal with unassign sequences classifying them in different groups according to a signature. Thus, these sequences can be included in downstream statistical analyses.

2.5.3. Database

There are public reference databases curated according to gene and taxa. For example, PR2 (Protist Ribosolaml Reference Database) (Guillou et al. 2013) for 18S, MIDORI2 (Leray et al. 2022) for COI, PFR2 (Morard et al. 2015) for planktonic foraminifera. Customised curated databases can be constructed from barcoded single-cell specimens (benthic foraminifera from Pawlowski's lab) or by regrouping public sequences of selection of taxa from NCBI nucleotide database, GenBank. There is no complete or extensive database, lack of references can be seen quickly by the high number of unassigned ASV/OTUs in metabarcoding studies. This occurs often on unexplored or little-known taxa or sites such as deep-sea as shown in Chapter 4.1 (Cordier et al. 2022).

2.6. Metabarcoding Foraminifera

2.6.1. Foraminifera primers

For the genetic identification of foraminifera, the 3' fragment of ribosomal RNA small subunit (SSU) gene is widely used. In this gene fragment, six variable regions have been identified: three appearing only in foraminifera 37/f, 41/f and 47/f; Pawlowski and Lecroq 2010), characterised by an expansion forming a loop in the RNA helix, and the other three corresponding to helices 43e, 45e, 49e, are common in eukaryotes (Fig. 2.6).
It should also be mentioned that recently another gene, COI (LeraXT foram), has been used to identify certain forams from rotaliids and miliolids orders (Macher et al. 2021, 2022; Girard et al. 2022).

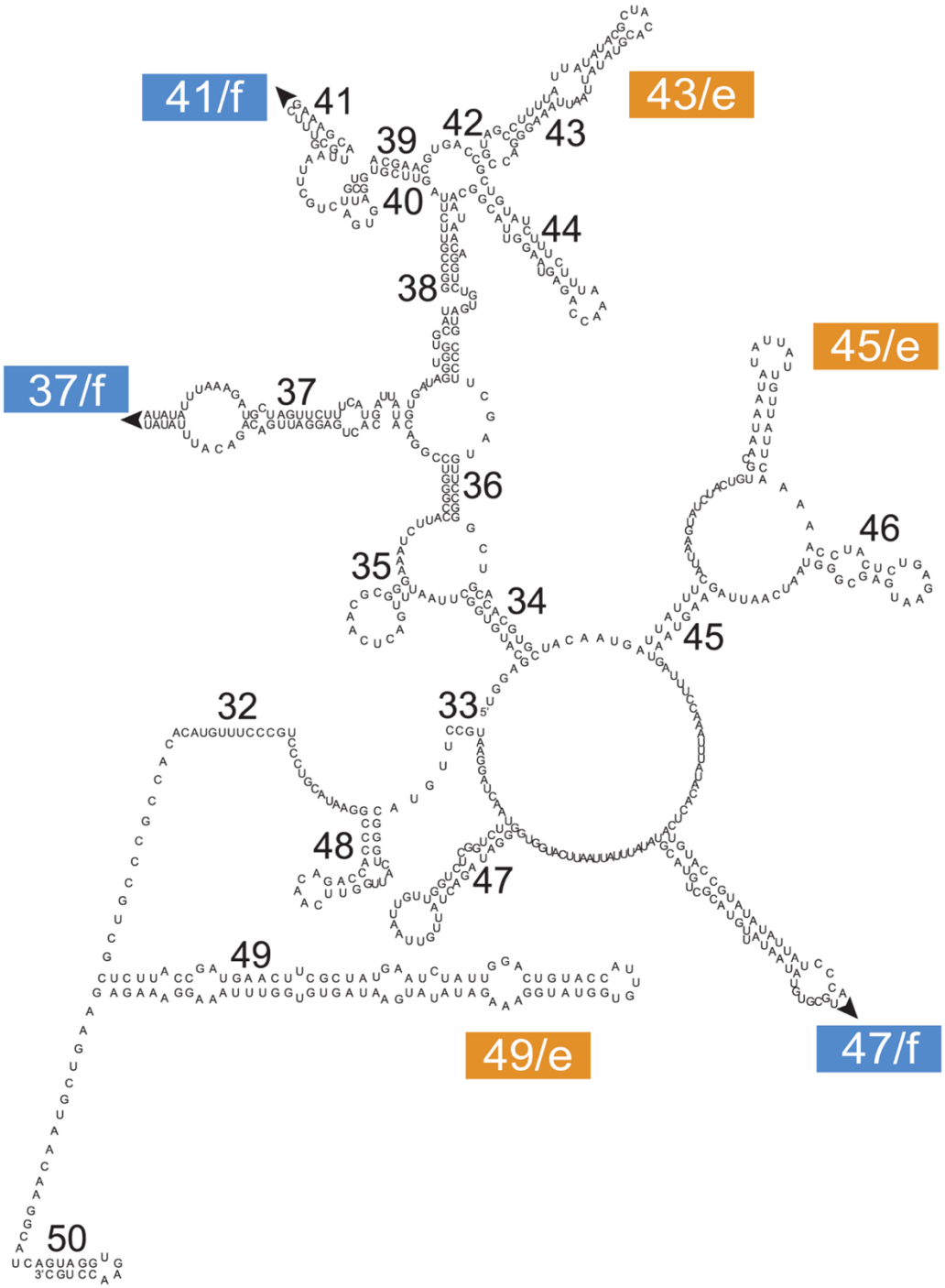

Figure 2.6: Predicted secondary structure of the small subunit (SSU) rRNA of Micrometula hyalostriata. The orange boxes with "e" marked the eukaryotic helices and those in blue with "f" the specific helices for foraminifera. Modified from (Pawlowski and Lecroq 2010).

The highly variable regions allow taxonomic resolution up to the species level and even cryptic species are identified (*G.glutinata I,II, III*). The entropy profile below shows how variable the region is and that it is flanked by conserved areas in which primers are designed (Fig 2.7).

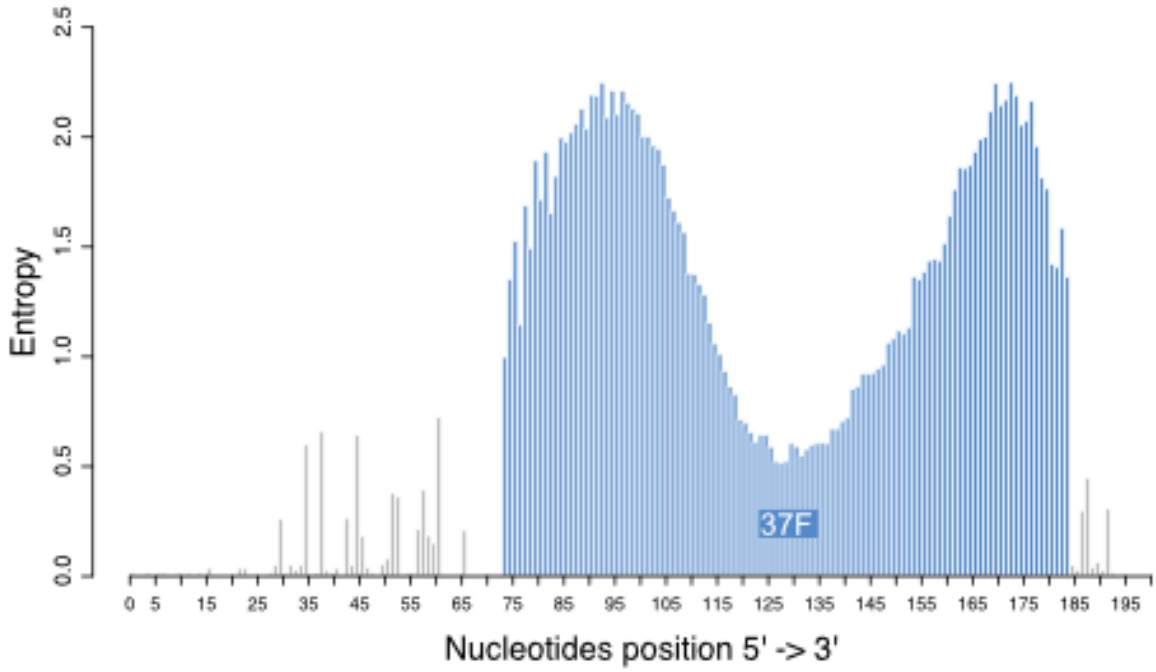

Figure 2.7: Entropy profile generated from an alignment of several sequences. The higher the values, the greater the genetic variation.

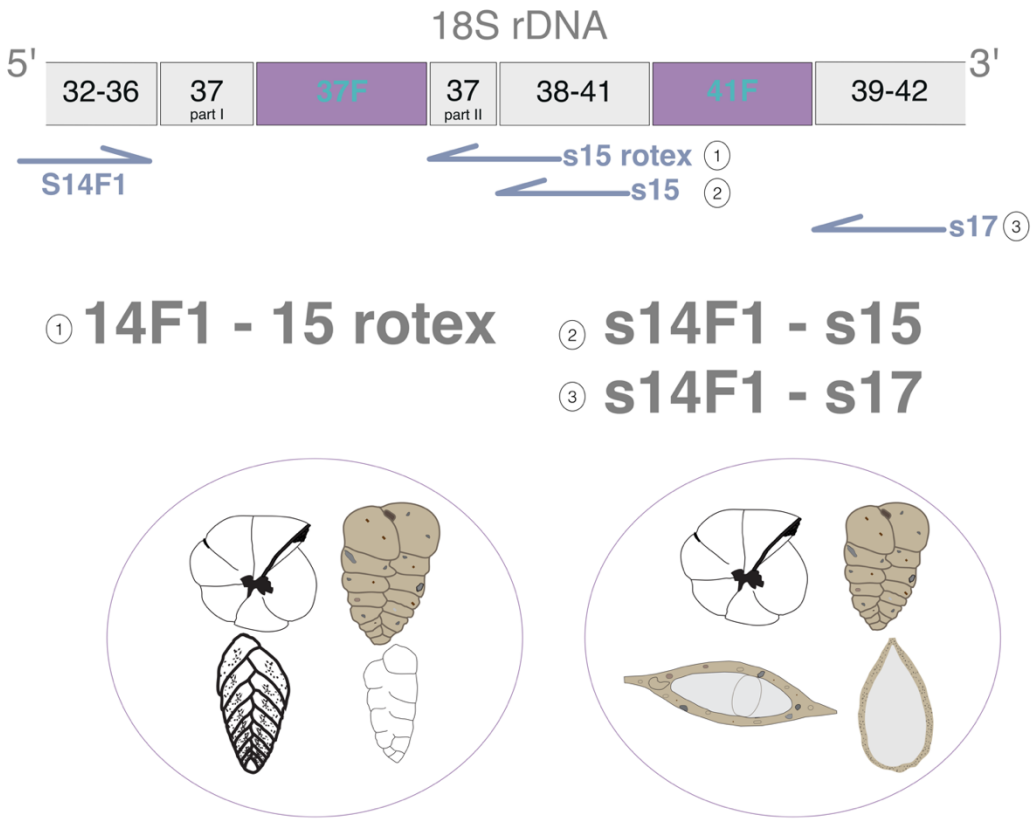

Figure 2.8: The SSU 18s rRNA fragment with the hypervariable regions specific to foraminifera (37F and 41F). The three primers are showed by half arrows and numbers. The arrows show the direction and estimated size of amplicons. The first primers amplify mainly calcareous, agglutinated hard-shelled foraminifera while the second and the third hard and naked soft-shelled foraminifera.

There are two so-called versatile primers for foraminifera, focusing in one short region or in two regions. The first pair, 14F1-s15, targets only the 37F region (90-170 bp) and the second, 14F1-s17, covers the 37F and 41F regions (190-410 bp). There is also a third pair of primers, 14F1-s15rotex (100-130 bp), that also target the 37F region but preferentially amplify hard-shelled foraminifera (rotalids and textulariids; Lejzerowicz et al. 2013a) (Fig. 2.9). These later primers can be used to suit morphological studies.

The short amplicons can be used for very ancient and degraded samples while the long amplicons can be used to better discriminate between species when clades are short in the region 37F, such as clade Y or for phylogenetic analysis.

1.1.1. Processing raw data

After obtaining a clustered or denoised sequences, we proceed to inspect the fasta file and apply some corrections to remove non foraminiferal reads. For the small region 37F, we keep reads having the beginning of 37 part I “GACAG” and the ending of 37 part II “TAGTCCCTT” or “TAGTCCTTT”. For long amplicons (37F + 41F), we keep sequences with “GGTGGT”. Depending on the research questions, for ancient eDNA, we prefer to keep the nucleotide level (ASV) and for recent eDNA when possible, we cluster the ASV and applied a LULU curation as recommend in (Brandt et al. 2021).

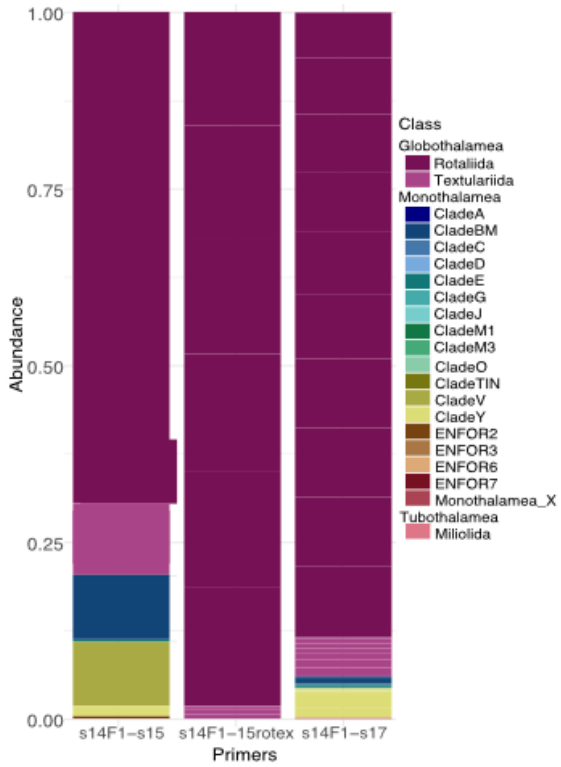

Figure 2.9: Distribution of foraminifera depending on the three types of primers. The same samples were here amplified with the long and shorts primers. 14-15rotex is the one showing preferentially hard-shelled groups.

Chapter 3: Planktonic foraminifera eDNA signature deposited on the seafloor remains preserved after burial in marine sediments

Ines Barrenechea Angeles, Franck Lejzerowicz, Tristan Cordier, Janin Scheplitz, Michal Kucera, Daniel Ariztegui, Jan Pawlowski and Raphaël Morard

3.1. Project description

The sediments were collected by Raphaël Morard off the coast of Newfoundland where cold and warm currents flow and intermingle. We first separated the surface sediments from the bottom sediments. We wanted to compare the morphological assemblages and see if, despite their dissimilarity, the surface distribution patterns are preserved. To do this, Janin Sheplitz and I counted and identified planktonic foraminifera, and I performed metabarcoding. For the bottom samples, I processed selected cores with both approaches (morphology and metabarcoding).

In this published paper, we present the data that allowed us to observe a change in the ratio of planktonic to benthic DNA in the cores and to better understand how planktonic DNA is preserved on the seabed.

A modified version of this chapter is published in *Scientific reports 10*(1), 1-12, (2020)

3.2. Abstract

Environmental DNA (eDNA) metabarcoding of marine sediments has revealed large amounts of sequences assigned to planktonic taxa. How this planktonic eDNA is delivered on the seafloor and preserved in the sediment is not well understood. We address these questions by comparing metabarcoding and microfossil foraminifera assemblages in sediment cores taken off Newfoundland across a strong ecological gradient. We detected planktonic foraminifera eDNA down to 30 cm and observed that the planktonic/benthic amplicon ratio changed with depth. The relative proportion of planktonic foraminiferal amplicons remained low from the surface down to 10 cm, likely due to the presence of DNA from living benthic foraminifera. Below 10 cm, the relative proportion of planktonic foraminifera amplicons rocketed, likely reflecting the higher proportion of planktonic eDNA in the DNA burial flux. In addition, the microfossil and metabarcoding assemblages showed a congruent pattern indicating that planktonic foraminifera eDNA is deposited without substantial lateral advection and preserves regional biogeographical patterns, indicating deposition by a similar mechanism as the foraminiferal shells. Our study shows that the planktonic eDNA preserved in marine sediments has the potential to record climatic and biotic changes in the pelagic community with the same spatial and temporal resolution as microfossils.

3.3. Introduction

Analyses of ancient DNA preserved in various archives have transformed our understanding of the evolution of species and ecosystems. Whilst earlier studies have concentrated on DNA extracted from taxonomically constrained samples (such as bones or frozen tissue), advances in high-throughput sequencing and bioinformatics now allow the analysis of ancient DNA extracted from sedimentary archives (Pedersen et al. 2015), so called sedaDNA. The accumulation and preservation of sedaDNA buried in land and lake sediments have been subject to active research and interpretation (Parducci et al. 2017). However, studying the deposition of DNA on the ocean floor and its preservation in marine sediments is more complex because the DNA has to travel through a water column of several kilometers(Armbrecht et al. 2019). Unlike in the terrestrial environment, with pervasive transport of subfossil biomass from land, the largest portion of the marine sedaDNA is derived from planktonic community, which is dominated by microbes and protists (Vargas et al. 2015). After the death of the surface plankton,

its DNA is subject to a transport through the water column, during which much of the associated organic matter is known to be consumed and respired (Iversen and Ploug 2010). This transport could take between 3 to 12 days depending on the size and morphology of test (Takahashi and Be 1984). However, it remains unclear how exactly the planktonic eDNA, defined as the total DNA present in the environment after (De Schepper et al. 2019), survives this transport, whether the degradation or transport are associated with sorting or lateral advection, and finally, whether the eDNA arriving at the seafloor is preserved in marine sediments without further distortion of its composition.

Despite the long exposure to degradation under oxic conditions during transport in the water column, and substantially lower concentration of organic matter on the seafloor, there is evidence that planktonic eDNA is preserved in marine sediments and contains exploitable ecological signal (Briggs 2020). Earlier studies have shown sedaDNA preservation in marine sediments deposited under anoxia with unusually high amounts of organic matter preserved (Morard et al. 2017), but later investigations indicate that sedaDNA can also be extracted from normal marine sediments, dominated by clastic or biogenic mineral fractions (Corinaldesi et al. 2007, 2008, 2011). In addition, the low temperature of deep-sea water (0 to 4 °C) ensures a good preservation of sedaDNA (De Schepper et al. 2019; Briggs 2020). Using planktonic foraminifera as a "Rosetta Stone", allowing benchmarking of sedaDNA signatures by co-occurring fossil tests of these organisms, 9[9] showed that the fingerprint of plankton eDNA arriving on the seafloor preserves the ecological signature of these organisms at a large geographic scale. This indicates that planktonic community eDNA is deposited onto the seafloor below, together with aggregates, skeletons and other sinking planktonic material. If this is true, sedaDNA should be able to record signatures of surface ocean hydrography, affecting the composition of plankton communities, with the same spatial resolution as the skeletal remains of the plankton. In addition, if the plankton eDNA is arriving on the seafloor in association with aggregates or shells, it is possible that it withstands the transport through the water column by fixation onto mineral surfaces. The same mechanism has been proposed to explain the preservation of sedaDNA in sediments (Corinaldesi et al. 2007, 2008, 2011), implying that the flux of planktonic eDNA encapsulated in calcite test arriving on the seafloor is conditioned for preservation upon burial.

Planktonic foraminifera sedaDNA is an ideal proxy both “horizontally” to assess the spatial resolution of reconstructing past surface ocean hydrographic features and “vertically”, to unambiguously track the burial of its signal throughout the sediment column. Indeed, the flux of planktonic foraminifera eDNA should be proportionate to the flux of dead foraminiferal shells sinking to the seafloor, allowing independent benchmarking of the eDNA signal. eDNA is powerful tool to study ecosystem because it does not require direct taxonomic knowledge thus allowing to gather information on every organism present in a sample. However, assignment of the eDNA sequences to known organisms is done via comparison with reference sequences (or barcodes) made available in public repositories or curated databases(Guillou et al. 2013). The taxonomy of planktonic foraminifera is well understood(Schiebel and Hemleben 2017) and barcodes exist allowing almost complete mapping of eDNA amplicons on the taxonomy based on foraminiferal test morphology (Morard et al. 2015, 2019). Importantly, the composition of planktonic foraminifera communities is closely linked to surface hydrography and this signal is preserved by fossil tests deposited on the seafloor(Rutherford et al. 1999; Siccha and Kucera 2017). Since foraminiferal eDNA accumulated in the ocean sediment can be recovered, it could be used to analyze changes in planktonic and benthic communities over time (Lejzerowicz et al. 2013b; Pawlowska et al. 2014, 2016; Szczucinski et al. 2016).

Here we take advantage of the planktonic foraminifera as a model system to investigate how planktonic foraminifera eDNA accumulates in sediments and to what extent the regional hydrographic features affecting the plankton are reflected in sedaDNA. We analyzed microfossil and molecular planktonic foraminifera assemblages in a series of short sediment cores, collected in the northwestern Atlantic, around the Grands Banks of Newfoundland, at the confluence of the cold Labrador and warm Gulf Stream Currents (Fig. 3.1). The southward Labrador Current (LC) exports cold and low salinity water from the Arctic Ocean in contrast to the northward Gulf Stream (GS) and North Atlantic Current (NAC) that bring warm and saline waters from the subtropical areas (Drinkwater 1996), creating a steep and seasonally stable ecological gradient. The strength of the LC has influenced changes in climate during the Holocene by slowing down the Atlantic Meridional Overturning Circulation (AMOC) (Sicre et al. 2014; Sheldon et al. 2015, 2016) which exports warm water masses to the east Atlantic.

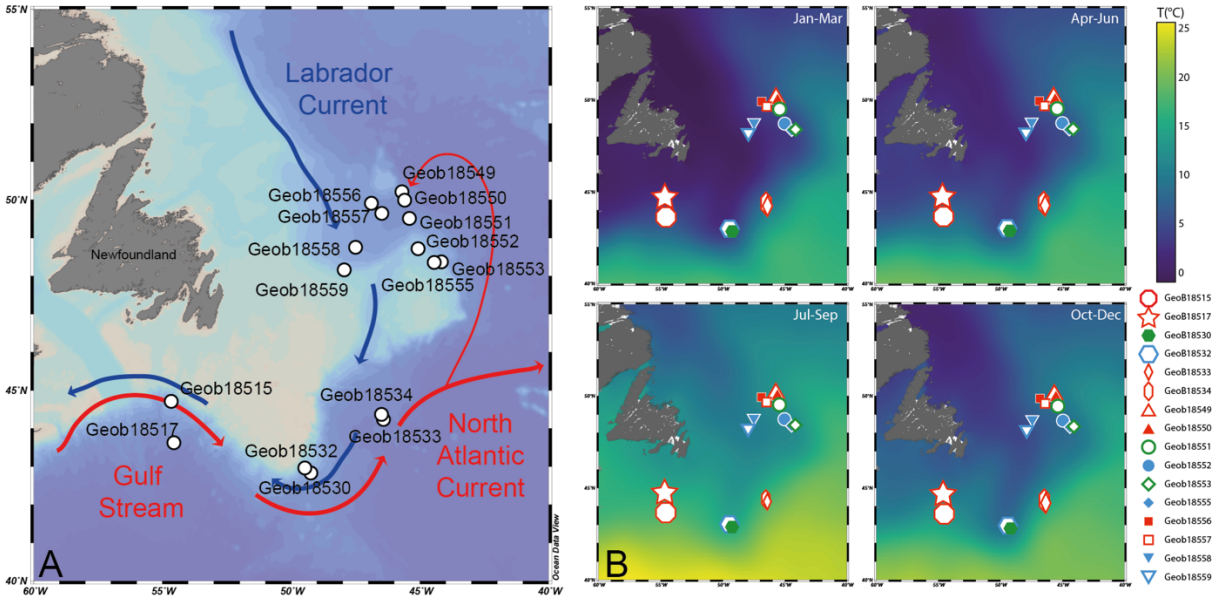

Figure 3.1: A. Location of the analyzed cores. Red ascending arrows represent the Gulf Stream (GS) and North Atlantic Current (NAC), blue descending arrows the Labrador Current (LC). B. Seasonal variability in sea-surface temperature observed in the sampling region. The data were extracted from the WOD13 database (Locarnini et al. 2013)*. Open symbols represent locations where the cores were analyzed for micropaleontological and metabarcoding analysis, filled symbols indicate locations where only the core top was used for micropaleontological analysis.*

3.4. Material and Methods

3.4.1. Sampling

The sediment samples were collected along the continental shelf off Newfoundland with multicorer (MUC) at 16 locations (Fig. 3.1) during the MSM39 cruise(Mulitza et al. 2015). The MUC allows the recovery of undisturbed sediment surface and the underlying 3-5 decimeters of the sediment. At each station, one small (Ø 5.6 cm) core was available for eDNA and micropaleontological sampling. The core was first gently pushed through the liner until the surface layer was at the level of the liner. The overlying water was removed using a clean syringe and the surface sediment was sampled with a sterile spatula and isolated into 1.5 mL tubes. Five replicates were taken for each MUC sampled. After the surface layer was sampled, the core was extruded in steps of 2-3 cm, the layer reaching above the liner was cut on the side and opened without touching the center. A sample from the pristine center of the slice was taken for DNA analyses with three to five replicates taken per layer using each time a different sterile spatula. The rest of the slice was used for micropaleontological analyses. The procedure was

repeated until the bottom of the core was reached. The sediment samples for DNA analyses were frozen at -80°C after the end of the collection, which took less than half an hour per core. After each sampling the spatulas were rinsed clean, left for at least 30 minutes in a bath with diluted H_2O_2 and finally flamed using 100% ethanol before the next collection.

3.4.2. Microfossil analysis

To assess the response of recent foraminifera community to the regional oceanographic settings, we selected 15 of the 16 locations for micropaleontological analyses of the core top samples. We subsampled the surface sediments and sieved them through 500, 150 and 63 µm size fractions. The fraction 150-500 µm was analyzed for census count of planktonic foraminifera of the surface samples. This size fraction was chosen to facilitate comparison with global datasets on planktonic foraminifera morphospecies abundances in seafloor sediments(Siccha and Kucera 2017). Based on the assemblage composition observed in the core tops (see results), we selected ten cores representative of the assemblage variability across the ecological gradient to perform a comparison between morphological and molecular assemblages buried in the sediment. In the selected cores, we processed every available layer until the deepest sampled sediment (42 cm in core GeoB18553). For the 10 cores selected for comparison between microfossil and metabarcodes analyses, we counted foraminifera in the fraction > 63 µm to capture compositional changes among small specimens and small morphospecies, in order to provide data more comparable to the bulk metabarcoding datasets (Morard et al. 2019).

Both sets of samples, the sediment was washed with freshwater and the residue was dried overnight at 40°C. The dried samples were weighted and split using a microsplitter to obtain a representative aliquot containing 200-300 foraminifera for census counts. Planktonic foraminifera tests were identified and counted to morphospecies level following 14 and benthic foraminifera were counted as well. Metadata regarding the collection, wet and dry sediment weight processed, and census counts are reported in Table S1.

3.4.3. eDNA extraction

We extracted the eDNA from all slices down to the 10 cm, from the 18-20 cm and 28-30 cm layers and in the longer core GeoB18553-3 also the layer 38-40 cm. In addition, in the cores GeoB18532-2 and GeoB18549-2, all available layers were processed, covering over 30 cm in both cores. We extracted between three to four replicates for the surface layers and between

one and two replicates for the down core samples, resulting in a total of 167 sediment samples for metabarcoding analyses from 84 layers of the 10 cores. DNA extractions were carried out in batches of nine samples and one empty vial acting as a negative control in a clean dedicated sedaDNA room at the University of Geneva. Approximately 0.5 g of sediment of each subsample was extracted following the protocol of PowerSoil® DNA Isolation Kit (MoBio) and the resulting DNA extracts were quantified with a fluorimeter Qubit™ (Invitrogen) and are reported in Table S2. Gloves and disposable spatulas were changed, and all surfaces were cleaned with bleach and RNase AWAY® solutions between each sample and quantifications as well as other steps performed in high molecular load environments were done using single-use aliquots.

3.4.4. PCR amplification and high-throughput sequencing

To enrich the foraminiferal signal in the DNA extractions, we carried out a PCR, making use of the relatively short and highly specific hypervariable region 37f (68-196 bp) (Pawlowski and Lecroq 2010). The region was targeted using a combination of primers s14F1 (forward) and s15 (reverse), as previously described (Lejzerowicz et al. 2014a). The primers were tagged with a unique combination of eight nucleotide identifiers attributed to each sample, allowing bioinformatic demultiplexing of the amplicons to their sample of origin. The tags' combinations correspond to different nucleotides attached to the primers which allow a multiplex of samples and were designed according to a Latin square matrix(Esling et al. 2015b). To ensure a better PCR yield from the DNA extracts, we opted for the use of bovine serum albumin (BSA) (Thermo Scientific™) and a polymerase from AmpliTaq Gold™ 360 Master Mix containing low-detection, heat-activated polymerase. The PCR mix contained: 15 µl of AmpliTaq Gold™ 360 Master Mix, 2 µl of BSA, 3 µl of combined primers at 0.2 M each, 10 ng of extracted DNA for samples having enough DNA, otherwise we added a maximum of extracted DNA as possible for those with low concentration, and H_2O to complete 30 µl. To avoid contamination after preparing this mix, extracted DNA was added in a dedicated hood located in a separate room. For each PCR session 21 samples in duplicate together with 7 controls were processed, including 5 PCR blanks to ensure that both primers and PCR mix were clean and 2 extraction blanks to monitor DNA samples contamination in the extraction room. The PCR reaction was performed as follows: pre-denaturation at 94°C during 1 min, then 60 cycles of denaturation at 94°C for 30 sec, annealing at 52°C for 30 sec and extension at 72°C for 30 sec, subsequently a final extension at 72°C for 2 min. Aliquots of the PCR products were migrated on a 1.5%

agarose gel for 20 minutes at 100 Volts, and quantified with Qubit™ fluorimeter (Invitrogen). The PCR products were pooled in equimolar mix with each duplicate located in a different pool to reach a total quantity of 100 ng of DNA. Each pool was purified using a High Pure PCR Cleanup Micro Kit (Roche) and quantified using the Qubit. One library was prepared for each pool following the instructions of the Illumina Truseq PCR-free® Library Preparation kit. The resulting eight libraries were quantified by qPCR using KAPA Illumina Library Quantification and diluted to a final concentration of 4nM. The diluted libraries were then pooled equimolar and sequenced on an Illumina Miseq® system. The raw sequence data can be downloaded from the European Nucleotide Archive under BioProject PRJNA668798

3.4.5. Sequence data analysis

The raw sequences obtained from the libraries were processed as in (Lejzerowicz et al. 2014a). The paired-end read pairs were quality-filtered by keeping only pairs having a mean quality score (phred score) above 30 and assembled with a minimum of 12 bases overlapping without mismatch. Those sequences were then demultiplexed based on the sequenced inline primer sequences, allowing a maximum of 2 mismatches to the reference tagged primer combinations. Chimeras were identified and removed with UCHIME 4.2(Edgar et al. 2011). The remaining sequences were further de-replicated to generate Individual Sequence Units (ISUs) as in [47], each ISU was aligned using the Needleman-Wunsch (NW) algorithm against a multiple sequence alignment of foraminiferal species and assigned to the consensus taxonomy of the sequences having the highest sequence identity level. ISUs without any alignment above 80% sequence identity with the reference database (See below) were left unassigned. To form OTUs, ISUs were pre-clustered based on their short 5'-end 37F hypervariable signatures (resolution described in Lecroq et al. 2010) and OTUs were delineated by average linkage clustering based on pairwise NW alignments distances and using thresholds defined for each pre-cluster based on the taxonomy of the ISUs to cluster.

In order to assign a taxonomy to OTUs, we assembled a custom reference multiple sequence alignment including planktonic and benthic taxa. We used the PFR² v.1.0 database (Morard et al. 2015) that includes planktonic foraminifera sequences only and added reference sequences of small planktonic foraminifera (Morard et al. 2019) that were published after the release of PFR². We merged the planktonic foraminifera reference sequences with those of benthic foraminifera species (Pawlowski and Holzmann 2014) coming from NCBI GenBank. The

taxonomy was structured into a 6-level hierarchical path that included the relevant level of the foraminifera taxonomy starting from the superorders (Pawlowski et al. 2013) until the genetic types for planktonic foraminifera (Darling and Wade 2008; Morard et al. 2016). The resulting alignment was trimmed to cover only the 37f region.

The OTUs were assigned using the assignment-table-vsearch module of the SLIM v0.4 web-application(Dufresne et al. 2019) at 95% of similarity against the local reference multiple sequence alignment. Because of the high specificity of the selected region for foraminifera, all the recovered OTUs can be considered as derived from foraminifera(Lecroq et al. 2011). OTUs that could not be assigned to the (almost) complete planktonic reference database was considered benthic. ISUs attributed to planktonic foraminifera by the assignment method were manually checked. Representative sequences of each planktonic foraminifera OTU were aligned with a selection of planktonic foraminifera reference sequences and only OTUs with clear similarity to the reference sequences were retained. The remaining ISUs were considered as benthic OTUs lacking close references in the multiple sequence alignment to be attributed with certainty. The result of the assignment of the ISUs are provided in Table S3.

The difference in number of reads recovered between libraries was normalized using the cumulative sum scaling method available on the metagenome-Seq Bioconductor package(Paulson et al. 2013b) in R(R Development Core Team 2014). The cumulative sum scaling corrects the biases induced by differential sequencing depths and uses a zero-inflated Gaussian distribution mixture model that accounts for technical zero values resulting from under-sampling. In order to compare the community composition in both morphological and molecular data, we used the *betadisper* function in the *vegan* package(Oksanen et al. 2007).

3.5. Results

3.5.1. Microfossil data

Sediment samples were collected along the continental shelf off Newfoundland with multicorer (MUC) at 16 locations (Fig. 3.1) that allowed the recovery of short sedimentary cores of 24 to 42 centimeters (Fig. 3.2). The cores were sampled to carry out micropaleontological and metabarcoding analysis in parallel. We carried out first a census count of the size fraction 150-500 μm that is typical for micropaleontological analyses of 15 selected core-tops. Based on the

obtained results, we carried out a census count of the samples of 10 cores representative of the compositional diversity at a size fraction in the fraction > 63 µm to capture compositional changes among small specimens and small species. In all samples, 23 morphospecies of planktonic foraminifera could be identified, representing a mixture of polar, transitional and rare sub-tropical species. The five most common morphospecies were *Neogloboquadrina pachyderma* (polar), *Globigerina bulloides* (transitional), *Neogloboquadrina incompta* (transitional), *Turborotalita quinqueloba* (subpolar) and *Globorotalia inflata* (temperate) accounting for ~97 % (83-100%) of the assemblages (Fig. 3.2). We also encountered the microperforate morphospecies *Globigerinita glutinata*, *Globigerinita uvula*, *Tenuitella fleisheri* and *Globigerinita minuta* that accounted for ~ 2 % of the assemblages. Finally, we observed rare occurrences of morphospecies which are normally encountered in temperate to subtropical areas: *Globigerinoides ruber albus*, *Globorotalia hirsuta*, *Globorotalia menardii*, *Globigerina falconensis*, *Pulleniatina obliquiloculata*, *Trilobatus sacculifer*, *Globigerinoides conglobatus*, *Globigerinoides elongatus*, *Globigerinella calida*, *Globorotalia truncatulinoides*, *Neogloboquadrina dutertrei*, *Dentigloborotalia anfracta*, *Globigerinoides ruber ruber* and *Globigerinella siphonifera* that accounted for ~1% of the assemblages. These morphospecies were grouped into a category named "WARM" as they likely represent the advection of non-resident foraminifera by the Gulf Stream into the local community.

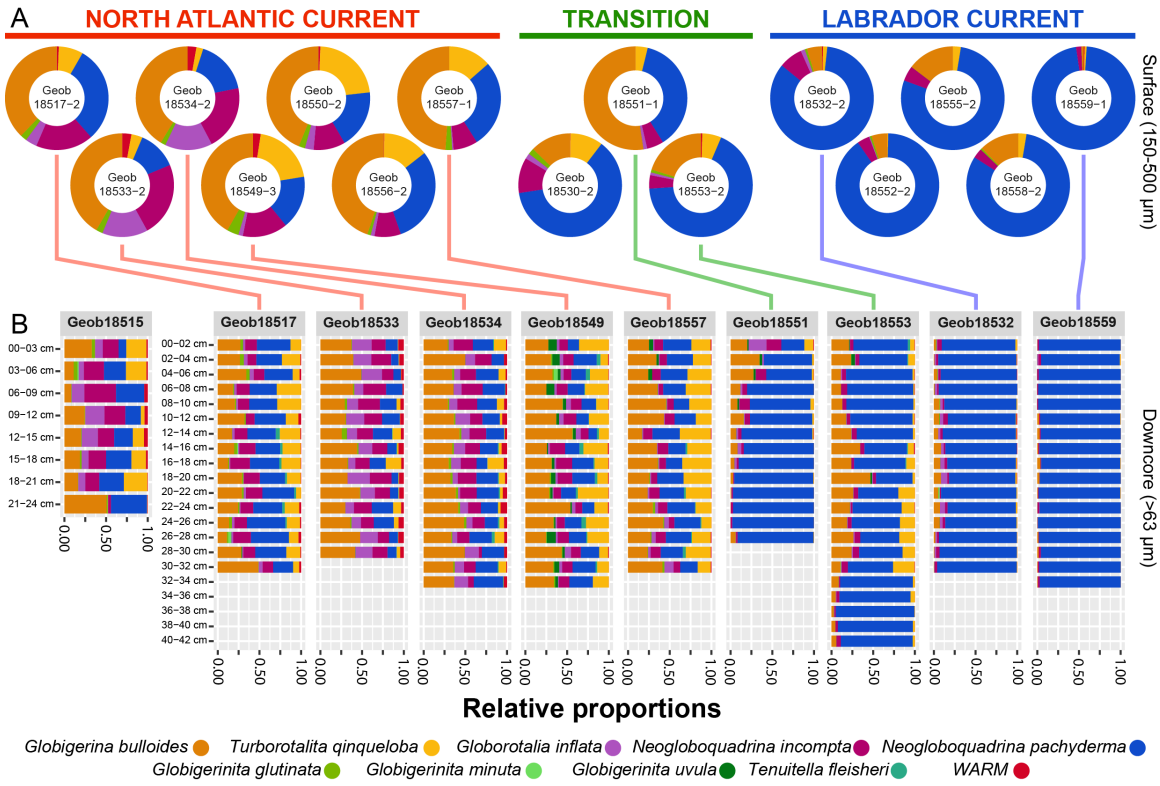

Figure 3.2: Census counts of planktonic foraminifera in the studied sediment cores (Fig. 1). A. Counts performed in the surface sediments in the size fraction 150-500 µm at 15 locations. B. Counts performed on the 10 cores selected for parallel census count and metabarcoding analysis. The cores are grouped according to the faunal signature of the regional hydrography preserved in the sediment.

Sedimentary assemblages record the westward retroflection of the NAC into the Labrador Sea that brings temperate and tropical species to the north of the Labrador Current (Fig. 3.1). Analysis of tests census counts in the surface samples of 15 sites (150-500 µm) that includes recent or subrecent specimens revealed that the planktonic foraminifera community at the sediment surface records the steep oceanographic gradient observed in the region (Figs. 3.2A, 3A). Indeed, we observed a strong polarization of the assemblages mostly driven by the relative proportion of *N. pachyderma,* a marker for cold temperatures, opposed to the transitional morphospecies *G. bulloides*, *T. quinqueloba, G. inflata, G. glutinata, N. incompta* and the WARM group. This compositional gradient largely follows the frontal zone between the NAC and the LC, including the counter-intuitive inversion of the temperature polarity in the Labrador Sea with samples collected to the South of Newfoundland having a "cold" signature (GeoB18530, GeoB18532) whilst samples collected Northeast off Newfoundland having a "warm" taxonomic composition (GeoB18549, GeoB18550, GeoB18551, GeoB18556, GeoB18557). This structure was largely conserved in the downcore analysis of the samples sieved at 63 µm (Figs 3.2B, 3.3B, 3.3C). At sites GeoB18551 and GeoB18553 we observe higher compositional variability downcore (Fig. 3.3 B), indicative of an episodically higher contribution of the "warm" species to these assemblages in the past, likely reflecting a shift of the front between the LSW and NAC influence in the past (Fig. 3.1). Surface and down-core sediments, and from both size fractions, could hence be classified into three zones reflecting a) polar conditions dominated by *N. pachyderma*, (LC zone) b) a mixed assemblage with higher proportion of the transitional species (transition zone) and c) stronger advection of species of the WARM group due to higher influence of the NAC (NAC zone, Fig. 3.2A).

3.5.1. Metabarcoding data

We successfully amplified foraminifera metabarcodes from 167 samples. The total DNA extracted from sediment ranged from below the instrument detection limit to ~9 $ng.\mu l^{-1}$, and showed an overall decrease in concentration with depth (Fig. 3.4A). This decrease was mirrored

by decreasing number of positive amplifications of foraminiferal DNA with increasing depth in the cores.

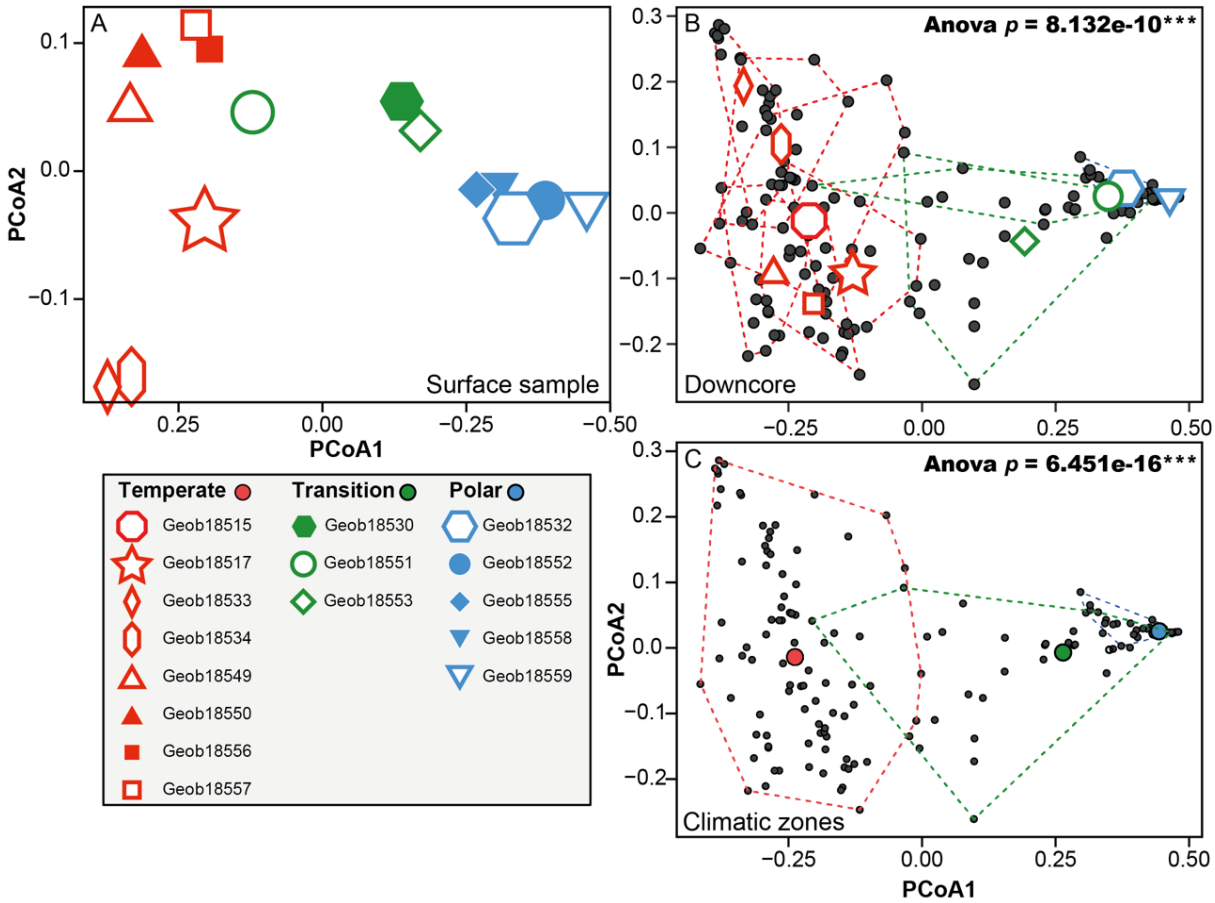

Figure 3.3: Principal coordinates analysis of the census counts of planktonic foraminifera carried out among A. all surface samples B. all downcore samples with each site treated as a group and C. all samples grouped by their hydrographic signature. The p value of the associated ANOVA is shown for B and C only because there is only a single surface sample per locality in A.

In total, 15,460,098 raw DNA amplicon reads were obtained from the 334 sequenced PCR products (2 replicates per samples) obtained from 167 samples. 8,306,918 reads were retained after quality filtering and clustered into 2,467 OTUs. Of these, 1,025 OTUs representing 7,908,878 reads occurred in more than three samples and were retained for subsequent analyses. From those, 10 OTUs accounting for 968,487 reads belonged to planktonic foraminifera and were assigned to the morphospecies *Neogloboquadrina pachyderma*, *Neogloboquadrina incompta*, *Globorotalia inflata*, *Globorotalia hirsuta*, *Globigerinita glutinata*, *Globigerinita uvula* and *Tenuitella fleisheri*. Three and two OTUs were attributed to the morphospecies G. uvula and N. pachyderma respectively but were considered as intragenomic variants. No

sequences of the two spinose planktonic foraminifera commonly encountered in the census counts *Globigerina bulloides* and *Turborotalita quinqueloba* were detected.

The ratio of planktonic to benthic foraminiferal reads changed with depth in the sediment. The relative proportion of planktonic reads was 3.5 % on average in the top 10 cm of the cores (with rare outliers) but increased between 10 and 20 cm up to 96 % of the total assemblages. Then, it dropped below 30 % between 20 and 30 cm but remained higher than for the top 10 cm (Fig. 3.4B).

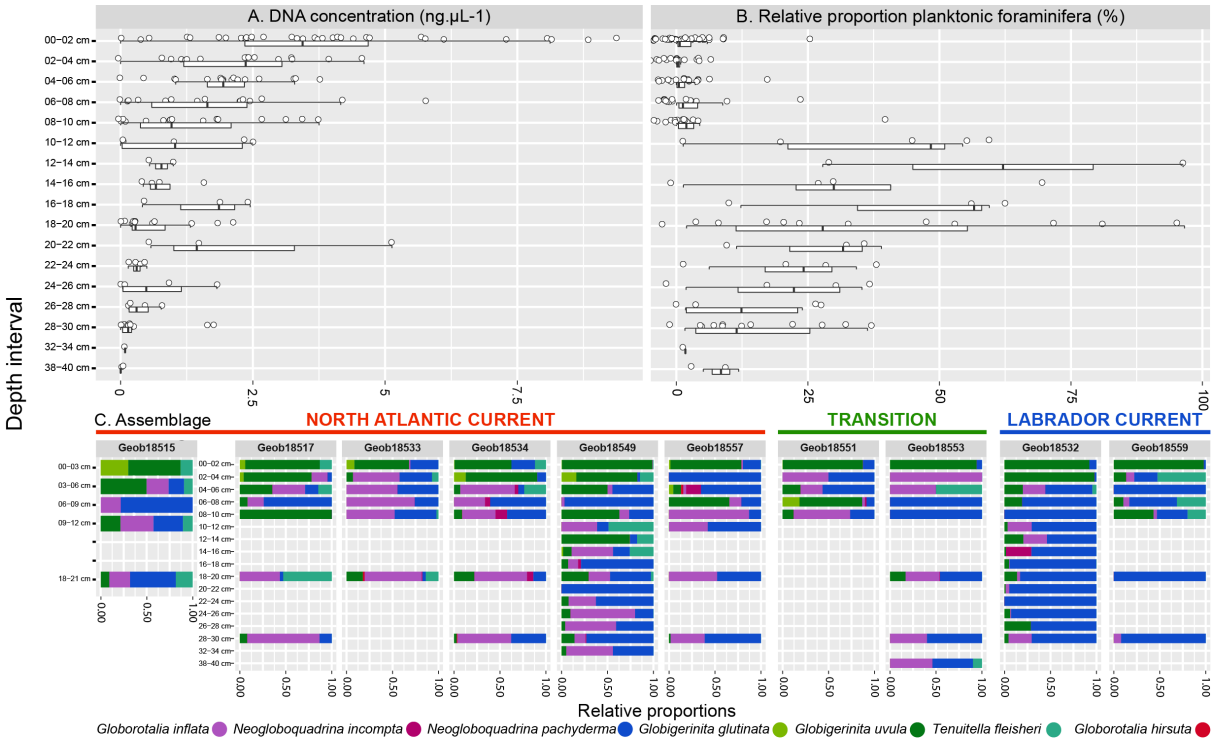

Figure 3.4: Results of metabarcoding analyses. A. box plot and jitter plot showing the DNA concentration in the sediment extract at all depths. B. Proportions of planktonic and benthic foraminifera reads observed in the metabarcodes. C. Relative proportions of reads assigned to individual species of planktonic foraminifera detected in the metabarcoding dataset.

In terms of the taxonomic composition, metabarcodes from the first four centimeters in all cores were dominated by reads assigned to the microperforate morphospecies *G. glutinata* and *G. uvula* (Fig. 3.4C). Further downcore, reads from these morphospecies become rare and the assemblage was dominated by *N. pachyderma* and *G. inflata.* Like in the micropaleontological samples, molecular assemblages of planktonic foraminifera in core GeoB18549 revealed substantial downcore variability, which was also present but less clearly developed in molecular assemblages in core GeoB18532 (Fig. 3.4C). Comparative analysis of microfossil and

metabarcoding community showed an overall conservation of the same biogeographic patterns but with a larger degree of overlap in the metabarcoding data (Fig. 3.5). The significant difference observed between sites with microfossils (Fig. 3.5A1) is not conserved in molecular assemblages (Fig. 3.5A2). When comparing the composition of assemblages under the temperate (NAC), transition and polar (LC) currents, the significant difference is also not conserved overall among the metabarcodes, but the comparison of the sites under the temperate and polar regimes displayed a significant difference both in microfossil and metabarcoding data (Figs 3.5B1, 3.5B2). To ensure that this observation was not only due to sampling size, we limited the comparison between the strength of the separation of the microfossils and metabarcoding datasets to the cores GeoB18532 (polar) and GeoB18549 (temperate) (Fig. 3.5C). Despite the presence of an overlap in the metabarcode compositional data (Fig. 3.5C2), compared to the complete exclusion of the assemblages of the two cores observed in the microfossil assemblages (Fig 3.5C1), the comparison of the distribution of the samples of the metabarcoding dataset returns a significant difference (p-value=0.0013) indicating that the two cores have distinct faunas.

3.6. Discussion

3.6.1. Planktonic foraminifera eDNA preservation in the sediment

The observed decrease in total DNA concentration with sediment depth in all cores is consistent with a model of progressive degradation of sedaDNA over time (Fig. 3.4A). In this model, the higher concentrations at the surface would reflect the combination of the flux of eDNA from the plankton and DNA from living, sediment-dwelling organisms, including the benthic foraminifera, whose population density decreases with sediment depth. As a result, eDNA from living benthic foraminifera dominates the pool of foraminiferal sedaDNA in the top 10 cm of the sediment (Fig. 3.4). This 10 cm limit is entirely consistent with direct observations of depth range of benthic foraminifera in North Atlantic sediments identified as living by vital stains (Corliss and Emerson 1990), and by the depth of the bioturbated mixed zone in similar sediments (Loubere 1989). The absence of living benthic foraminifera below 10 cm could also explain the remarkable sharp increase of the relative proportion of planktonic foraminifera sedaDNA immediately below this level. In the zone inhabited by living benthic foraminifera, eDNA from the living organisms is pristine and abundant, and thus more prone to be amplified and sequenced than the planktonic sedaDNA fraction (Lejzerowicz et al. 2013b). Below the

inhabited zone, all foraminiferal DNA is environmental, and its amount reflects the flux of the involved populations. Since the flux of planktonic foraminifera is much higher than that of benthic foraminifera (as also seen in the concentration of their shells in the sediment with a ratio ~100:1), the composition of foraminiferal sedaDNA is skewed towards the plankton.

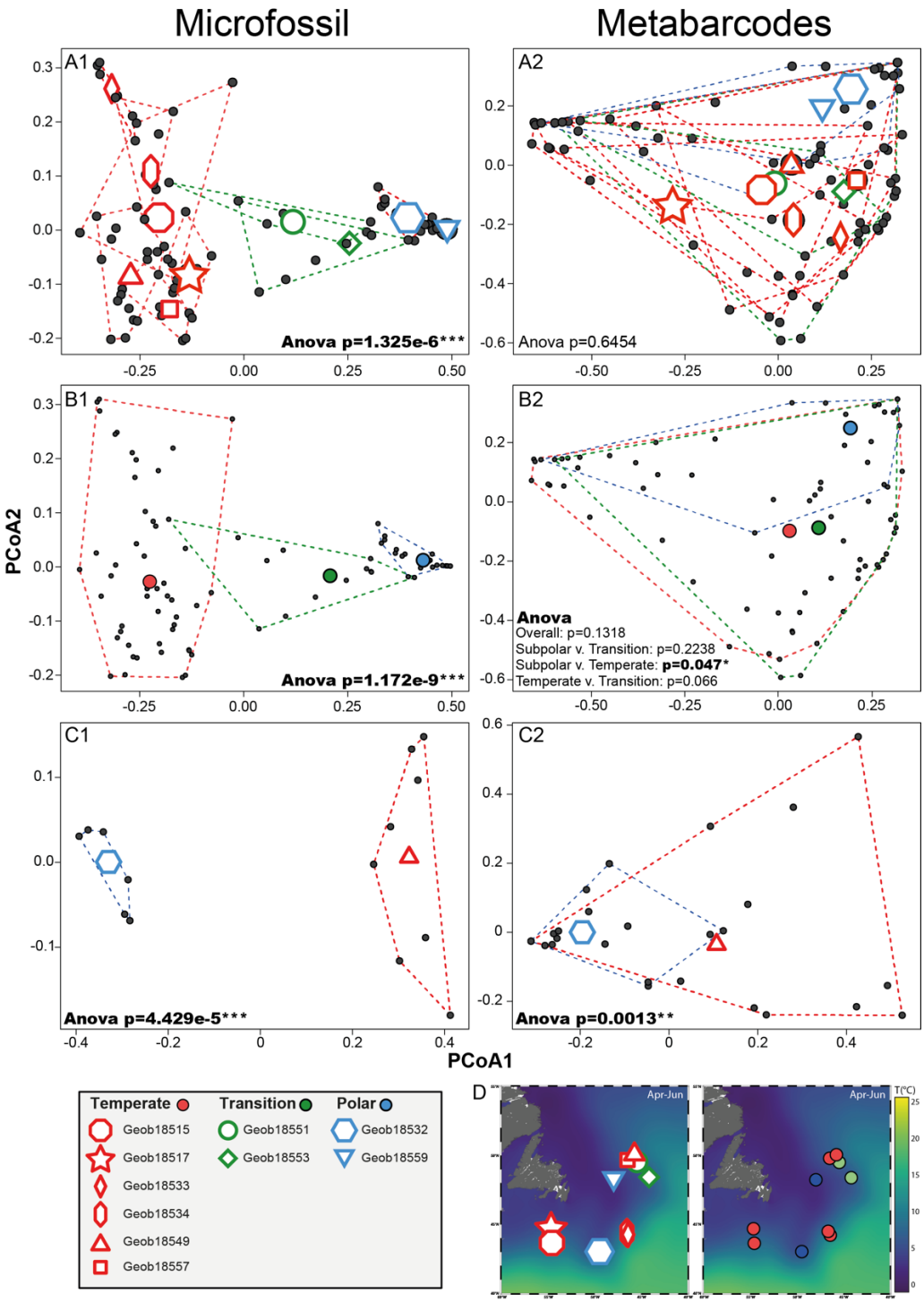

Figure 3.5: Comparison of principal coordinates analysis performed on the samples for which both morphological and metabarcoding analyses have been performed. The analyses have been performed

with each site as a group (A.), each hydrographic signature as a group (B.) and only for the cores GeoB18532 and GeoB18549 (C.) but without the layers between 0-4 cm that are dominated by Globigerinita uvula. The p value of the associated ANOVA is provided, and we provide each pairwise group for the (B2.) analysis to show that the Polar vs Temperate is significant. Maps showing the position of each core and the currents signature are provided (D.).

To explain the observed pattern of relative abundance of planktonic and benthic reads, we considered that the benthic foraminifera community had an absolute abundance of 99 (arbitrary unit) at the surface and follows a linear decrease with depth to reach 1 at 10 cm which is the bottom of the inhabited zone in the sediment. Next to the living benthic foraminifera DNA, we considered a pool of planktonic eDNA exported from the surface that follows an exponential decay:

$$N(t) = N_0 e^{-\lambda\tau}$$

where N_0 is the initial abundance of planktonic foraminifera eDNA, λ is the decay constant and τ the mean lifetime, all set to 1. Below the inhabited zone (10 cm), we considered that there are no more living benthic foraminifera, and therefore its decay profile follows the same formula as for the planktonic foraminifera eDNA. However, we considered a lower decay constant for the benthic foraminifera DNA set to 0.9 because unlike the plankton, the decay profile of benthic DNA does not have to pass through the water column and is buried in the sediment directly with their cells. The resulting modeled relative proportions are shown on Fig. 3.6, indicating that the invoked process can at least theoretically reproduce the observed pattern.

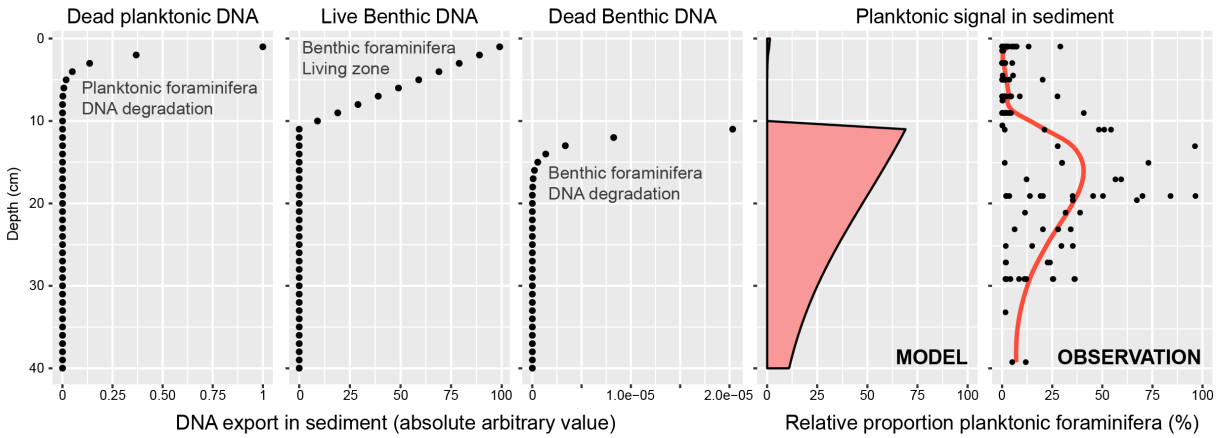

Figure 3.6: A conceptual model explaining the process of conservation of planktonic and benthic DNA with burial in the sediment. The flux of planktonic eDNA starts to degrade directly after the delivery

onto the seafloor and is in competition with the DNA from living benthic foraminifera inhabiting the first 10 cm of the sediment. Below the living depth of benthic foraminifera, only their eDNA is preserved and therefore the relative abundance of planktonic foraminifera amplicons increases. The benthic DNA degrades at a lower rate than the planktonic DNA that leads to a decrease of relative abundance of planktonic DNA as observed in Fig 2B. See discussion for detailed explanation of the model.

In the deepest samples, the sedaDNA concentrations are low, in some samples even below the detection limit of the Qubit fluorometer used to quantify the DNA extract (Fig 3.4A). However, the slope of both concentration and the relative abundance of benthic and planktonic eDNA appear to stabilize (Fig 3.4B). The Holocene sedimentation rates on the continental slope off Newfoundland, including at some of the same locations where the studied cores were taken, vary between 15 and 25 cm/kyr (Leng et al. 2018), indicating that the age of the oldest layers of the analyzed cores is likely > 2000 years. The consistent low concentration and consistent benthic/planktonic ratio in the oldest sedaDNA indicate that below the zone inhabited by the benthic foraminifera, there is no, or only limited, leakage of genomic DNA downcore, as shown already by previous studies (Torti et al. 2015, 2018). Since we were able to recover a taxonomically consistent foraminiferal eDNA signature even in the samples with the lowest overall DNA concentration, a substantial part of this DNA must be present in amplifiable strands of at least 100 to 200 nucleotide length (microbarcode and flanking conserved region length). These results agree with previous studies showing good preservation of foraminiferal eDNA in a variety of marine sediments (Lejzerowicz et al. 2013b; Pawlowska et al. 2014, 2016). The observation of good preservation of longer DNA fragments is also consistent with studies of Pleistocene (older than 12,000 years) sedaDNA in marine sediments, which were based on PCR and sequencing of long barcodes (Coolen et al. 2013; Kirkpatrick et al. 2016; De Schepper et al. 2019)

3.6.2. Microfossil vs metabarcoding record

The morphospecies composition and relative abundances recorded in the microfossils, both at the surface and downcore, reflect fauna typical for the region (Siccha and Kucera 2017), mirroring in its composition remarkably well the steep hydrographic gradient along the continental margin off Newfoundland (Fig. 3.2). The metabarcoding community shows a similar geographical structure (Figs. 3.4, 3.5) and contains only taxa that were identified among the microfossils. This indicates that the sedaDNA both at the surface and downcore does not

appear contaminated by long-range lateral transport and that no contamination likely occurred during laboratory handling of the samples. However, compositional composition of the metabarcoding dataset differs from the census count dataset in two key aspects.

First, we observe systematic differences in the presence/abundance of some taxa that could be attributed to PCR-induced biases. The most striking is the absence in metabarcoding community of the spinose Globigerinidae (Weiner et al. 2016; Morard et al. 2017) such as *G. bulloides* and *T. quinqueloba* that are common among the microfossils in the studied material. This is due to a 1-3 nucleotide difference in the forward primer region between the spinose clade and other foraminifera. As a result, the recovered metabarcoding community contained no amplicons that could be assigned to spinose morphospecies. Considering the rapid evolution of the SSU rDNA gene in planktonic foraminifera (Darling et al. 1997), it is difficult to design universal planktonic foraminifera primers for the hypervariable region 37F, which we deemed was the ideal target for studying ancient DNA because of its short length. Other foraminiferal metabarcoding studies also showed a primer bias but were able to target all planktonic morphospecies (Morard et al. 2018, 2019) using a longer 18S fragment, that we judged likely more difficult to amplify in sedaDNA samples.

Another PCR-related bias concerns the difference in species abundances between metabarcoding and microfossil data. This type of biases is due to either preferential PCR amplification or differences in gene copy number among species(Weber and Pawlowski 2013; André et al. 2014). In our study this is best seen in the disproportionally high abundance of amplicons from the large *G. inflata* (Fig. 3.4), where we speculate that the species may have more gene copies per cell, since nothing indicates that the species should be preferentially amplified. Other than PCR biases or differential gene copy number see no evidence for systematic bias in sedaDNA preservation due to different sizes, shapes or thicknesses of the shells with which the DNA is likely associated. The most abundant amplicons were derived from OTUs assigned to the thick-walled medium-sized *N. pachyderma*, but many of the samples were dominated by amplicons from OTUs representing the thin-walled and tiny *G. uvula* and *T. fleisheri* (Fig 3.4). The second aspect in which the metabarcoding community composition differs substantially from the co-deposited microfossils is the high abundance of *G. uvula* in samples from the top 4 cm of the sediment (Fig. 3.4). We rule out primer and other

sequence-related biases, that would result in a systematic preferential enrichment of this particular taxon at all depths.

It is also unlikely that this morphospecies, which has a broad ecological and geographic distribution, was not recorded consistently in microfossils samples. The species is small, but its shells have been recovered at all depths, albeit at smaller quantities. Therefore, we retain the possibility that the high abundance of this species in surface metabarcoding datasets reflects a genuine signal. Indeed, it has been shown that the ongoing climate warming induces large latitudinal shifts in planktonic foraminifera communities globally (Jonkers et al. 2019), and higher abundance of *G. uvula* in recent plankton samples from the North Atlantic has been noted by two independent studies(Schiebel et al. 2017; Meilland et al. 2019). We speculate that a potential recent increase of *G. uvula* in the studied region due to ongoing global change could in theory indeed be first observed in the eDNA on the sea floor. This is because the eDNA delivered during the last few years is the "freshest" and should amplify preferentially over older eDNA at the sediment surface. The most recent aDNA is likely completely mixed on short time scales over a few cm, explaining the abundance of *G. uvula* reads down to about 4 cm. Given the sedimentation rates reported in the region(Leng et al. 2018), the top 4 cm of the studied sediments should represent the average deposition over ~ 200 years. This layer would be still numerically completely dominated by shells deposited before the recent plankton shift, but the aDNA in this layer would be biased towards the most recent signal. In this scenario, sedaDNA would act as a more sensitive recorder of recent changes in plankton flux than the fossil assemblage.

3.6.3. Spatio-temporal patterns in community composition

Besides the two key differences described above and ascribed to PCR biases and bias due to amplification of the signal of a recent community shift, the composition of the metabarcoding community shows similar patterns of spatial and temporal variability as recorded by than of the fossil assemblage. Due to the lack of amplicons from spinose morphospecies, the difference between "warm" assemblages influenced by North Atlantic Current and "cold" assemblages deposited underneath the Labrador Currents are expressed more strongly in microfossil than metabarcoding data, but the spatial ordination of the sediment samples (surface and downcore) is similar (Fig 3.5). The ANOVA performed on the microfossil assemblages returned highly significant values. The same analysis performed on the metabarcoding assemblages showed a

larger degree of overlap between sites, but the pairwise ANOVA comparing metabarcoding and microfossil composition was still significant, showing essentially the same spatial ordination of the communities, irrespective of whether represented by microfossil counts or metabarcodes.

Importantly, these spatial patterns clearly remain preserved during burial. This is shown by the analyses of microfossil and metabarcode compositions in downcore samples from cores GeoB18532 and GeoB18549, representing deposition under colder and warmer conditions (Figs 3.2, 3.4). Samples from both cores are clearly separated with respect to their microfossil assemblages and this separation remains preserved in the metabarcoding data, despite the lack of spinose amplicons (Fig. 3.4). Moreover, the downcore records also show similarities in the amount of compositional variance within each core between the microfossil and metabarcoding data. The microfossil community in the "cold" core GeoB18532 is strongly dominated by *N. pachyderma* (Fig. 3.2) and the metabarcoding data in this core is also dominated by this morphospecies (Fig. 3.4).

We selected the cores Geob18532 (Polar) and GeoB18549 (temperate) were selected for metabarcoding analysis on their entire length specifically because the microfossil assemblage polarity is opposite to the overall latitudinal gradient, with warmer fauna found to the North of Newfoundland, reflecting the NAC retroflection (Figs. 3.1, 3.2). The fact that the microfossil and metabarcoding communities both preserved the inverted regional features implies that both signals are congruent. Whereas foraminifera shells, made of calcite, sink to the seafloor within days, the DNA could, in theory, reside among the dissolved organic pool and its deposition could be delayed, allowing long-range mixing and advection. The fact that we observe highly congruent patterns, retrievable despite various sources of bias identified in the metabarcoding community, implies that the sedaDNA likely represents molecules delivered and preserved in association with the flux of planktonic foraminifera tests. Thus, sedaDNA preserved in marine sediments deposited under normal, oxic conditions, has the potential to record climatic and biotic changes in the pelagic community above the site of deposition with the same spatial and temporal resolution as microfossils. This conclusion certainly applies to the planktonic foraminifera but can be likely extended to other planktonic taxa producing skeletal remains as well as those whose remains are deposited in the form of aggregates.

3.7. Conclusion

The pool of sedaDNA can be considered as a reliable source of information that not only contains genomic remains, but also records paleoceanographic changes. We show that sedaDNA retains the fine scale regional oceanographic features, and we speculate that sedaDNA could detect rapid changes such as the recent oceanographic warming not yet recorded in the morphological samples. The same approach that we used for foraminifera could be applied to any group of plankton, even those that do not leave fossil remains and without requiring taxonomic expertise, provided that reference databases are complete enough to assign sedaDNA metabarcodes to known taxa. However, the metabarcoding approach has major pitfalls, since the PCR step can bias the community composition, and as it is shown in our case, omit an entire clade that is abundant in the samples. The careful design of primers is thus a prerequisite and when possible, a parallel validation with fossil content of the samples should be mandatory.

Supplementary Material is accessible online under: https://www.nature.com/articles/s41598-020-77179-8

Table S1. List, processing details and taxonomic counts of the selected samples of the study for morphologic analyses.

Table S2. List and processing details of the selected samples of the study for metabarcoding analyses.

Table S3. Taxonomic assignment of the ISU for the 334 PCR products

Chapter 4: Evolution of marine eukaryotes and terrestrial plants over 300'000 inferred from a marine sedimentary aDNA collected off Papua New Guinea

4.1. Project description:

A marine sedimentary core from the Bismarck Sea off Papua New Guinea stored for 14 years was provided by Luc Beaufort from the CEREGE. The core was previously dated and record more than 340 Ka, showing interval of warm and cold periods.
Two major challenges:

- Amplify ancient sediments DNA from a tropical core with a potential less preservation of DNA due to SST of 25°C or 30°C or UV radiation on tropics
- Work with a long-time stored core at 4°C, deal with contaminants during sampling and the storage.

Goal of this study were:

- Evaluate the possibility of the use of sedimentary cores stored in many repositories
- How far in the past an ancient eDNA can be recover in a good preservation
- Which organisms can be targeted.

I present here a compilation of the results of metabarcoding of foraminifera, radiolarians, eukaryotes, and plants. This project is not yet complete, and we plan to include morphological assemblages of foraminifera, radiolarians and coccoliths to support our findings with the sedaDNA approach.

In prep.

4.2. Abstract

Over the past 15 years, sedimentary ancient DNA (sedaDNA) has become a new proxy for paleoenvironmental studies that provide information on a wide range of fossilised and non-fossilised taxa. In general, the sedaDNA studies focus on temperate or polar regions, rather than on tropical environments as they are exposed to DNA damaging factors such as high temperature or UV radiation. Moreover, sediment samples for ancient eDNA studies are usually frozen immediately after collection. However, many marine sediment cores were collected years before the advent of palaeogenomics and are kept cold at 4°C in core repositories around the world. Here, we test the limits of sedaDNA preservation by analysing a core retrieved from the Bismarck Sea, off New Papua Guinea, where the mean annual temperature is about 29°C and stored at 4°C for 14 years. We analysed 20 samples at glacial to interglacial intervals identified by isotopic measures of $\partial^{18}O$ of benthic foraminifera and spanning the past 385'000 years. We applied a metabarcoding approach using specific 18S rRNA gene primers for eukaryotes, foraminifera, radiolarians and TrnL primers for vascular plants. The data obtained by using universal eukaryotic marker (V9) show that the samples are strongly contaminated by fungal and amoeba DNA, probably originating from collection or storage. However, the results obtained using specific foraminifera, radiolaria or plants markers showed the presence of typical tropical communities. Even if the number of DNA declines down the core, the patterns of successional changes in species communities of these taxonomic groups are well archived at least until 200 Ka. Our study indicates the decrease of terrestrial and marine plankton diversity since the latest glacial period. However, the preservation of sedaDNA in the refrigerated core was insufficient to follow the plankton and plant community evolution during the previous glacial and interglacial periods.

4.3. Introduction

Marine and lake cores are commonly used as archives of past climate and biodiversity (Tzedakis et al. 1997; Westerhold et al. 2020; Morlock et al. 2021). Usually, one half of the core is used for geochemical and micropaleontological studies and the other half remains intact and is stored as an archive at 4°C in temperature-controlled rooms. Thousands of sediment cores have been collected for several decades and are available in cores repositories.

The micropalaeontological records consist of the hard remains of organisms that may have fossilised. The most common microfossils are foraminifera, diatoms, radiolarians, and pollen. They are used to reconstruct the climate and biodiversity of the past. However, this approach only reflects the evolution of a fraction of biodiversity at a given time (Yasuhara et al. 2017; Frenzel 2019). The soft-bodied, non-fossilized species are excluded, thus missing a large proportion of living organisms from paleoecological studies. Another disadvantage of using only fossils is the partial compilation of past biodiversity when hard shells are dissolved or fragmented or too small and not considered.

To complement micropaleontological proxies, ancient sedimentary DNA (sedaDNA) has been used for more than 15 years, giving access to more complete overview of past biodiversity. The sedimentary DNA record often includes bacteria, viruses, archaea, and eukaryotes (soft and non-soft bodies) (Pedersen et al. 2015; Thomsen and Willerslev 2015; Capo et al. 2021). So far, climate and biodiversity changes have been assessed with this tool mainly in high latitudes. Temperate and especially polar regions offer ideal conditions for DNA preservation, such as low temperatures and little UV radiation. For example, the vegetation history of Scandinavia has been reconstructed using ancient DNA (Rijal et al. 2021; Nota et al. 2022), diatom assemblages in Siberia lakes (Dulias et al. 2017).

Few sedimentary DNA studies have been conducted in tropical sites, which are known to be the most faunistically and floristically diverse regions of the world. Seven studies have been conducted in tropical African and Mexican lakes or swamps. In Ethiopia, eukaryotes were targeted (Krueger et al. 2021), in Uganda prokaryotes and eukaryotes (Dommain et al. 2020) and plants (Boessenkool et al. 2014), in Kenya rotifers (Epp et al. 2010) and diatoms (Stoof-Leichsenring et al. 2012) and in Benin plants (Bremond et al. 2017). In Mexico prokaryotes and eukaryotes were sequenced (Moguel et al. 2021). To our knowledge, one tropical marine core has been used for sedaDNA studies in the Arabian sea (More et al. 2018). Protists communities' changes over 43ka due to variations of oxygen minimum zone (OMZ) were reported. There is also a reef core in Australia to know the communities inhabited in the coral reef the past 750 years (del Carmen Gomez Cabrera et al. 2019), which could be consider as sediment core too.

Those cores were collected for sedaDNA analysis, which was not the cases for cores with a geological analysis purpose. However, a recent study reported successful recovery of ancient metagenomes from a core after several years of storage (Selway et al. 2022).

Here, we conducted for the first time the paleogenomic analysis of a marine sediment core collected in a tropical area, in the Bismarck Sea, off the coast of New Papua-Guinea where the sea surface temperature (SST) is around 29°C. The core has been stored at 4°C for 14 years at the CEREGE core depot in Aix-en-Provence, France. Dating shows that it has recorded the last 385 000 years (Tachikawa et al. 2011). In addition, isotopic measurements of ∂18O from benthic foraminifera identified glacial to interglacial intervals. To evaluate whether sedaDNA can be recovered from such a long-stored core and to gain insight into the evolution of marine organism and plant diversity during interglacial and glacial cycles, we used metabarcoding approach targeting small fragments of ribosomal (18S rRNA) and chloroplast (TrnL) genes. We obtained the metabarcoding data for marine eukaryotes, radiolarians, foraminifera, and vascular plants and we analysed their evolution across the past 300'000 years.

4.4. Material and methods

4.4.1. Study area and the core

The core MD05-2920 was collected during the MD148 IMAGES XII PECTEN cruise in 2005. The core collection site is about 100 km off Papua New Guinea (PNG), in the Bismarck Sea, (2°51′48S, 144°32′04E) at a water depth of 1843 m (Beaufort et al. 2005) (Fig. 4.1). Two main rivers, Sepik and Ramu, after passing through highlands, lowlands and swamps discharge terrigenous sediments into the sea. Tachikawa et al. 2011 confirmed the terrestrial origin of the sediment particles and calculated the sedimentation rate, which varies between 6 and 13 cm/ka. The core was dated using ^{14}C AMS on planktonic foraminifera (G. ruber) and d1^{8}O on benthic foraminifera (C. wuellestorfi and U. peregrina). The 35 m long core recorded the last 384 kyr. Benthic foraminiferal $d^{18}O$ values permitted the identification of interglacial and glacial periods to corresponding to eleven Marine Isotopic Stage (MIS) (Fig. 4.2).

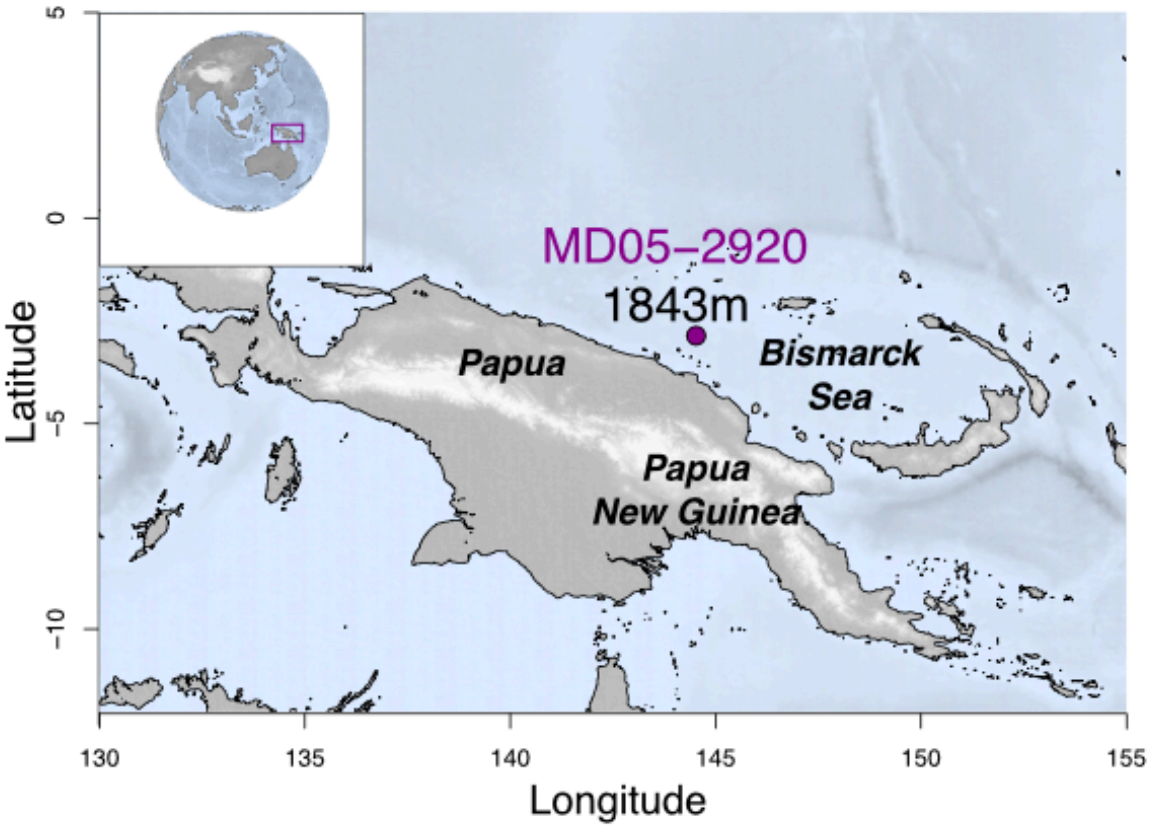

Figure 4.1: Core MD05-2920 location (2°51'48S, 144°32'04E). The core was collected in Bismarck Sea at -1843 m depth, off Papua New Guinea. The purple box shows the location of Papua New Guinea.

4.4.1. Sediment sample collection

The archive half core (lengthwise split) was preserved intact at 4°C at the CEREGE core repository. In 2019, based on chronostratigraphy of the core, twenty layers were selected at the interglacial and glacial periods (Tachikawa et al. 2011).With a spatula, the thin surface at the subsampling sites were scrapped and sediment was collected in a 15ml tube with a sterile spoon avoiding the edges of the core. Two samples of approximatively 5g were taken per layer. To reduce DNA/RNA damage, the tubes of sediments were filled with 10ml of Lifeguard buffer. Patches of mold were visible to the naked eye on the tops of some cores, we were aware not to touched them cores during the subsampling.

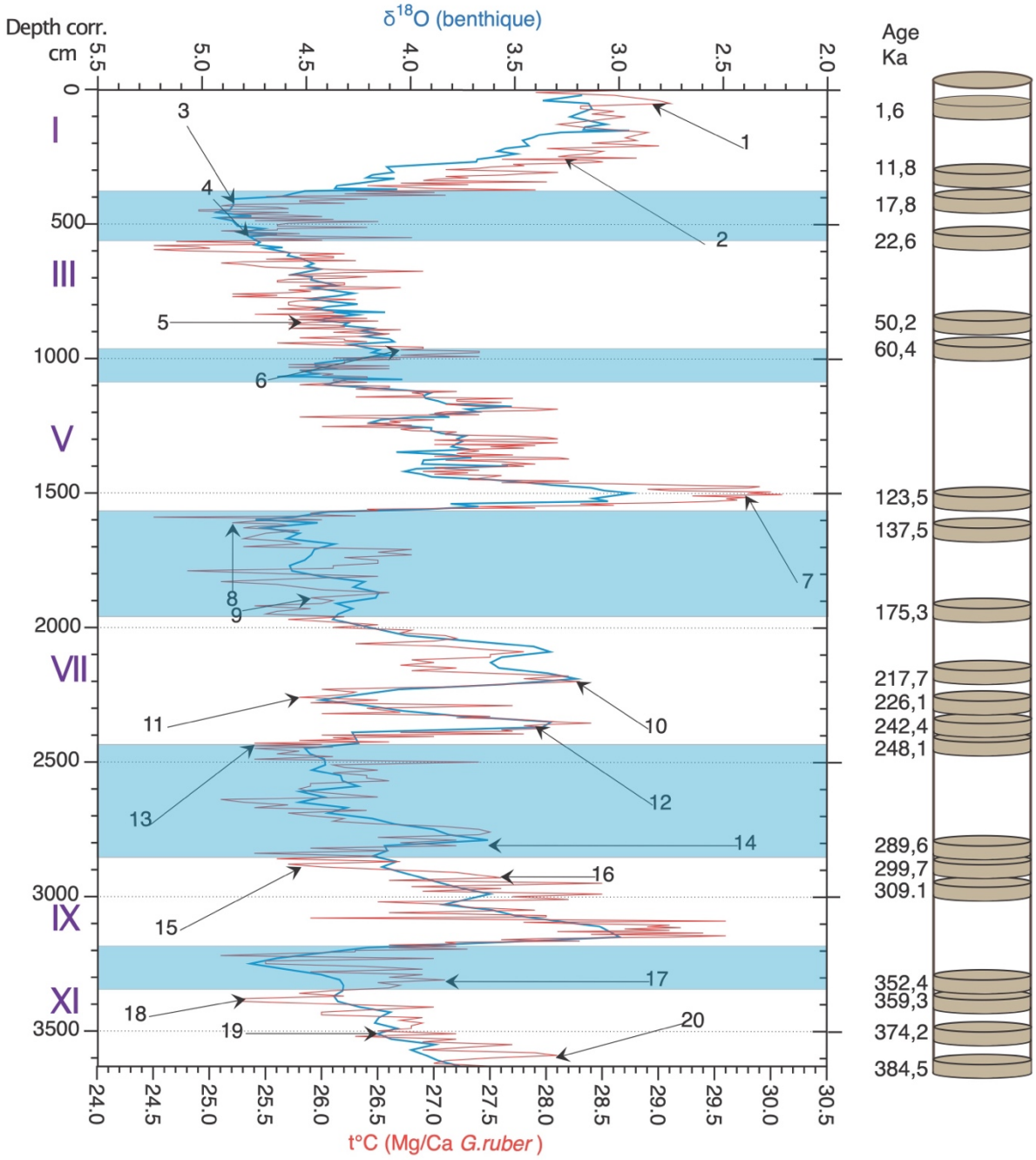

Figure 4.2: MD05-2920 time series of MIS, which are here in roman numbers (I-XI). The blue line shows the δ ^{18}O benthic foraminiferal values obtained from U. peregrina and C. wuellerstorfi over the core. In red line, the Mg/Ca ratio of G. ruber used to reconstruct sea surface temperatures (SST). MD05-2920 MIS time series, which are here in Roman numerals (I-XI). The blue line shows the δ 18O values of benthic foraminifera obtained from U. peregrina and C. wuellerstorfi over the last 384 ka. The red line shows the Mg/Ca ratio of G. ruber used to reconstruct sea surface temperatures (SST). The numbers correspond to the twenty layers that were subsampled in the core. The schematic drawing of the core on the right shows these layers with their respective ages.

4.4.2. Sediment DNA extraction, amplification, and sequencing

For each sample about 5g of sediments was extracted using DNeasy PowerMax Soil kit following the manufacturers instruction. The final elution was concentrated and precipitated with 0.2 ml of 5M NaCl and 10.4 ml of cold EtOH and finally resuspended in 400 µl of elution solution. Extractions were done in batches of 4 samples (two layers) to which two extraction blanks were added to control risks of contamination. These extraction blanks were subsequently processed in the same way as the other samples.

We wanted to get an overview of the eukaryotes that can be obtained after such a long storage time of the core but also specific planktonic taxa and plants. As the DNA was several 100ky old, it was appropriate to use primers that amplify small fragments. Therefore, we selected those targeting the 18S RNA and chloroplastic genes. The 18S V9 hypervariable region (~100-130 bp) was amplified to generate eukaryotes amplicons using 1389F and 1510R primers (Amaral-Zettler et al. 2009). Similarly, for radiolarians we amplified the V4 region (~ 250 bp), using the specific primers S879 (forward) and modified s32Jmod reverse after (Decelle et al. 2012a) to obtain both spumellarians and acantharians. For foraminifera, the 18S 37F hypervariable region (90 - 120 bp) was amplified using for s14F1 and s15 primers (Pawlowski and Lecroq 2010). Finally, for plants we used TrnL-g and TrnL-h primers (Taberlet et al. 2007) to amplify the TrnL P6 loop locus (~ 39 – 77 bp).

The PCR reactions for V9, radiolarians and plants were performed in three replicates and for diatoms and foraminifera in eight replicates. The forward and reverse primers were tagged with eight unique nucleotides at 5’ extremities to enable the multiplexing of samples (Esling et al. 2015). Per each couple of reverse and forward tagged primers, a blank PCR was also included in the amplification. All the replicates were pooled and quantified with QIAxcel system (Qiagen, a high-resolution capillary electrophoresis. Based on the quantification, two 1.5ml tubes were prepared per marker with samples mixed equimolarly and purified to remove dimers. We used the High Pure PCR product Purification kit (Roche) removing under 100bp for 18S amplicons and for plants the MinElute PCR Purification kit which removed under 70 bp. The two libraries per marker were prepared using TruSeq DNA PCR-Free Library Preparation Kit (Illumina) and quantified by qPCR using the Kapa library quantification kit (Roche). The sequencing was achieved on a MiSeq system (Illumina) using paired-end sequencing with 300 cycles for V9, foraminifera, radiolarians, and plants libraries.

4.4.3. Bioinformatic analysis

All bioinformatics analyses were performed using the SLIM platform (Dufresne et al. 2019), which has a set of tools for generating Amplicon Sequences Variant (ASV) data. Briefly, the raw data was demultiplexed and then analysed with DADA2 which removes primers, merges sequences, and removes chimeras. The generated ASVs, were then taxonomic assigned at 95% of identity using Vsearch and PR2 database for V9-eukaryotes, and radiolarians. For foraminifera we used a customised database containing more than 4100 reference benthic sequences and PFR^2 (Morard et al. 2015), a planktonic foraminifera reference database. For plants, we first created a TrnL reference database by collecting sequences from NCBI and we used blast+ to assign the sequences.

Additional filtering steps were conducted to get curated table and sequences. The non-targeted taxa were removed from all datasets (e.g., Bacteria and Archaea in V9). For plants, cultivated plants were eliminated and only sequences referring to plants endemic to Papua New Guinea or the surrounding islands were retained. To determine which plants are endemic we used a list published by Cámara-Leret et al. 2020 and Plants of the World Online (POWO). Sequences present in extraction controls were suppressed with microdecon (McKnight et al. 2019) or manually. Moreover, sequences with less than 1000 reads were removed. This is to ensure that the sequences obtained are real, not just a trace or an artefact of PCR or sequencing.

All datasets were normalized before alpha and beta analysis. In some cases, due to the low sequencing depth, some samples were removed from the dataset.

4.5. Results

4.5.1. Sequence data

After all filtering steps, we observed good DNA preservation in all markers up to 200 Ka, where the average number of reads per sediment layer varied between markers from 139’003 to 928’440. Below 200 Ka, a few taxa were sporadically amplified (Fig. 4.3). The average number of reads per layer ranged from 25’187 to 467’005 reads.

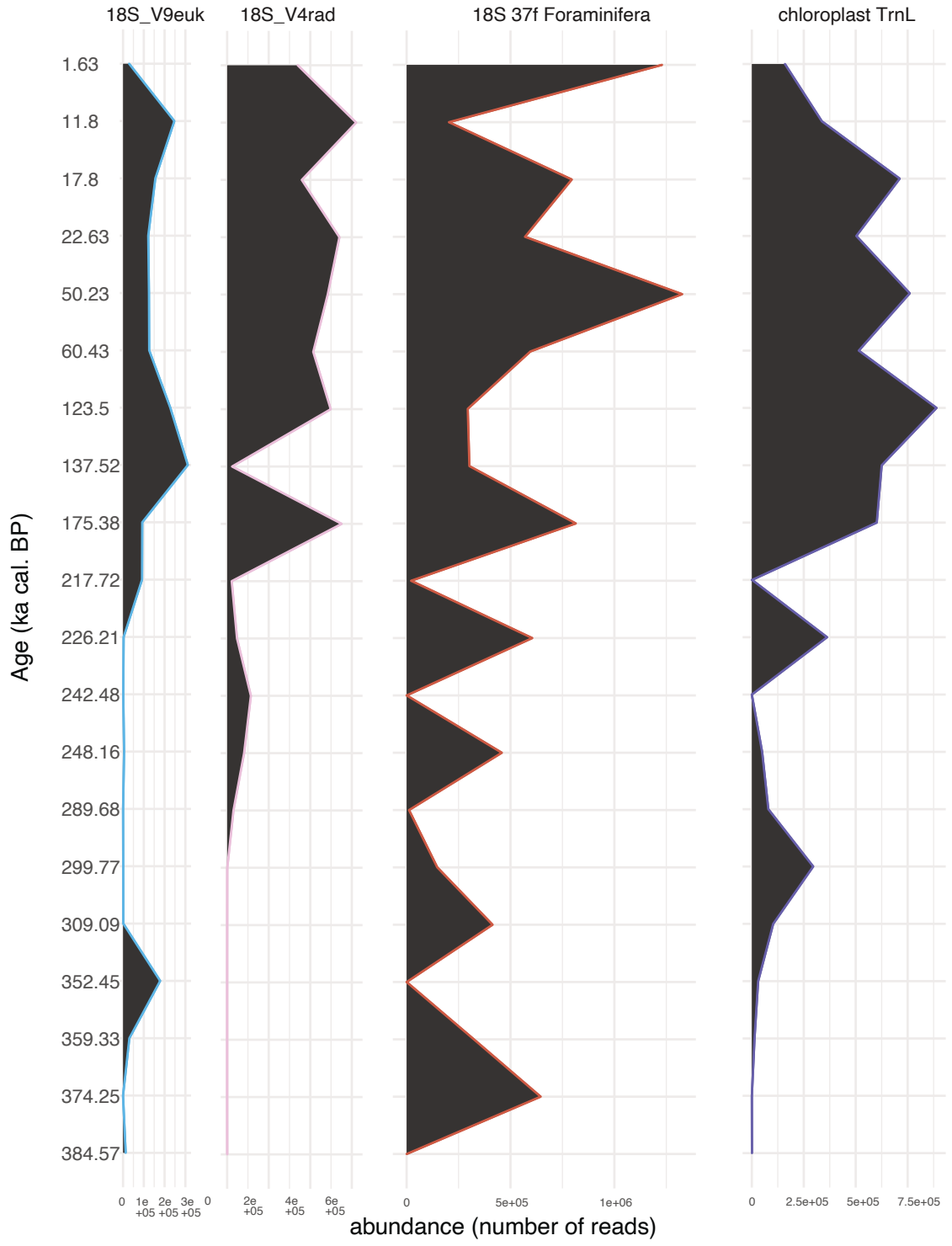

Figure 4.3: Number of reads variation throughout the core per marker: 18S-V9 for eukaryotes, 18S-V4-rad for radiolarians, 18S-37F for foraminifera and TrnL P6 loop for vascular plants

4.5.2. Taxonomic assignment (Fig. 4.4)

4.5.2.1. V9- eukaryotic marker

After removal of reads appearing in controls 336 ASVs. The V9 dataset was dominated by the sequences assigned to an amoebozoan *Acanthamoeba* sp (Fig. 4.5 A.). Fungi AVSs were also present along the core. Neither amoebozoan nor fungal sequences were present in the blanks suggesting that they represent a contamination of the core rather than acquired during its processing. We supposed that the fungal ASV could come from mould patches observed on the top of some. All these sequences were removed from taxonomic and composition analysis. After excluding Fungi and Amoebozoa, the V9 dataset was composed of Alveolata, Archaeplastida, Excavata, Opisthokonta (mainly metazoans), Rhizaria and Stramenopiles (fig. 2) until 359 Ka (MIS11) (Fig. 4.5 B.). After 200 Ka (MIS1-6), one or two phyla dominate all the layer's composition, probably due to poor DNA conservation. Before 200 Ka, all phyla were represented, and we observed variation between glacial and interglacial layers. For instance, the proportion of Alveolata decreased after each glacial period, while those from Rhizaria increased. The alpha diversity indexes (Fig. 4.5 C.) showed higher values in glacial samples following the interglacial for observed and Shannon and Simpson indexes. The multidimensional scaling (MDS) and hierarchical clustering (HC) (Fig. 4.5 D.-E.) indicated groups with similar compositions during glacial or interglacial periods. However, it was also observed some samples at the boundary of these periods still maintain a composition close to the previous period, as if it was a slow transition.

4.5.2.2. Radiolarians marker

We could amplify radiolarian with a high sequencing depth down to 289 Ka (MIS 8). The mean number of reads per sample was 179'311 corresponding to 1465 ASV. At species taxonomic level, 30 species (29 genera) could be identified (Fig. 4.6 A.). The radiolarian class Polycystinea tended to dominate throughout the core. Acantharea and Rad-A were mostly present in glacial samples. Spumellarida-Group I, a genus from Polycystinea class was the most common species across all layers (>70% per sample).

Interestingly, glacial (e.g., MIS2, MIS4) samples exhibited a higher diversity than those from interglacial period (e.g., MIS1) (Fig. 4.6 B.). The MIS2 have the greatest diversity. However, Simpson index showed an evenness of diversity between MIS1 to MIS4 and less difference in diversity with older samples (Fig. 4.6 B.). The MDS plot displayed three groups, samples from

MIS1 to MIS4 were regrouped but glacial and interglacial samples were slightly different in composition (Fig. 4.6 D.). The hierarchical clustering separated the samples into two main groups, curiously those from MIS 6-8 are together except samples at 137 Ka and those from MIS 1-5 form another group. The first group did not display any community variation accordingly with the dendrogram (Fig. 4.6 D.). The average SST of this group was 26°C in contrast to the second group where the average SST was 27°C (Fig. 4.2).

4.5.2.3. Foraminifera marker

4.5.2.3.1. Planktonics

After a dataset curation, we kept 809 ASV, including the unassigned ASV as they presented the characteristics of foraminifera 37f amplicons. 190 ASV belonged to the group of planktonic foraminifera, which represented 13 species (Fig. 4.4), and some cryptic species were also identified such as *N. dutertrei I*., *G. glutinata III*, *G. uvula I*, *G. scitula I*, *P. obliquiloculata I*, *IIa* and *IIb*. The planktonics were identified in all layers except in those under 352Ka. Knowing that planktonics have a specific niche related to sea surface temperature, the identified species were categorized as either restricted to temperate or warm or living in both regions. As expected in glacial periods, warm species (e.g., *C. nitida*, *N. dutertrei*, *P. obliquiloculata*, *G. conglomerata*, *G. tumida*, *S. deshicens* and *G. vivans*) had to share the area with temperate and warm-temperate species (e.g., *G. uvula*, *G. glutinata*, *G. minuta*, *G. hirsuta*, *G. scitula* and *T. iota*) (Fig. 4.7A-B.).

Alpha diversity indices indicated again variations between samples from MIS1 - MIS4 and between MIS5 - MIS8. The first group showed a higher diversity as with other markers. MIS3, an interglacial period had the highest observed number of species but according to the Shannon and Simpson values it was not the most diverse period (Fig. 4.7 C.). The MDS and dendrogram regrouped periods into 3 groups, here again glacial and interglacial periods were mingled (Fig. 4.7 D-E.).

4.5.2.3.2. Benthics

The remaining 619 ASV were ascribed to benthic foraminifera. Among them 408 ASV were unassigned and 216 ASV belonged to Globothalamea, Monothalamea and Tubothalamea (Fig. 4.8 A.). Within globothalamids, rotaliids and textulariids were encountered in all layers until 217 Ka. Textulariids showed a higher proportion in MIS2 and MIS6 (glacial periods). In

monothalamids, Clades BM, V and Y were present across the core, and together with some unassigned detected also in the oldest layers. Miliolids were detected only at 50 Ka. As it was shown in Fig. 4.4 these unknown foraminifera were detected in almost all layers and in some of them, they were the most abundant (>25%). Usually, as they do not have any ecological meaning they are removed from dataset. However here, we kept the five most abundant unassigned ASVs (ASV5, ASV8, ASV16, ASV19 and ASV21) as they account for more than 15% of all benthic reads.

Alpha diversity indexes showed again higher diversity in samples from MIS1 to MIS4 and glacial periods were characterized by higher diversity than the preceded interglacial periods (Fig. 4.8 B.). The MDS and dendrogram plot showed three groups. Samples belonging to MIS10 - MIS 5 were regrouped, with the exception 384.57 Ka and 242.48 Ka. The younger samples from MIS1 - MIS4 displayed the third group (Fig. 4.8 C-D.).

4.5.2.4. Vascular plants marker

The three classes of plants Magnoliopsida, Pinopsida and Polypodiopsida were amplified with the TrnL marker down to 359 Ka (MIS 11). We excluded plants reads of plants considered as food and cultivated plants. The Magnoliopsida were the most abundant across the core (> 87%) (Fig. 4.4). A total of 66 species (90/400 ASV, 1,713,471 reads) were considered as endemic to PNG, while 111 (178/400 ASV) genus and one (1/400 ASV) family were reported as native to surroundings islands of PNG (Fig. 4.9 A.). In further analysis we kept only the endemic ones. We regrouped the plants accordingly to their functional groups into climber, epiphyte, fern, herb, and woody (tree and shrubs). Climber plants were more abundant during glacial periods except at 299 Ka corresponding to the interglacial MIS9 where more reads were detected. Epiphytes were present in small proportion only in glacial periods. Ferns had a peak of abundance (>15%) at 137 Ka (MIS6). Herbs found in glacial and interglacial samples. The peak abundance was in the oldest sample at 359 Ka (MIS 11) and in more younger layers it was at 123 Ka (MIS5). Woody (bushes and trees) type had a maximum abundance at 248 Ka (MIS 8), but in younger samples the peak was at 11.8 Ka (MIS1).

To summarize, we noticed that during glacial periods, vascular plants from any type had higher proportion that during interglacial periods. Likewise, in the classification of plants by biomes, there was a significative presence of subtropical plants, more than 25% at 137 Ka (MIS6). The alpha indexes confirmed again a greater diversity during glacial periods. The MDS and HC displayed a spatial layout of glacial samples having similar composition except for sample at

50 Ka (MIS3) and 299 Ka (MIS9). These two exceptions are at the boundaries of glacial-interglacial.

4.6. Discussion

4.6.1. Use of long-time stored sedimentary cores as molecular archives

Many cores have been deposited in repositories after their collection by oceanographic campaigns, usually a half lengthwise core is still stored intact as archived core. This represent a material that can be used even though the sampling and subsampling for ancient eDNA rules were not applied. The cold storage temperature of 4°C could limit the DNA damage as it corresponds to bottom water temperature. Studies reported successfully metabarcoding and shotgun sequencing on such cores where the subsampling was done at least 5 years after collection (Clarke et al. 2020; Selway et al. 2022).
In our case, despite 14 years storage, we were able to amplify from the youngest to the oldest layers ranging from 1.6 Ka to 384 Ka. We excepted a low DNA concentration in oldest layers due to sampling conditions and above all the storage time, which was long enough for microbial and fungal growth.

In addition to bacteria or fungal DNA, we found those of Acanthamoebas. Their presence in seawater in free-living form or as cysts is rather common (Hussain et al. 2022); we could assume that their DNA is contemporary to the sedimentation, but we cannot exclude contamination during core recovery as the blank extraction samples were free of them. It is well known that modern DNA can compete with ancient DNA during PCR amplification (Webster et al. 2003; Giguet-Covex et al. 2019). So, if the Acanthamoebas were introduced during the coring, their fresh DNA diluted the ancient ones which explains such a large amount of reads not only in eukaryotes but also in the foraminiferal and radiolarian datasets.
To avoid dealing with exogenous taxa, we opted for taxon-specific primers as we assumed that targeting planktonics and tropical plants allows to control external contamination.
Since the older is the sample, the more the DNA is fragmented, these markers targeting short gene fragments (<300 bp) are preferential for. From our results and those of Armbrecht et al. 2020, 2021, we conclude that only primers amplifying less than 150 bp are suitable for ancient eDNA studies and are short enough to expect to amplify the maximum number of species.

However, even with short primers, it must be considered that the taxonomic resolution is not necessarily very high and not all short markers gave high sequencing depths in oldest samples either.

There exist other techniques such as shotgun sequencing allowing the ancient DNA authentication by size selection, but here again bacteria or fungi will overcome other taxa. Probably hybrid capture, used more and more (Murchie et al. 2021; Armbrecht et al. 2021a) to enrich for specific taxa like plants (Nota et al. 2021) in old samples would be ideal in the case of long-time stored cores. With metabarcoding, another possibility would be to considerably increase the number of PCRs per sample and sequence them separately, as proposed by Ficetola et al. 2015.

4.6.2. Protists DNA assemblages' evolution during climate changes

Planktonic composition reflects sea surface conditions and therefore SST is usually calculated based on planktonic foraminiferal tests. Knowing the glacial and non-glacial intervals recorded in the core with temperature variations up to 5°C, changes in the planktonic community were expected. Radiolarians are known to have a specific niche from tropics to poles. According to Boltovskoy and Correa 2017 and Hernández-Almeida et al. 2017, the distribution of modern species is highly influenced by SST and reach species richness in the tropics.

In the past, during the Neogene (5 Ma) a remarkable cooling led to the migration of temperate and polar species to the tropics (Trubovitz et al. 2020). This may explain the greater diversity in cold periods (MIS 2,4,6) following warm periods. In the case of a global increase in temperature, the radiolarian niche only enlarges, the richness may increase but not the diversity. In our core, the 4°C -5°C variation between warm and cold periods in was significative to lead change the diversity of radiolarians. This result can be extrapolated to the current and future warming.

During cold periods, planktonic foraminifera communities had to share their niche with subtropical and temperate species. The shifts towards the tropics of the latter increments the diversity in term of species. On other hand the variation of SST across the time contributed to diversify planktonic foraminifera in the past. In fact, some species adapted rapidly to their environments giving new species or cryptic species. For instance, *P. obliquiloculata* species had three cryptic species which we were able to identify: *P. obliquiloculata I*, *P. obliquiloculata IIa* and *P. obliquiloculata IIb*. *P. obliquiloculata I* is a tropical species globally distributed and diverged around 3.1 Ma. Types *IIa* and *IIb* diverged around 1.4 Ma. The *IIa*

habitat is in more warmer temperatures (25°C – 29°C) while *IIb* in a larger range of temperature (13°C – 29°C) (Ujiié and Ishitani 2016; Pearson and Penny 2021). Those small changes on DNA sequence were also reported in polar areas in Pawłowska et al. 2020 where potentially subspecies of *N. pachyderma* were identified across the time.

Benthic foraminiferal communities changed between glacial and interglacial periods. Their diversity values were higher in cold periods as were those of planktonic protists. The parameters controlling the abundance and diversity of benthic foraminifera are food availability and oxygen variations (Thomas and Gooday 1996; Gooday 2001). We suppose that the decrease in diversity during interglacial periods is due to the decrease in primary productivity. All markers considered; the observations were the same: diversity was greater during cold periods. This finding differs from studies in the high latitudes where marine and terrestrial species could only colonise and develop after the ice caps had retreated (Pedersen et al. 2016; Nota et al. 2022). It seems that in tropical areas, a higher temperature in the sea or on land is not favourable to biodiversity.

4.6.3. Evolution of plant biodiversity in PNG

The choice of use plant markers for a marine core was based on the presence of freshwater foraminifera and streptophyta in eukaryote markers indicating a high input of terrigenous sediments in the core. Moreover, the Tachikawa et al. 2011; Savranskaia et al. 2021 confirmed the origin of the sedimentary particles by the concentrations of ratio Ti/Ca and $^{10}Be/^{9}Be$. We assumed that plant remains (a piece of leaf, a root) could also be transported, and thus that plant DNA could be preserved in the same way as the remains of marine organisms.

After (Cámara-Leret et al. 2020) over 13'000 plant species have been recorded recently in PNG, the majority of which were identified as endemic. The number of species obtained in our study identified as endemic is close to 90 species. This is higher than in the Northern Hemisphere metabarcoding studies (Parducci et al. 2015; Clarke et al. 2019; Rijal et al. 2021) and even than studies conducted in the tropics (Bremond et al. 2017). The high number of species confirms by eDNA that the island has a very high diversity and presumably also in the past.

Since not all species in the catalogue have been sequenced, for 59% of genera found in PNG (Cámara-Leret et al. 2020). It can be assumed that the additional 150 species (111 genera) identified as native to the surrounding areas, i.e., northern Australia or the islands of Indonesian-Malesia, could exist on PNG or at least species from the same genus.

We found greater diversity during glacial intervals, which is contrary to studies in temperate and polar zones, where a recolonization of land was observed during postglacial period (Tollefsrud et al. 2008; Tsuda et al. 2016; Nota et al. 2022). As few plants ancient eDNA studies have been carried out in the tropics (Boessenkool et al. 2014; Dommain et al. 2020), we can speculate from our data that a few degrees less in the lowlands tropics favors the growth of some plants. It should be remembered that during these glacial and intergalactic periods, the SST varied by 4 to 5 degrees (i.e., from 24°C to 30°C), so same variation in the ambient temperature on the island at least in the lowlands could be expected. A review on effects of warming in humid tropical alpine, mentioned the biodiversity could suffer as some species cannot adapt to changing conditions due to rainfall, soils, and CO2 storage changes (Buytaert et al. 2011). Nevertheless, in highlands PNG the glaciers extension during last glacial maximum (LGM) was until 3400 m (Prentice et al. 2005). Pollen data on Hogayaku lake PNG cores documented the evolution of vegetation between an interglacial (MIS1) and glacial (MIS2) periods (Prentice et al. 2005). As the temperature increased, alpine vegetation was gradually replaced by subalpine vegetation and other plant types. In the highlands, the vegetation was more diverse during the interglacial, due to retreat of glaciers. This change follows more the pattern of vegetation changes in temperate regions.

4.7. Conclusions

In conclusion, in our study we have shown the potential of using a long-stored tropical core to assess past biodiversity. The specific primers used prevent the amplification of modern and exogenous DNA (modern contaminants). Our metabarcoding data show a high preservation of DNA up to 200 Ka, enough to follow some glacial and interglacial stages. The global biodiversity on land and sea was higher during cold periods in the tropics.

Apart from the challenges associated with core storage, its tropical location, we note a lack of database in tropical regions, at least in plants.

.

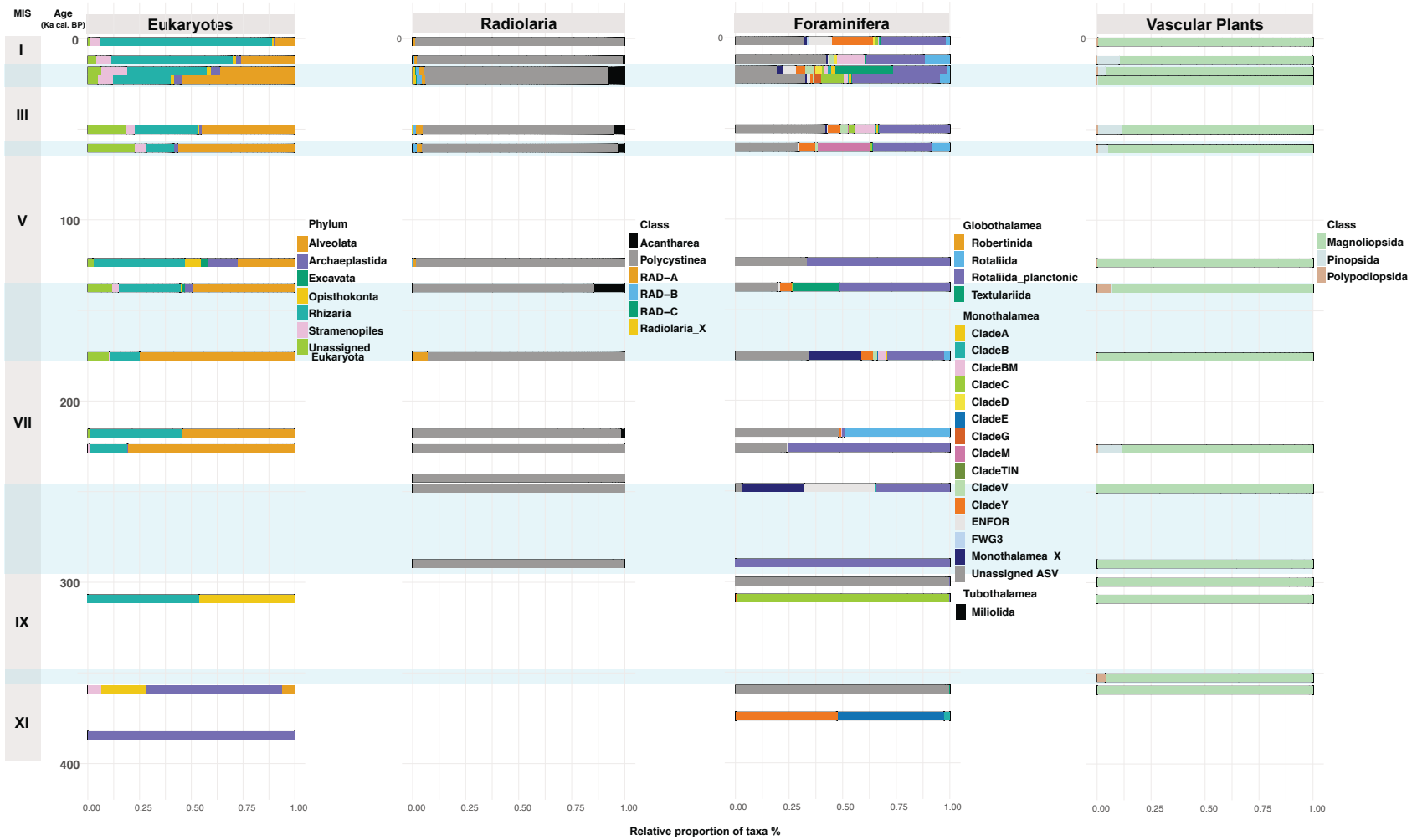

Figure 4.4:Paleocommunitiy composition per maker at higher taxonomic level. Eukaryotes at phylum level without the Amoebozoa, radiolarians at class level, foraminifera at class and order level and finally vascular plants at class level. The Marine isotope stages (MIS) are in roman numbers. The MIS corresponding to glacial periods are in light blue.

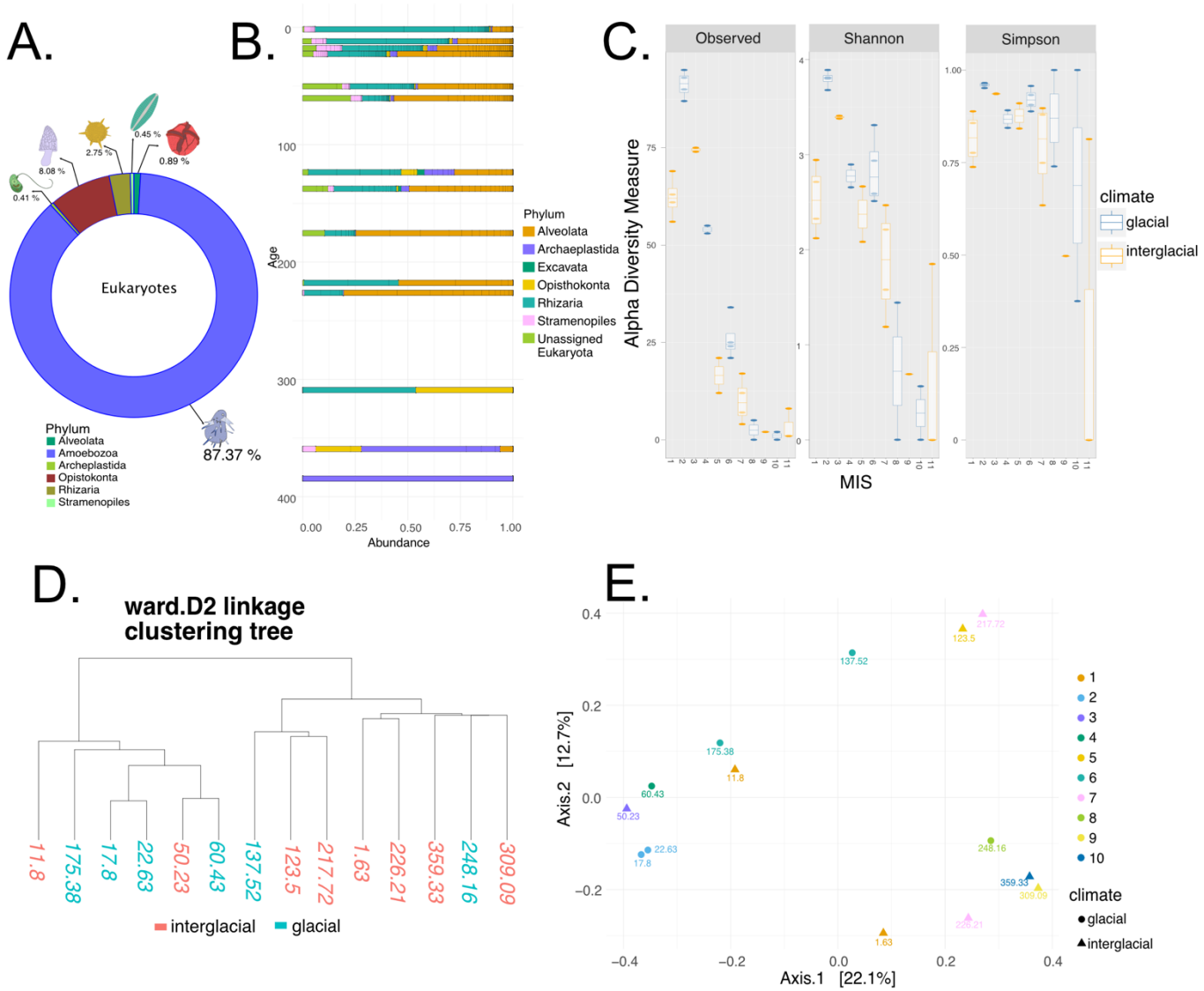

Figure 4.5: Eukaryotes assemblage. A. Taxonomic composition of all eukaryotic sequences. This graph shows the percentage of amoebozoans in the dataset that were deleted in subsequent analyses. B. Taxonomic composition at phylum level across the core. Only 14/20 samples that amplified well are shown here. C. Alpha diversity (Shannon and Simpson) at each MIS, here until MIS11. Blue boxplots correspond to glacial periods and orange to interglacial periods. D. Hierarchical clustering of samples with colored ages corresponding to glacial or interglacial periods. E. Multidimensional scale (MDS), colored by MIS, here 10. The periods corresponding to glacial are in the round and those corresponding to interglacials in the triangle.

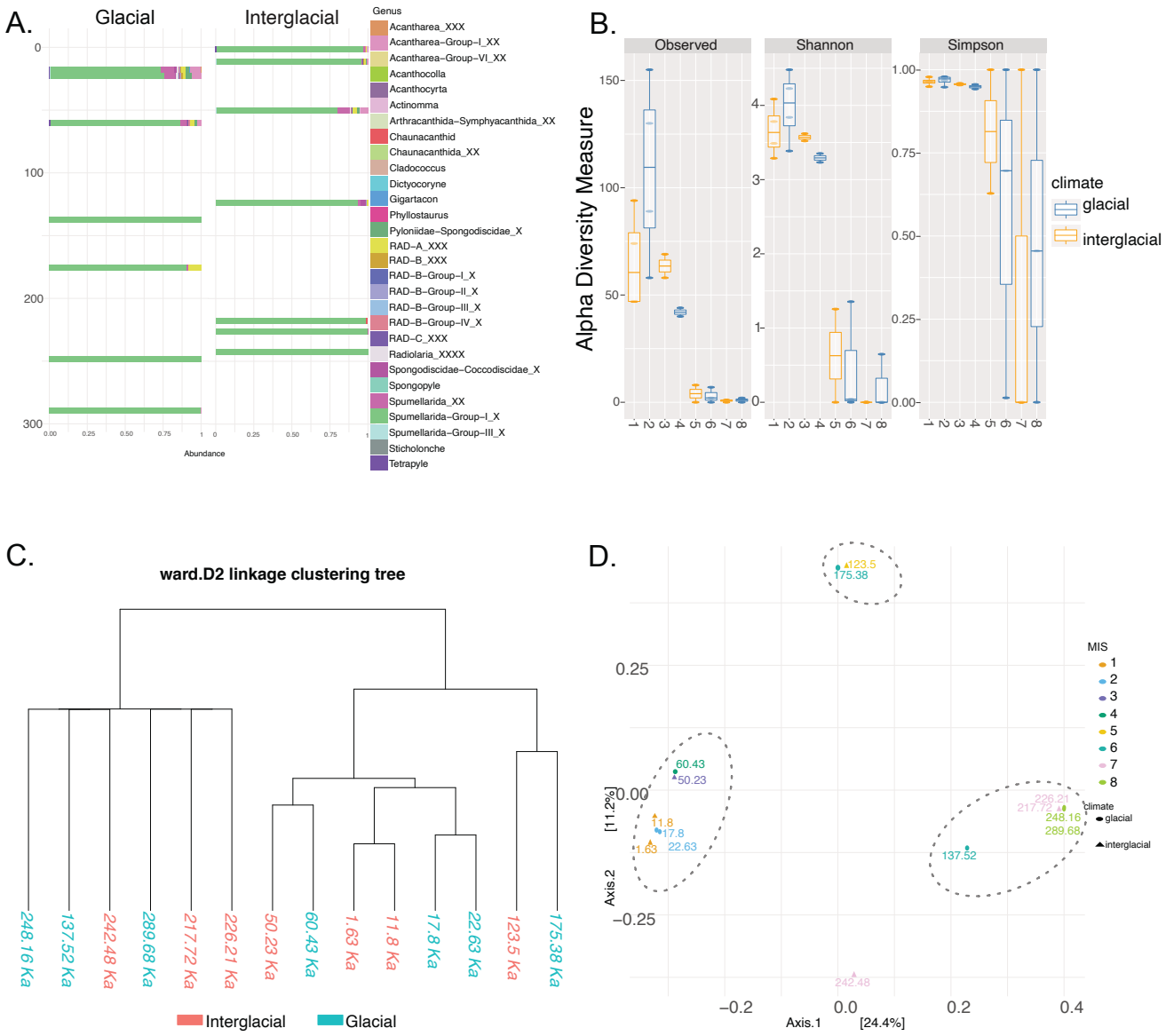

Figure 4.6: Radiolarians assemblage. A. Taxonomic composition at genus level across the core split into glacial or interglacial. Only 14/20 samples that amplified well are shown here. B. Alpha diversity (Shannon and Simpson) at each MIS, here until MIS8. Blue boxplots correspond to glacial periods and orange to interglacial periods. C. Hierarchical clustering of samples with colored ages corresponding to glacial or interglacial periods. D. Multidimensional scale (MDS), colored by MIS, here 8. The periods corresponding to glacial are in the round and those corresponding to interglacials in the triangle.

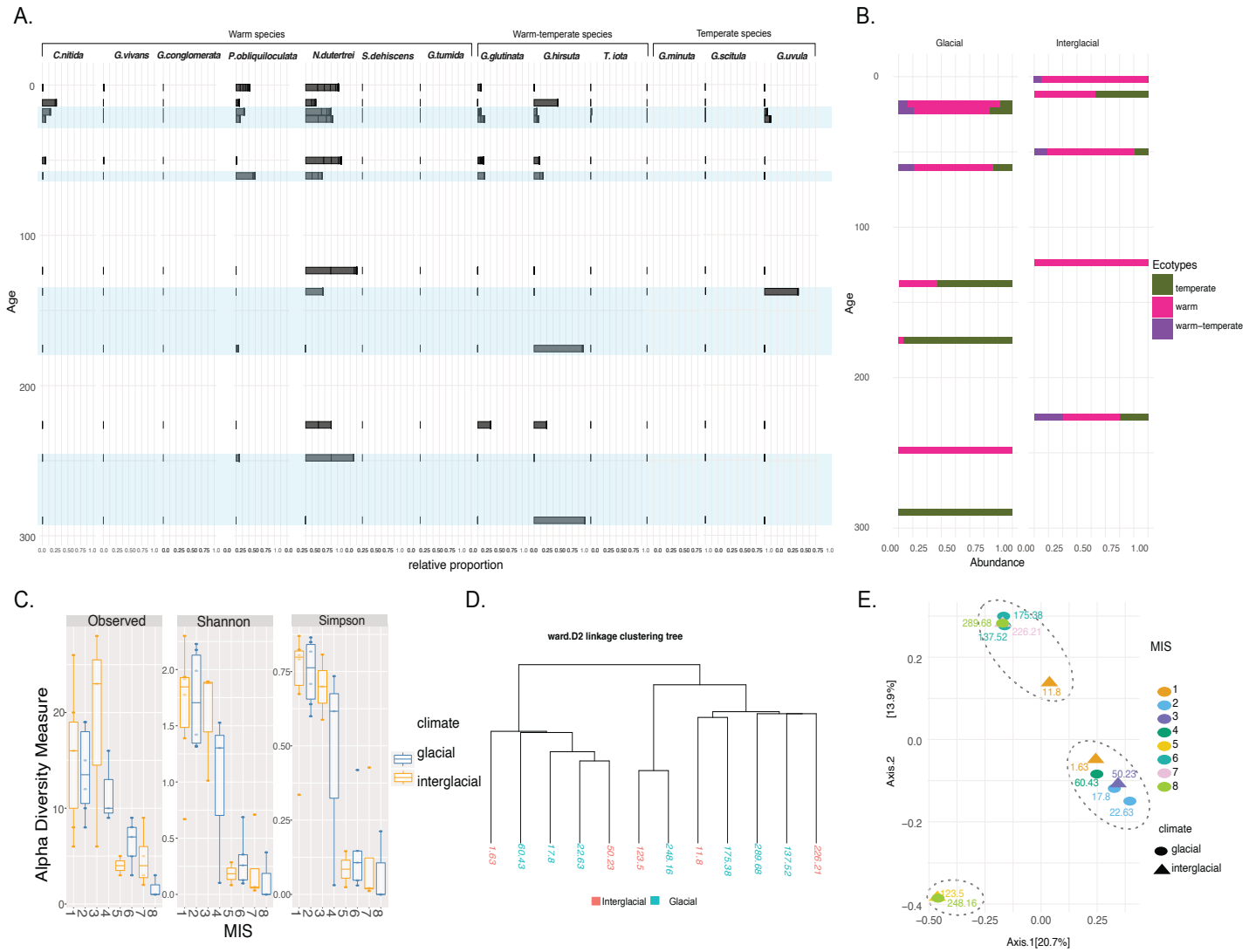

Figure 4.7: Planktonic foraminifera assemblage. A. Taxonomic composition at species level across the core. Species were regrouped by ecological groups (warm, warm-temperate, and temperate). Only 12/20 samples that amplified well are shown here. B. Ecological groups distribution during the last 300 ka. C. Alpha diversity (Shannon and Simpson) at each MIS, here 8. Blue boxplots correspond to glacial periods and orange to interglacial periods. D. Hierarchical clustering of samples with colored ages corresponding to glacial or interglacial periods. D. Multidimensional scale (MDS), colored by MIS, here until MIS8. The periods corresponding to glacial are in the round and those corresponding to interglacials in the triangle.

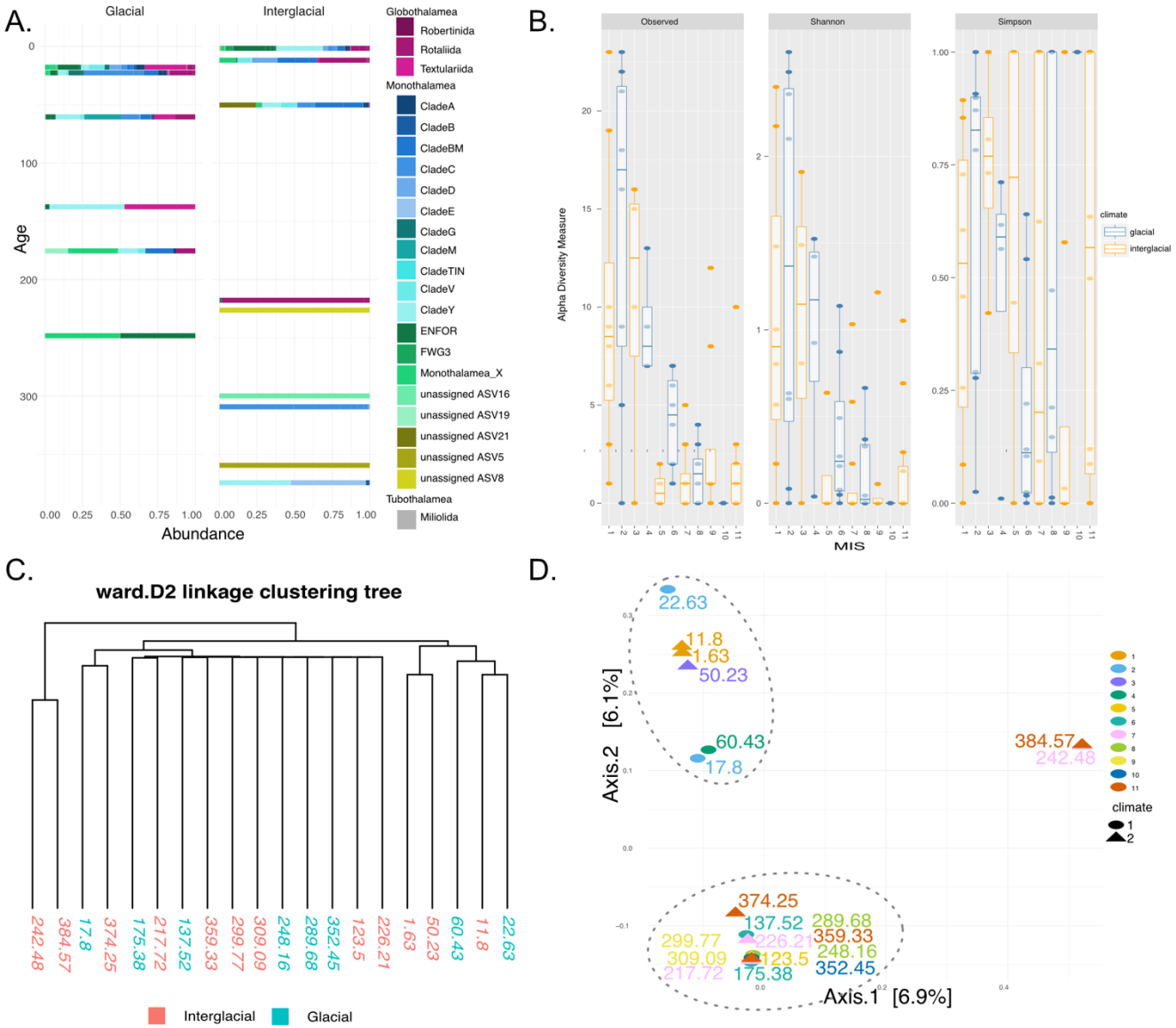

Figure 4.8: Benthic foraminifera assemblage. A. Taxonomic composition at class and order level split into glacial and interglacial. Only 15/20 samples that amplified well are shown here. B. Alpha diversity (Shannon and Simpson) at each MIS. Blue boxplots correspond to glacial periods and orange to interglacial periods. C. Hierarchical clustering of samples with colored ages corresponding to glacial or interglacial periods. D. Multidimensional scale (MDS), colored by MIS, here until MIS11. The periods corresponding to glacial are in the round and those corresponding to interglacials in the triangle.

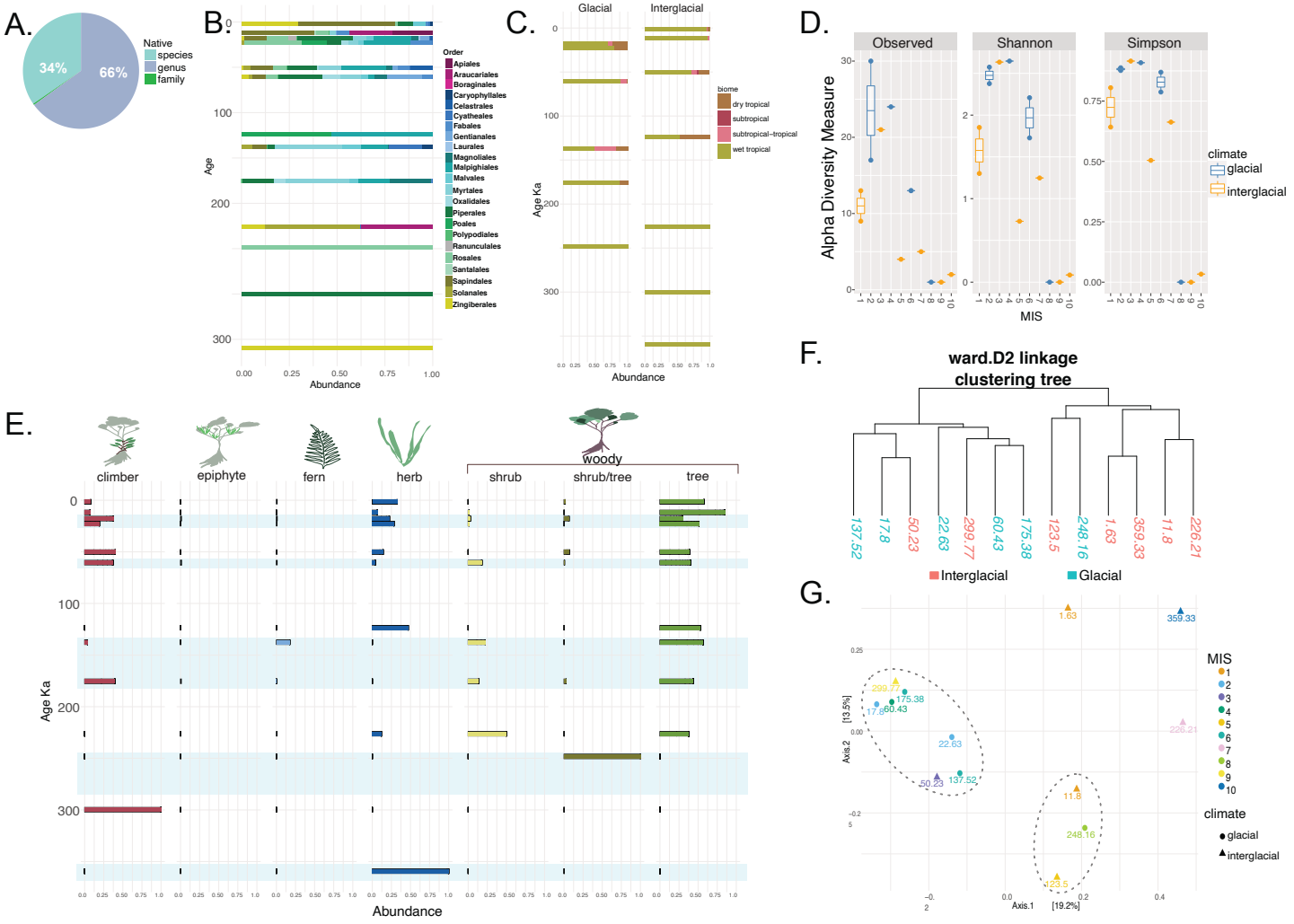

Figure 4.9:Vascular plants assemblage. A. Proportion of species/genus/familly identified as native. B. Taxonomic composition at order level. C. Biome composition across the core split into glacial and interglacial. D. Alpha diversity (Shannon and Simpson) at each MIS, here until MIS10. E. Ecotypes composition across the core, the blue rectangle in the background marks the ice ages. F. Hierarchical clustering of samples with colored ages according to glacial or interglacial periods G. Multidimensional scale (MDS), colored by MIS, here 10. The periods corresponding to glacial are in the round and those corresponding to interglacials in the triangle.

Chapter 5: Patterns of eukaryotic diversity from the surface to the deep-ocean sediment

Tristan Cordier, Inès Barrenechea Angeles, Nicolas Henry, Franck Lejzerowicz, Cédric Berney, Raphaël Morard, Angelika Brandt, Marie-Anne Cambon-Bonavita, Lionel Guidi, Fabien Lombard, Pedro Martinez Arbizu, Ramon Massana, Covadonga Orejas, Julie Poulain, Craig R. Smith, Patrick Wincker, Sophie Arnaud-Haond, Andrew J. Gooday, Colomban de Vargas, Jan Pawlowski

5.1. Project description

This project started when I completed the sequencing of 18S V9 fragments from samples collected in many deep-sea locations, mainly by Franck Lejzerowicz. Tristan Cordier, Jan Pawlowski and I realised that, as in the previous studies (chapters 3.1 and 3.2), many planktonic sequences were detected in deep-sea sediments. To distinguish between pelagic and benthic sequences, we needed V9 amplicons. We first contacted members of the TARA expeditions (from Vargas) to obtain their sequences. This later turned into a global study, with the participation of many groups from 15 expeditions that have conducted metabarcoding studies in surface waters and deep sediments in recent years.

A modified version was published in *Sciences Advances* 8(5), eabj9309.

5.2. Abstract

Remote deep-ocean sediment (DOS) ecosystems are among the least explored biomes on Earth. Genomic assessments of their biodiversity have failed to separate indigenous benthic organisms from sinking plankton. Here, we compare global-scale eukaryotic DNA metabarcoding datasets (18*S*-V9) from abyssal and lower bathyal surficial sediments and euphotic and aphotic ocean pelagic layers to distinguish plankton from benthic diversity in sediment material. Based on 1685 samples collected throughout the world ocean, we show that DOS diversity is at least threefold that in pelagic realms, with nearly two-thirds represented by abundant yet unknown eukaryotes. These benthic communities are spatially structured by ocean basins and particulate organic carbon (POC) flux from the upper ocean. Plankton DNA reaching the DOS originates from abundant species, with maximal deposition at high latitudes. Its seafloor DNA signature predicts variations in POC export from the surface and reveals previously overlooked taxa that may drive the biological carbon pump.

5.3. Introduction

Deep-ocean sediment (DOS) ecosystems cover more than half of Earth's surface and remain one of the least explored ecosystems on the planet. This vast and heterogeneous environment provides habitats for diverse biological communities that support fundamental ecological processes and services, such as nutrient recycling for the healthy functioning of ocean ecosystems and carbon sequestration for the regulation of Earth's climate over geological time scales (Thurber et al. 2014). The DOS is exposed to growing anthropogenic pressures, notably from climate change (Levin and Bris 2015; Breitburg et al. 2018), deep-sea mining (Smith et al. 2020), oil and gas exploitation, and bottom trawling (Pusceddu et al. 2014), making a scientifically informed protection of its biodiversity a matter of the highest importance (Barbier et al. 2014; Mengerink et al. 2014; Levin et al. 2020).

For more than 50 years, a considerable effort has been devoted to understanding the diversity and biogeography of benthic organisms thriving in the DOS (Rex and Etter 2010). However, the enormous extent of this habitat and its remote location under several kilometers of water means that only a minute proportion has ever been sampled. Most previous studies have focused on morphological analyses of the macro- and mega-fauna, which typically show high levels of a diversity and small-scale faunal patchiness (Grassle and Morse-Porteous 1987; Rex

and Etter 2010; McClain and Rex 2015), and have recently been proposed as biological indicators for deep-ocean monitoring and conservation (Danovaro et al. 2020a). Less attention has been paid to the microbial and meiofaunal organisms that numerically dominate DOS communities (Gooday et al. 2020; Ingels et al. 2020) but can hardly be identified using classical morphotaxonomic approaches. Studying planktonic organisms sinking to DOS is hampered by similar technical limitations related to great depths (resulting in poor spatial coverage of sinking plankton datasets) and limited morphological identification (but for shell-building taxa that can keep distinctive features once in the sediment). Their study is hence often approached indirectly, using sediment traps to capture the sinking flux of taxa over time that contribute most to the biologically driven carbon sequestration in the deep ocean before they reach the sediment (Thiel et al.; Billett et al. 1983; Silver and Gowing 1991).

The development of high-throughput environmental genomics has begun to fill these gaps in knowledge, revealing substantial unknown diversity among viruses (Zheng et al. 2021) and prokaryotes (Bienhold et al. 2016; Danovaro et al. 2016; Shulse et al. 2017; Hoshino et al. 2020) from DOS. Yet, the use of genomics to explore DOS eukaryotes has been limited and focused mostly on particular taxonomic groups (Lecroq et al. 2011; Lejzerowicz et al. 2014; Sinniger et al. 2016a) or geographic regions [(Danovaro et al. 2010; Shulse et al. 2017; Fonseca et al. 2017), but see (Lejzerowicz et al. 2021)]. One major challenge in interpreting molecular data from DOS is to distinguish DNA reads that belong to indigenous benthic eukaryotes from those originating from pelagic organisms that sink through the water column and leave their DNA traces in the sediments (Pawlowski et al. 2011; Morard et al. 2017; Lindh et al. 2017; Laroche et al. 2020).

Here, we tackle these problems by comparing a newly generated, global-scale DNA metabarcoding dataset of total eukaryotic diversity from deep oceanic surficial sediments (418 samples collected during 15 oceanographic cruises from 2010 to 2016; table S1) to comparable published datasets from euphotic (1160 samples from the *Tara* Oceans expeditions) (de Vargas et al. 2015; Ibarbalz et al. 2019) and aphotic (138 samples from the *Tara* Oceans and Malaspina expeditions) (Obiol et al. 2020) zones across the world ocean. Together, these represent the first consistent molecular meta-dataset spanning the three main open-ocean realms (pelagic euphotic, pelagic aphotic, DOS) at a global scale across 447 sampling sites (Fig. 5.1A). We assembled ~2.42 billion DNA reads (table S2), produced by polymerase chain reaction (PCR) amplification of the V9 region of the 18*S* ribosomal RNA gene, and processed them using the DADA2 workflow to infer amplicon sequence variants (ASVs). On the basis of taxonomic annotations of ASVs using the SILVA and PR^2 sequence databases and on the occurrence of a

highly conserved DNA sequence motif across eukaryotes, we discarded prokaryotic, plastidic, and mitochondrial ASVs, as well as technical artifacts, allowing us to focus on eukaryotic diversity.

5.4. Materials and methods

5.4.1. DOS sample collection

DOSs have been collected during two main projects (deep_sea and eDNAbyss). For the deep_sea project, sediment samples were collected at abyssal depths during eight expeditions to the Arctic, Atlantic, Southern, and Pacific Oceans (table S1). We used disposable sterile spoons to subsample the top surface sediment centimeter (c.a. 2 g from 0 to 1 cm) following a nested sampling design: up to three pseudo-replicates per core, up to two cores per deployment (multicorer), and up to three deployments per station (detailed list in table S7). Sediment samples were placed in sterile falcon tubes with (VEMA, SYSTCOII, KuramBio I, MANGAN'16, ABYSSLINE) and without (MSM39, DIVA3, and BIONOD) 6 ml of Lifeguard Preservation Solution (QIAGEN) before being frozen onboard at −20°C. The samples were shipped within −20°C containers to the University of Geneva (Switzerland). Upon arrival, sediment samples were stored at −80°C until extraction of nucleic acids.

For the eDNAbyss project, lower bathyal and abyssal sediments were collected in the Mediterranean Sea, North Atlantic, and Arctic Oceans during five cruises (Arctic: MarMine; North Atlantic: MEDWAVES; Mediterranean: PEACETIME, CANHROV, and ESSNAUT; table S1). For each station, triplicate cores (10 cm in diameter) were collected with a multicorer or with a remotely operated vehicle. Surface sediment (0 to 1 cm) was collected using metallic spatulas previously sterilized with bleach or DNA Exitus, rinsed with ethanol 96° and then nanopure water, and transferred into sterile zip-locked bags, homogenized by mixing and flattened to be stored at −80°C until DNA extraction. When possible, other layers (1 to 3, 3 to 5, 5 to 10, and 10 to 15 cm) were also collected from sediment cores. An empty zip lock bag from the stock used served as a blank sampling and extraction control in several stations along each cruise.

5.4.2. Nucleic acid extractions, PCR amplification, and illumina sequencing

For each of the 320 surface sediment samples collected in abyssal plains of the deep_sea project, we extracted the total RNA and DNA contents of c.a. 2 g of material as in (Lejzerowicz et al. 2015), and we generated cDNA from deoxyribonuclease-treated RNA as in (Pawlowski

et al. 2014a). We controlled that no carried-over DNA molecules remained in the RNA extracts based on the absence of PCR products after 60 cycles. We amplified by PCR the V9 hypervariable region of the ribosomal 18*S* gene with the following primer pair: the forward 1389F (5′-TTGTACACACCGCCC-3′) and the reverse 1510R (5′-CCTTCYGCAGGTTCACCTAC-3′) as designed in (Amaral-Zettler et al. 2009). Tag-encoded versions of the primers (a unique 8-nt sequence was added in the 5′ end of each primer) were used to multiplex up to 40 samples per sequencing library. Each sample was amplified in duplicate PCR reactions, and each PCR was performed in a total volume of 25 µl as follows: 19.4 µl of H_2O, 2.5 µl buffer (FastStart, Roche), 0.5 µl of bovine serum albumin (20 mg/ml; Invitrogen Ultrapure), 0.5 µl of 10 mM dNTPs (deoxyribonucleotide triphosphate) (Roche), 0.1 µl of FastStart DNA Polymerase (5 U/µl; FastStart, Roche), 0.5 µl of forward and reverse primers at 10 mM, and lastly, 1 µl of DNA or RNA template (or 1.5 µl for some samples that did not amplify with 1 µl). All DNA and RNA samples were measured using the double-stranded DNA (dsDNA) High-Sensitivity Assay Kit and the RNA High-Sensitivity Assay Kit on the Qubit 4 fluorometer (Thermo Fisher Scientific) and diluted at 7 ng/µl prior PCR amplification. The PCR reaction conditions were as follows: predenaturation step at 94°C for 3 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 57°C for 1 min, extension at 72°C for 1.5 min, and a final extension at 72°C for 2 min. A PCR-negative control for each unique combination of tag-encoded primers was verified by agarose gel electrophoresis. The two PCR replicates for each sample were combined and quantified using high-resolution capillary electrophoresis (QIAxcel System, QIAGEN). The PCR products were pooled in equimolar concentration within each multiplexed library. Each pool of PCR products was purified using a High Pure PCR Product Purification kit (Roche), following the manufacturer's instructions. The sequencing libraries were prepared using the TruSeq DNA PCR-Free Library Preparation Kit (Illumina), following the manufacturer's instructions. The libraries were quantified by quantitative PCR (qPCR) using the Kapa Library Quantification Kit for Illumina Platforms (Kapa Biosystems) and sequenced on a MiSeq instrument (Illumina) using paired-end sequencing for 300 cycles with kit v2.

Within the project eDNAbyss, DNA extractions were performed on about 10 g of sediment using the PowerMax Soil DNA Isolation Kit (QIAGEN, Hilden, Germany), following the manufacturer protocol, except for the last step where incubation of the elution buffer was prolonged 10 min on the spin filter membrane to increase the DNA yield. The first solution of the kit was poured into empty field control ziplock bags, before being extracted along with sediment samples, following the exact same protocol. All DNA extracts were then stored at

−80°C (and transported to Genoscope on dry ice) until PCR amplifications. The V9 hypervariable region of the 18*S* ribosomal RNA (rRNA) gene was amplified by PCR using the same primer pair (1389F and 1510R). Each sample was amplified in triplicates, and each PCR reaction was performed in a total volume of 25 µl with the Phusion High-Fidelity PCR Master Mix with GC buffer (Thermo Fisher Scientific), 0.4 µM final concentration of each primer, 3% of dimethyl sulfoxide, 1× Phusion Master Mix, and 2.5 ng of template DNA (less for few extracts with very low DNA concentration). The PCR reaction conditions were as follows: predenaturation step at 98°C for 30 s, followed by 25 cycles of denaturation at 98°C for 10 s, annealing at 57°C for 30 s, extension at 72°C for 30 s, and a final extension at 72°C for 10 min. PCR products were purified using 1.8× AMPure XP beads cleanup (Beckmann Coulter Genomics). Aliquots of purified amplicons were then run on an Agilent Bioanalyzer using the DNA High Sensitivity LabChip kit to check their lengths and quantified with a Qubit Fluorometer to check their quality and concentration. Amplicons generated were then used for preparation of sequencing libraries. Amplicons (100 ng) were directly end-repaired, A-tailed, and ligated to Illumina adapters on a Biomek FX Laboratory Automation Workstation. Library amplification was then performed using a Kapa Hifi HotStart NGS library Amplification kit with the same cycling conditions applied for previous steps and cleaned up by AMPure XP purification (1 to 1 volume). All libraries were then quantified first by Quant-it dsDNA HS (high-sensitivity) assay using a Fluoroskan Ascent instrument (Thermo Fisher Scientific) and then by qPCR with the KAPA Library Quantification Kit for Illumina Libraries (Kapa Biosystems) on an MXPro instrument (Agilent Technologies). Library profiles were checked using high-throughput microfluidic capillary electrophoresis system (LabChip GX, PerkinElmer, Waltham, MA). Libraries were then 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 no. 15006165). PhiX DNA spike-in was adapted for some libraries (20% instead of 1%) to minimize the loss of data due to low nucleotide diversity at the beginning of the sequencing run. Libraries were sequenced on HiSeq4000 or HiSeq2500 instruments (Illumina) on a paired-end mode. The raw sediment sequencing data have been deposited to the European Nucleotide Archive (ENA) under project accessions PRJEB33873 (eDNAbyss) and PRJEB48517 (deep_sea).

5.4.3. Public 18*S*-V9 rDNA sequencing datasets

We gathered published datasets (table S2) targeting the V9 hypervariable region of the eukaryotic 18*S* rRNA gene and using the same PCR primers pair (1389F and 1510R) used here

for the DOS samples. These datasets were produced by studies sampling the euphotic/aphotic zones (Lie et al. 2014; de Vargas et al. 2015; Xu et al. 2017; Ibarbalz et al. 2019; Obiol et al. 2020; Vargas et al. 2020) and the DOS (Pawlowski et al. 2011; Lie et al. 2014).

5.4.4. Environmental variables

Although some environmental variables and sediment descriptors were collected during the oceanic expeditions from which we collected sediment samples, their heterogeneity led us to extract more homogeneous environmental layers from the Global Marine Environment Datasets (http://gmed.auckland.ac.nz) to standardize our concatenated dataset across multiple studies. These variables included the surface calcite (calcite, in mole per cubic meter), surface nitrate (nitrate, in micromole per liter), surface silicate (silicate, in micromole per liter), surface phosphate (phosphate, in micromole per liter), average photosynthetically active radiation (PAR_mean, in Einstein per square meter per day), surface pH, average sea surface temperature (sst_mean, in celcius), variation in sea surface temperature (sst_range, in celsius), average surface currents strength (srf_current, in meter per second), primary production (primprod, in $mgC \cdot m^2/day/cell$), average stock of particulate inorganic carbon (PIC_mean, in mole per cubic meter), average stock of POC (POC_mean, in mole per cubic meter), total suspended matter (tsm_mean, in grams per cubic meter), seabed slope (slope, degree), seabed nitrate (sb_nitrate, in micromole per liter), seabed silicate (sb_silicate, in micromole per liter), seabed-dissolved oxygen (sb_o2dissolve, in milliliter per liter), seabed-utilized oxygen (sb_o2utilized, in milliliter per liter), seabed temperature (sb_temp, in celsius), seabed salinity [sb_salinity, practical salinity scale (PSS)], and average temperature in the water column (wat_col_temp, in celsius). We also extracted the estimated POC export at 100 m depth below the surface (POC_export, g C_{org}/m^2 per year) (Henson et al. 2011) and the POC fraction reaching the seafloor (POC_seafloor, g C_{org}/m^2 per year) (Lutz et al. 2007; Sweetman et al. 2017). The values of each environmental variable for each sample analyzed in this study were extracted from the environmental layers with their Global Positioning System (GPS) coordinates (table S7).

5.4.5. Raw sequencing data processing

For the deep_sea dataset, the sequencing libraries were demultiplexed using Double Tag Demultiplexer (DTD) software (https://github.com/yoann-dufresne/DoubleTagDemultiplexer) to screen the R1 and R2 files of each library and retrieve unique tag-encoded primer combinations associated to each sample (allowing no mismatches). We thus produced pairs of

fastq files for each sample. For all other illumina datasets (see table S2), we obtained at least one pair of fastq files (paired-end) per sample (some samples were sequenced several times to obtain enough reads that were subsequently merged before statistical analysis). For the datasets produced with the 454 sequencing technology, we obtained one fastq file per sample. We used two R scripts (for paired-end illumina datasets and for 454 dataset, see the "rds_pipeline_illumina.R" and "rds_pipeline_454.R" scripts) to process all the fastq files by batch of 10 samples per job on a High-Performance Computing cluster (Baobab, University of Geneva). The R scripts implemented the key steps of the DADA2 workflow (Callahan et al. 2016) and additional quality filtering steps (see below). The scripts performed the quality filtering with the filterAndTrim function of the DADA2 v1.12.1 R package with default settings, the trimming of primers using the cutadapt v2.4 software (Martin 2011), the filtering of any read that still contain traces of primers (*fastqFilterPrimersMatchs* function in the fastqUtils.R script), the filtering of any read below 20 bp (*fastqFilterWidth* function in the fastqUtils.R script), the training of errors models using the *learnErrors* function of DADA2 with default settings, the inference of ASVs using the *dada* function with default settings (but for the 454 data, for which we used the HOMOPOLYMER_GAP_PENALTY = −1, BAND_SIZE = 32 options, as recommended by the DADA2 package developing team), and the merging for the overlapping paired-end reads using the *mergePairs* function with the option "trimOverhang." Last, we exported the output of the DADA2 workflow, i.e., the ".rds" files that contain all the ASV sequences and their counts for each sample. We also collected summary statistics on each processing step, the trained errors models, and processing time for each sample.

5.4.6. Combining the datasets into a single ASV-to-sample table, taxonomic and functional annotations, matrix curation

We reimported the rds files into R to build an ASV-to-sample table. We first produced an ASV table per dataset, filtered chimeric ASVs with the option "consensus" within each dataset, and aggregated the replicated libraries per biological sample for the deep_sea, tara, and tara_polar datasets. We also aggregated the reads obtained from DNA and RNA libraries generated for each deep_sea sediment samples, since a comparative analysis revealed that the diversity and biogeographic patterns of eukaryotic communities are mostly similar between DNA and RNA [fig. S12; in line with (Guardiola et al. 2016; Lejzerowicz et al. 2021)]. We filtered the ASVs detected in the negative controls of the eDNAbyss dataset across the eDNAbyss ASV table. Last, we concatenated each dataset-based ASV-to-sample table into a single one and exported

all the ASV sequences into a fasta file for taxonomic annotations. We used the "assignment-fasta-vsearch" module of the SLIM v0.6 software (Dufresne et al. 2019) that wraps the vsearch v2.2.2 software (Rognes et al. 2016). The ASVs were compared to a custom version of PR^2 (de Vargas et al. 2015) that focus on the 18*S* V9 region and that include functional annotations (available at https://doi.org/10.5281/zenodo.3768950) and with the SILVA v138 database (Quast et al. 2013). Taxonomic annotations were the consensus among up to three candidate reference sequences that are above 85% similarity with the query or directly assigned to the reference sequence if the query had a similarity of at least 99%. We also performed another search without restricting a minimum similarity threshold to match an entry reference sequence in the custom V9 version of PR^2, to identify non-18*S* V9 sequences. We focused our analysis on the eukaryotic diversity by discarding any prokaryotic, plastidic ASVs, or any other artifactual ASV. We used taxonomic annotations obtained with the SILVA database to discard prokaryotic ASVs and the annotations obtained with PR^2 to discard organelle-derived ASVs. All ASVs that only loosely match any V9 reference sequence (i.e., <20% similarity) were considered as non-18*S* V9 sequence and were discarded. We also filtered ASVs that did not contain the "GTCG" motif in the first four nucleotides in the 5′ end. This motif is widely conserved across eukaryotes, whereas prokaryotes have a "GTCA" motif highly conserved in those positions. We lastly used the length distribution of eukaryotic and prokaryotic ASVs to discard any possible prokaryotic unassigned ASVs (filter set at 116 bp; fig. S13). For downstream taxonomic analyses, we used PR^2-based annotations using the 85% minimum similarity threshold. We also inferred the trophic mode of pelagic ASVs (phototrophic/photosymbiotic/parasitic/heterotrophic protists and zooplankton/other metazoan) based on the matching candidate reference sequences in PR^2 (up to three candidates per queried ASV). We ascribed our ASVs to functional groups only if the functional attributes across candidates above 95% similarity were unambiguous, i.e., all the candidates for taxonomic assignment share similar functional attributes. We also used the different size fractions of the *Tara* Oceans samples to infer the size of pelagic ASVs, by using a weighted average of relative abundances across the size fractions of plankton samples (using the lowest mesh size, e.g., from the 20- to 180-µm size fraction, we used 20 µm in our calculation for this size fraction).

5.4.7. Classifying eukaryotic ASVs into pelagic or benthic taxa

We considered the ASVs being detected in pelagic samples as planktonic (or nektic), the ASVs detected exclusively in sediment samples as benthic, and the ASVs detected in both pelagic

and sediment datasets as sinking plankton [although 29 vertebrates ASVs, comprising most of the nekton, represented ~1.29% of the sequences detected in both pelagic and sediment samples (table S3), we hereafter refer only to sinking plankton]. However, because multiple benthic groups have meroplanktonic larvae and hence could be detected in pelagic samples, we manually curated the ASVs assigned to metazoans within the sinking plankton fraction, based on their known lifestyles. This was, for instance, the case for some polychaetes, molluscs, echinoderms, or harpacticoid copepods that were "forced" into benthic diversity but not for pteropods that were left in the sinking plankton. Of the 546 metazoan ASVs in the sinking plankton fraction, 224 were curated as benthic (see table S8 for the details of this manual curation).

5.4.8. Eukaryotic community diversity and structural analysis

For a and b diversity analyses, we used functions of the vegan R package v2.5-3 (Oksanen et al. 2007), unless specified differently. Because the size fractionation of the pelagic samples from the *Tara* Oceans datasets has a strong effect on a and b diversity measures, we compared eukaryotic diversity patterns across pelagic and benthic realms by considering only the richest nano- (3 to 20 µm) and pico- (0.2 to 5 µm) size fractions of pelagic samples. The eukaryotic ASV accumulation curves as a function of sampling effort were computed with the *specaccum* function with the "random" method. We calculated the Shannon diversity for each sample and compared the distribution of sample diversity across both pelagic euphotic and aphotic with the strictly benthic diversity using the *stat_compare_means* function of the ggpubr R package v0.2.5 (Kassambara A. 2020) with default settings. For b diversity analysis, we removed samples with less than 1000 reads and discarded ASVs represented by less than 100 reads throughout the dataset. We then normalized the ASV-to-sample matrix with the cumulative sum scaling (CSS) method (Paulson et al. 2013a) and computed a Bray-Curtis dissimilarity matrix between pairs of samples. The dissimilarity matrix was used to perform a nonmetric multidimensional scaling (NMDS) ordination on two axes. Sampling depth and absolute latitude variables were fit to the NMDS as smooth surfaces using the *ordisurf* function. The dissimilarity matrix was also used as input of the *adonis* function for PERMANOVA models testing for differences between eukaryotic compositional structure between realms (pelagic euphotic, pelagic aphotic, and sediment) and along a gradient of absolute latitude (nested in type of realm and restricting permutations within type of realm with the "strata" option), using 999 permutations. Last, we measured the b diversity dispersion within each realm using the *betadisper* function and compared the distances distribution to

group centroids between realms using the *stat_compare_means* function of the ggpubr R package.

For a diversity and b diversity analyses of the deep-ocean benthic communities, we focused on oceanic samples only, i.e., we did not consider the samples from the Mediterranean Sea nor the ones from the Gulf of California, to avoid potential effects from coastal ecosystems. We calculated the normalized ASV richness per sample for the overall benthic communities and for selected benthic groups (nematodes, foraminifera, platyhelminths, polychaetes, molluscs, and ciliates) by rarefying each benthic sample at the lowest remaining sequencing depth (after removing planktonic ASVs and after focusing on a given benthic taxonomic group). We used generalized additive models (GAMs) to investigate the possible nonlinear variation of richness and Shannon diversity along gradients of latitude, primary production, and POC export from the surface and reaching the seafloor using the *gam* function of the mgcv R package (https://cran.r-project.org/web/packages/mgcv/) and the smoothing parameter set to 3. For b diversity analyses, we used a similar approach than detailed above (CSS-normalized and Bray-Curtis dissimilarity matrix), although here, we did not filter rare ASVs. We used the *pcoa* function of the ape R package (Paradis and Schliep 2019) to perform a principal coordinate analysis of the Bray-Curtis dissimilarity matrix and calculate the structural variation explained by the first two axis of the ordination. We used the *ordisurf* and *envfit* functions to respectively fit the absolute latitude and a selection of environmental variables (seabed variables: salinity, temperature, silicate, nitrate, dissolved oxygen, POC reaching the seafloor, and pelagic variables that connect the surface to the DOS, namely, the primary productivity and the POC export from the surface) to the ordination. PERMANOVA models were used to test for differences in benthic composition between abyssal postulated biogeographic provinces (Watling et al. 2013) and along a gradient of absolute latitude using 999 permutations. Then, we used the selected environmental variables in a stepwise model building for constrained ordination (distance-based redundancy analysis) using the *ordi2step* function in a forward direction and using 999 permutations, to explain the observed benthic community structure. We calculated the proportion of shared ASVs between pairs of benthic samples to investigate the decrease of shared ASV proportion as a function of increasing spatial distance (calculated from GPS coordinates, see the "companionFunctions.R" script). We also calculated key distance-decay parameters as in (Soininen et al. 2007), i.e., the initial similarities (Sørensen similarities between pairs of samples distant to each other by less than a kilometer), the slope of distance-decay relationship (here in a log-linear regression form), and the halving distances, i.e., the spatial distance after which the initial similarities are halved. We calculated these

parameters on an average Sørensen dissimilarity matrix calculated over 10 rarefaction draws at the minimum sequencing depth possible (the sequencing depth of the sample with the lowest number of reads) and by considering the full benthic community or by focusing on selected benthic groups only, e.g., polychaetes, molluscs, or platyhelminths (macrofaunal size classes); nematodes or foraminifera (meiofauna); and amoebae or ciliates (microbes). We used the *mantel* function to test for correlation between spatial distance and community dissimilarities using 999 permutations. We used the *betadisper* function to calculate the b dispersion of benthic communities at increasing sampling spatial scale (between replicates samples of a sediment core, between cores of the same deployment, between deployments at a given station, and within a given abyssal basin). Last, we aggregated all samples at the station scale and fitted neutral community assembly models as in (Sloan et al. 2007) to investigate whether the distribution of ASVs within the pelagic and benthic realms are less or more geographically widespread than expected by neutral models.

We compared the inferred functional attributes (size and trophic mode) of sinking pelagic ASVs with their nonsinking counterparts to explore whether these traits could explain their transfer to the DOS. We also explored the variation of functional groups and size classes of the sinking planktonic communities in the sediment along the gradient of latitude. We investigated the spatial pattern of planktonic abundance on the seafloor by fitting a GAM on the proportion of planktonic DNA reads in the sediment as function of latitude (with smoothing parameter set to 3). Then, we aggregated the planktonic DNA reads of all sediment samples at the station scale and used random forest models to predict the POC export from the surface and the POC reaching the seafloor in a leave-one-out cross-validation approach. We used the *ranger* function of the ranger R package (Wright and Ziegler 2017)in a regression mode, growing 300 trees and setting the "mtry" parameter at one-third of the total number of features (number of sinking pelagic ASVs). Linear models were used to measure the performance of predictive models. Last, we used a sparse partial least square regression [mixOmics R package (Rohart et al. 2017)] to identify the pelagic ASVs detected in the sediment that are best correlated with the variation of POC export and POC reaching the seafloor and with primary productivity and latitudes. We then focused on the pelagic ASVs that were reported with a correlation coefficient above 0.3 with POC export and POC reaching the seafloor and presented them in a clustered heatmap.

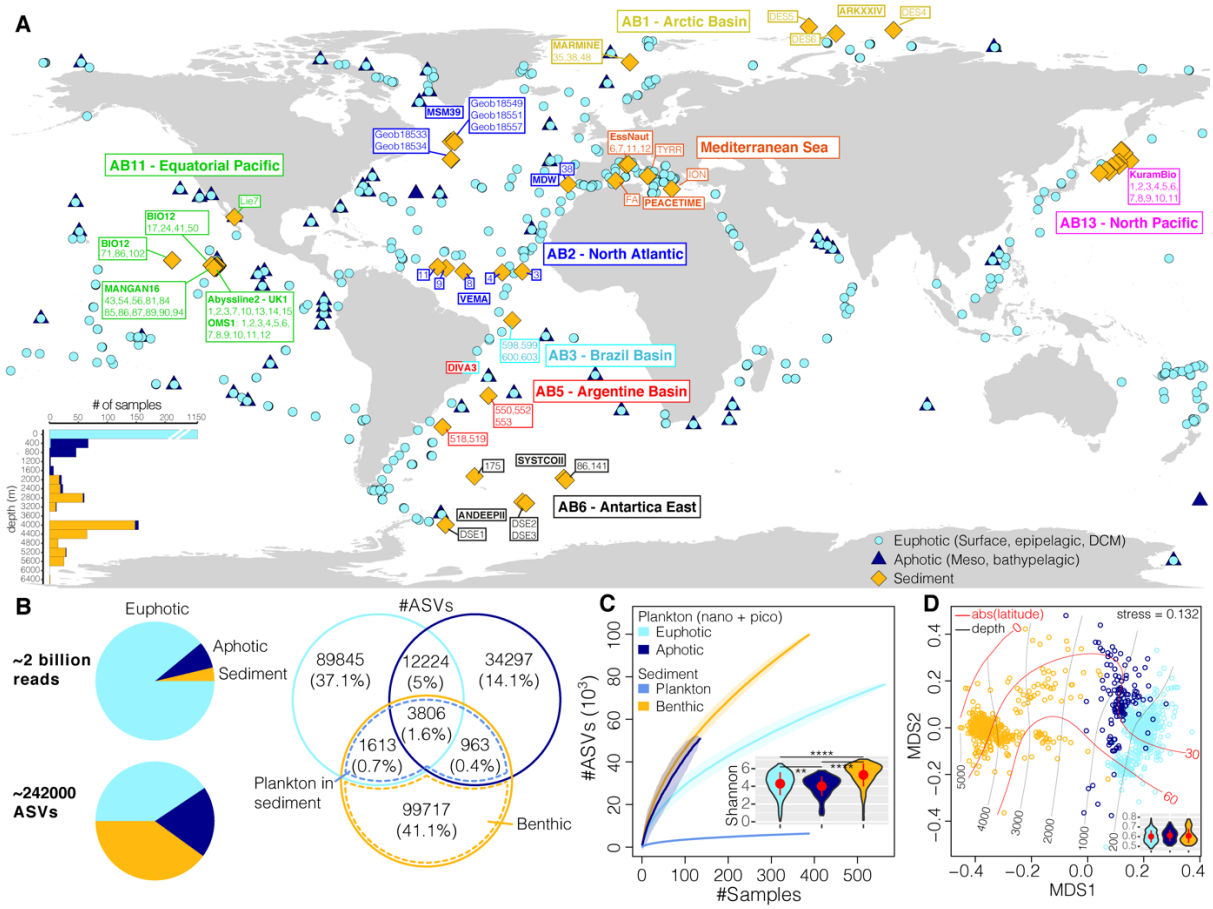

Figure 5.1: ***Eukaryotic ribosomal DNA diversity in the pelagic euphotic and aphotic zones, and on the deep ocean floor.*** *A) Geographic distribution of the stations (n = 447) from which the samples (n = 1,685) analyzed in this study were collected. The color of the sediment sampling stations tags indicates approximate correspondence with Abyssal Provinces* (Watling et al. 2013)*, and the names of deep-sea cruises is indicated in bold or within text boxes. The bottom left inset represents the depth distribution (in meters) of the samples. B) Number of eukaryotic 18S V9 rDNA reads and of Amplicon Sequence Variants (ASVs) for the three sampled realms. A Venn diagram represents the distribution of ASVs richness and their proportions within and across the three realms. The intersection of the pelagic and sediment datasets is used here to separate the indigenous benthic organisms from the sinking plankton. C) ASV accumulation curves as a function of sampling effort for the pelagic euphotic, pelagic aphotic, sinking plankton and for the benthic eukaryotes. For pelagic realms, we calculated the curves by focusing only on the nano and pico planktonic size fractions (see methods). The bottom right inset indicates the distribution of Shannon diversity for pelagic and benthic communities. The red dots and bars within violin plots represent means and standard deviations and horizontal black bars indicate significant differences (Wilcoxon tests, **: $p < 0.01$, ****: $p < 0.0001$). D) Non-metric Multidimensional Scaling analysis of the Bray-Curtis dissimilarity matrix computed from the pelagic (only the nano and pico fractions) and sediment datasets normalized with the Cumulative Sum Scaling method. The red and black lines on the ordination represent respectively*

the absolute latitude and depth as fitted surfaces to the ordination. The stress value of the NMDS is indicated on the plot. The bottom right inset represent the community dispersion within each realm (i.e. distribution of Bray-Curtis distances to the group centroid, higher values indicate more compositional variation). The red dots and bars within violin plots represent means and standard deviations and no significant differences between realms were detected (Wilcoxon tests, $p > 0.05$).

5.5. Results and discussion

5.5.1. Eukaryotic diversity from the ocean surface to the DOS

We obtained a total of 242,465 eukaryotic ASVs represented by ~1.95 billion DNA reads (Fig. 5.1 B). Only 3806 (1.6%) of these ASVs were detected in all three realms, while 6382 pelagic ASVs were detected in DOS. These ASVs were assumed to correspond to sinking pelagic organisms, mainly plankton, although 29 ASVs (representing 1.29% of the reads of these ASVs) could be ascribed to nekton (e.g., dead vertebrates), which also contribute to the downward flux of organic matter. From the metazoan fraction of sinking pelagic organisms, we curated benthic animals with known meroplanktonic larvae (224 ASVs; see Materials and Methods). The number of ASVs found exclusively in DOS, here assumed to correspond to indigenous deep-sea benthic organisms, was comparable to that found in the pelagic realms, although there were 25 times more pelagic DNA reads in our meta-dataset (Fig. 5.1 B). To account for this variation in sequencing effort, we subsampled each aggregated dataset per realm 1000 times at identical sequencing depths (1 Mio reads) and analyzed the diversity of ASV, together with their distribution and abundances within and across the pelagic (euphotic and aphotic zones) and DOS (sinking pelagic and benthic organisms; fig. S1) realms. This indicated that, although nearly half of eukaryotic DNA reads represent sinking planktonic ASVs, the ASV richness in the DOS could be more than three times higher than in pelagic habitats, with more than 60% of ASVs being exclusively benthic.

The unique size fractionation of the pelagic samples from the *Tara* Oceans dataset has a strong effect on a diversity (richness and evenness) and b diversity (compositional variation) measures (fig. S2) (de Vargas et al. 2015). The samples from the micro- (20 to 180 µm) and meso- (180 to 2000 µm) plankton collected by 20- and 180-µm net tows concentrate mostly on copepods and collodarians that lower a diversity and inflate b diversity measures when compared to other pelagic samples. Since these taxa were also detected in the lower plankton size fractions, we confined our analyses to the richest nano- (3 to 20 µm) and pico- (0.2 to 5 µm) plankton

fractions to compare a and b diversity patterns across *Tara* Oceans, Malaspina, and DOS samples.

Both the ASV accumulation curves as a function of sampling effort and Shannon diversity values confirmed that benthic eukaryotic diversity is much higher than that in the water column (Fig. 5.1 C). The benthic accumulation curve is similar to that obtained for the pelagic aphotic zone, which may indicate that diversity in aphotic waters is also very high and largely undersampled [but see (Pernice et al. 2015)]. Clustering of eukaryotic communities by their compositional similarity revealed a clear separation of the pelagic and DOS realms and changes along a gradient of absolute latitude (Fig. 5.1 D and fig. S3). This was confirmed by permutational multivariate analysis of variance (PERMANOVA), both for type of realm (R^2 = 0.144, $P < 0.001$) and for absolute latitude ($R^2 = 0.031$, $P < 0.001$). The degree of eukaryotic community differentiation (b diversity dispersion) within each realm was similar (Wilcoxon test, $P > 0.05$; Fig. 5.1 D). Our results therefore indicate that eukaryotic communities of the DOS are both more diverse and sharply different compared to those of pelagic realms.

The taxonomic compositions of eukaryotic assemblages were clearly different in the pelagic euphotic, pelagic aphotic, and DOS realms (Fig. 5.2 A; see fig. S4 for relative abundances and table S3 for details). While diversity in the euphotic zone is dominated taxonomically by Alveolata (30.8%), notably dinoflagellates (Dinophyceae, 19.4%), the aphotic zone is extremely rich in Diplonemea [46.5%; see (Lara et al. 2009; Morgan-Smith et al. 2013; Flegontova et al. 2016)], mainly heterotrophic nanoflagellates in the family Eupelagonemidae (46.2%). The taxonomic composition of the deep, exclusively benthic, eukaryotic assemblage is very different, comprising various groups that do not occur or rarely occur in the water column (fig. S5 and table S3), e.g., Dactylopodida amoebae (6.5%), Chromadoria nematodes (5.3%), Monothalamid foraminifera (4.4%), and Oligohymenophorea ciliates (3.7%). Nearly two-thirds (60.1%) of the benthic eukaryotic ASVs (representing 47.8% of the reads) could not be taxonomically annotated using current reference taxonomic databases and a similarity cutoff of 85%, and many of them matched a reference sequence with less than 80% similarity (Fig. 5.2 B). By comparison, the proportion of unassigned ASVs in the pelagic samples is 24.7% (2.6% of the reads) in the euphotic zone and 13.9% (4.1% of the reads) in the aphotic zone (Fig. 5.2 C and table S3).

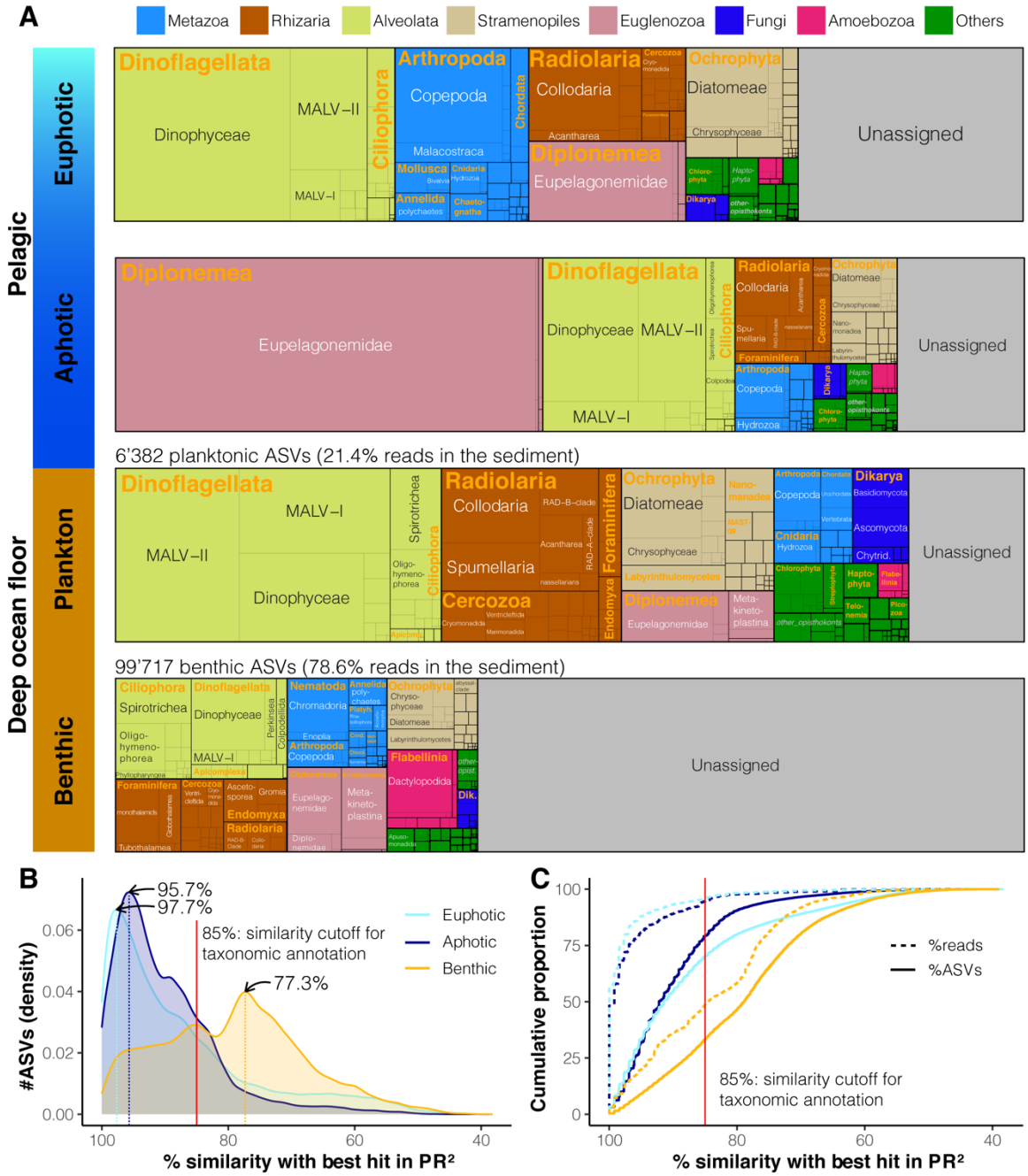

Figure 5.2: ***Taxonomic composition of eukaryotes (ASVs richness) in the pelagic euphotic, pelagic aphotic and deep-ocean floor (sinking plankton and benthic communities) realms****. A) ASVs richness of eukaryotic groups (see figure S4 for relative abundances and Table S3 for details). The number of ASVs and their relative abundance in the sediment is shown for ASVs of pelagic origin, as well as those derived from indigenous benthic taxa. B) Number of ASVs (represented as density) as a function of similarity with the best hit with a reference sequence in the PR² database. The peaks in density for each realm are highlighted on the plot and the corresponding similarity level with reference sequences is indicated. C) Cumulative proportion of ASVs and reads abundance as a function of similarity with best hit with a reference sequence in PR². The vertical red lines indicate the similarity cutoff (85%) for taxonomic annotation used in this study.*

To better characterize the taxonomic breadth of the unknown eukaryotic diversity in the ocean, we clustered all unassigned eukaryotic ASVs into operational taxonomic units (OTUs) at decreasing similarity thresholds (fig. S6). This revealed that more than 10,000 benthic OTUs are formed with a 90% similarity cutoff, well below the species/genus threshold levels (Forster et al. 2019). These results indicate that previously unknown high-rank eukaryotic groups with diverse and abundant sublineages likely make up most of the diversity thriving in DOS. A similar number of 90% cutoff OTUs is formed in the pelagic euphotic zone, but their relative abundance (2.6% of the reads) is much lower than for benthic diversity (47.8% of the reads). Many of these unassigned pelagic ASVs may thus correspond to rare unknown eukaryotes or rare intraspecific/intragenomic variants of known eukaryotes with unusually high polymorphism (Zhao et al. 2019). Among the known taxa (39.9% of benthic ASVs and 52.2% of benthic reads), our data for selected typical deep benthic macrofaunal and meiofaunal groups show that some are relatively well represented in the current databases (e.g., polychaetes and nemerteans; fig. S7), while others remain poorly represented (e.g., foraminifera and nematodes).

5.5.2. Biogeography of deep-ocean benthic eukaryotes

Analysis of the strictly benthic eukaryotic diversity revealed global biogeographic patterns among DOS communities. The overall richness of benthic ASVs tends to decrease with increasing latitude (fig. S8). Benthic richness follows a bell-shaped trend with increasing export flux of particulate organic carbon (POC) from the surface and particularly with increasing POC reaching the seafloor (the latter explaining up to 10.7% of the variation in overall benthic richness). This pattern is not consistent across benthic groups, with nematodes, foraminifera, and molluscs being notably more diverse at higher latitudes and at sites with higher POC flux reaching the seafloor (fig. S8). The compositional structure of deep benthic communities is in broad agreement with abyssal biogeographic provinces (Watling et al. 2013) (PERMANOVA $R^2 = 0.136$, $P < 0.001$) and somewhat structured along a gradient of absolute latitude at a global scale ($R^2 = 0.051$, $P < 0.001$), although polar regions are separated on the ordination (Fig. 5.3 A). We used a selection of environmental parameters (see Materials and Methods) in a stepwise model building for constrained ordination to explain the observed pattern. The model explained up to 15.1% of the benthic compositional variation, with seabed nitrate, POC export from the surface, and POC reaching the seafloor together explaining 11.2%

(table S4), in line with previous findings on the role of POC export in shaping benthic prokaryotic communities (Danovaro et al. 2016).

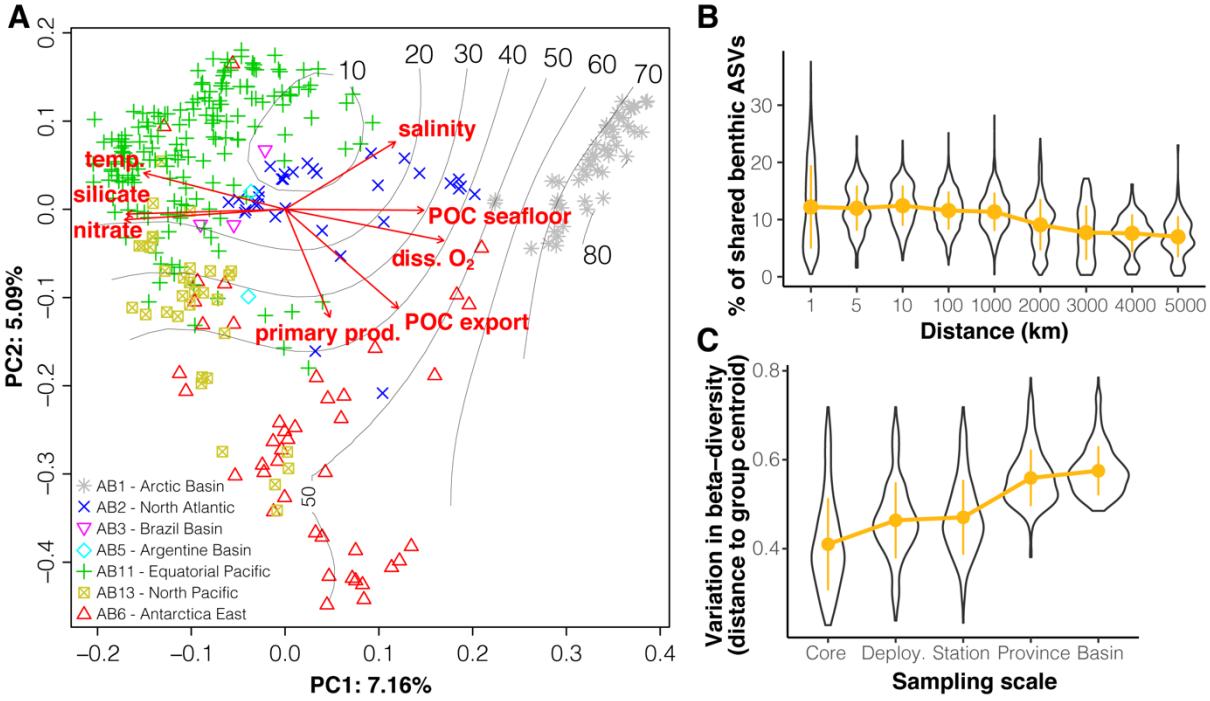

Figure 5.3: ***Biogeography of the deep ocean benthic eukaryotic communities.*** *A) Principal Coordinates Analysis of the Bray-Curtis dissimilarity matrix computed from the reads counts of indigenous benthic ASVs normalized with the Cumulative Sum Scaling method. The proportion of variance explained by the first two axes is indicated on the plot. The grey lines and numbers indicate the absolute latitude as a fitted smooth surface on the ordination, and the red arrows are fitted environmental parameters to the ordination. Colors and symbols indicate location of sampling sites in the abyssal biogeographic provinces (AB1 – AB13) postulated by* (Watling et al. 2013). *B) Proportion of shared benthic ASVs as a function of increasing distance between pairs of samples. The proportion of shared ASVs was computed only within the same oceanic basins. C) Variation in beta diversity, i.e., distribution of samples distances to group centroids, as a function of increasing spatial sampling scale. Higher values indicate higher compositional variation. Orange dots and bars represent means and standards deviation.*

The average proportion of shared ASVs between pairs of samples as a function of their geographical distance was relatively stable up to 10 km, after which it decreased steadily (Fig 5.3 B), possibly indicating dispersal limitation among benthic taxa or shifts in environmental drivers. We calculated key distance-decay parameters for selected benthic groups (table S5), i.e., the rate of decrease in their community similarity with increasing spatial separation. The initial similarities (i.e., similarity at 1 km distance that provides a measure of the local presence/absence patchiness) indicate that benthic macrofaunal groups (molluscs and

polychaetes) tend to have a stronger local turnover than meiofaunal or protistan groups (nematodes, foraminifera, and amoebae). The former also generally have a steeper distance-decay, as indicated by the greater slope of linear models and lower halving distances, i.e., the distance after which the initial similarity is halved (table S5). These results indicate that dispersal limitation or environmental filtering (or a combination of both) may be stronger for macrofaunal benthic organisms than for meiofaunal or microbial eukaryotes, although macrofaunal taxa are usually thought not to be limited in their dispersion, owing to their common planktonic larval phases (McClain and Rex 2015).

We lastly compared spatial structures and distance-decay parameters for whole benthic eukaryotic communities with those for water column communities (table S5). Overall, benthic communities are more spatially structured (mantel: $r = 0.454$, $P < 0.001$) than pelagic communities (euphotic: $r = 0.147$, $P < 0.001$; aphotic: $r = 0.228$, $P < 0.001$) and have lower initial similarity, steeper distance-decay, and smaller halving distances (table S5 and fig. S9), consistent with previous findings for benthic compared to pelagic communities of bacteria in the world ocean (Zinger et al. 2011). This was also shown by fitting neutral community assembly models to each realm dataset, indicating that taxonomic groups of the pelagic euphotic zone tend to be more geographically widespread, while benthic groups tend to be less widespread than expected by neutral models (fig. S10). The importance of benthic community patchiness at local scales is reinforced by the variation of b diversity as a function of increasing sampling scale (Fig. 5.3 C). The compositional variations within single sediment cores (8 to 10 cm in diameter) and between sediment cores collected from the same multicorer deployment (30 cm to 1 m apart) were comparable to that between deployments at a single station. This confirms the high degree of deep benthic community variation at local scales observed before in the case of selected groups of macro- and meiofauna by both morphological (Zeppilli et al. 2016) and DNA-based (Lejzerowicz et al. 2014) studies.

5.5.3. Eukaryotic plankton DNA signature on the DOS

Our study provides the first DNA-based insight into the qualitative and semiquantitative importance of the eukaryotic plankton diversity reaching the DOS at a global scale and thus driving the biological transfer of atmospheric carbon to the seafloor. The taxonomic composition of the 6382 planktonic eukaryotes in the DOS is roughly similar to that in the pelagic euphotic and aphotic zones (Fig. 5.2). In terms of relative abundance, however, planktonic DNA reads from sediment samples are mainly distributed among diatoms (15.4%) and various groups of rhizarians (26.9%) but include relatively few copepods (1.3%) and

dinoflagellates (4.7%), which numerically dominate the plankton in upper ocean layers (fig. S4 and table S3). Plankton DNA on the DOS also include abundant ASVs assigned to the diplonemids (2.3%) and fungi (4.2%) that are common in the aphotic zone, supporting previous findings that the DOS accumulates DNA from organisms occurring throughout the entire water column (Laroche et al. 2020).

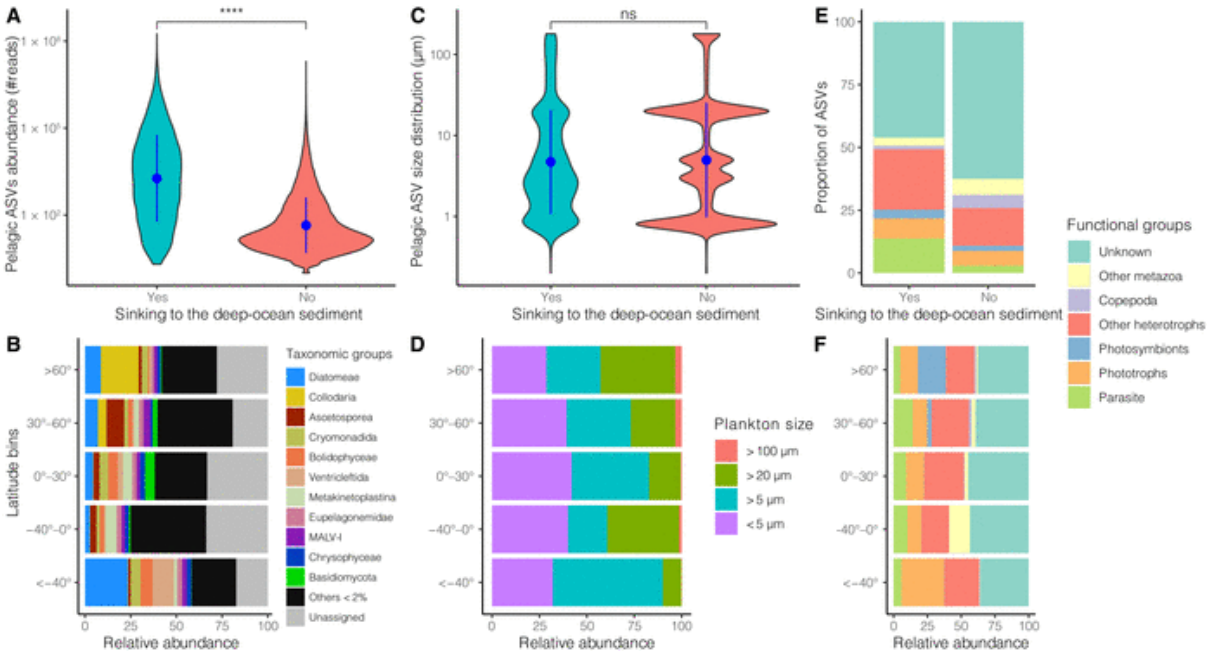

Figure 5.4: ***Abundance and functional attributes of the sinking plankton compared to their nonsinking counterparts and the taxonomic and functional abundance breakdown of plankton DNA in DOSs*** *Comparison of the DNA-based log-transformed read abundance (A), the inferred size distribution (C), and the functional breakdown (E) of pelagic ASV sinking to the DOS with their nonsinking counterparts in pelagic realms (euphotic and aphotic datasets have been combined). Results of Wilcoxon tests in (A) and (C) are indicated as follows: ****P < 0.0001; not significant (ns), P > 0.05. Breakdown of taxonomic (B), inferred size classes (D), and functional (F) relative abundances of plankton DNA in DOS as a function of latitudinal bins.*

The pelagic ASVs reaching the DOS are generally among the most abundant planktonic eukaryotes in the water column (together representing 75.8% of the reads in the euphotic and 79.3% of the reads in the aphotic), although not all abundant pelagic ASVs are present in the DOS (Fig. 5.4 A). We explored whether their occurrence in the DOS could be explained by their size distribution and by their trophic modes. We found no evidence that larger planktonic taxa are more likely to reach the DOS than smaller taxa (Fig. 5.4 C), reinforcing the idea that most sinking plankton is transferred to the sediment through organic aggregates and not as individual organisms. However, the relative abundance of large planktonic taxa was higher in high-latitude sediments, especially in the Arctic (Fig. 5.4 D), consistent with the trend of

increasing sea-surface plankton size with increasing latitude and nutrient content (Barton et al. 2013; Acevedo-Trejos et al. 2018). Notably, the proportion of parasitic protists among sinking pelagic ASVs (13.7%) is greater than among nonsinking pelagic ASVs (2.9%; Fig. 5.4 E), indicating that pelagic parasites are more likely to reach the DOS. Their relatively higher abundance in temperate and tropical latitude sediments suggests their ecological importance at these latitudes (Fig. 5.4 F). Greater transfer of parasites to the DOS could reflect their ability to infect and kill larger hosts and/or the massive amounts of resistant and relatively dense propagules that they typically release after host infection and that could persist in sinking aggregates (Guidi et al. 2016; Boeuf et al. 2019; Preston et al. 2020; Poff et al. 2021).

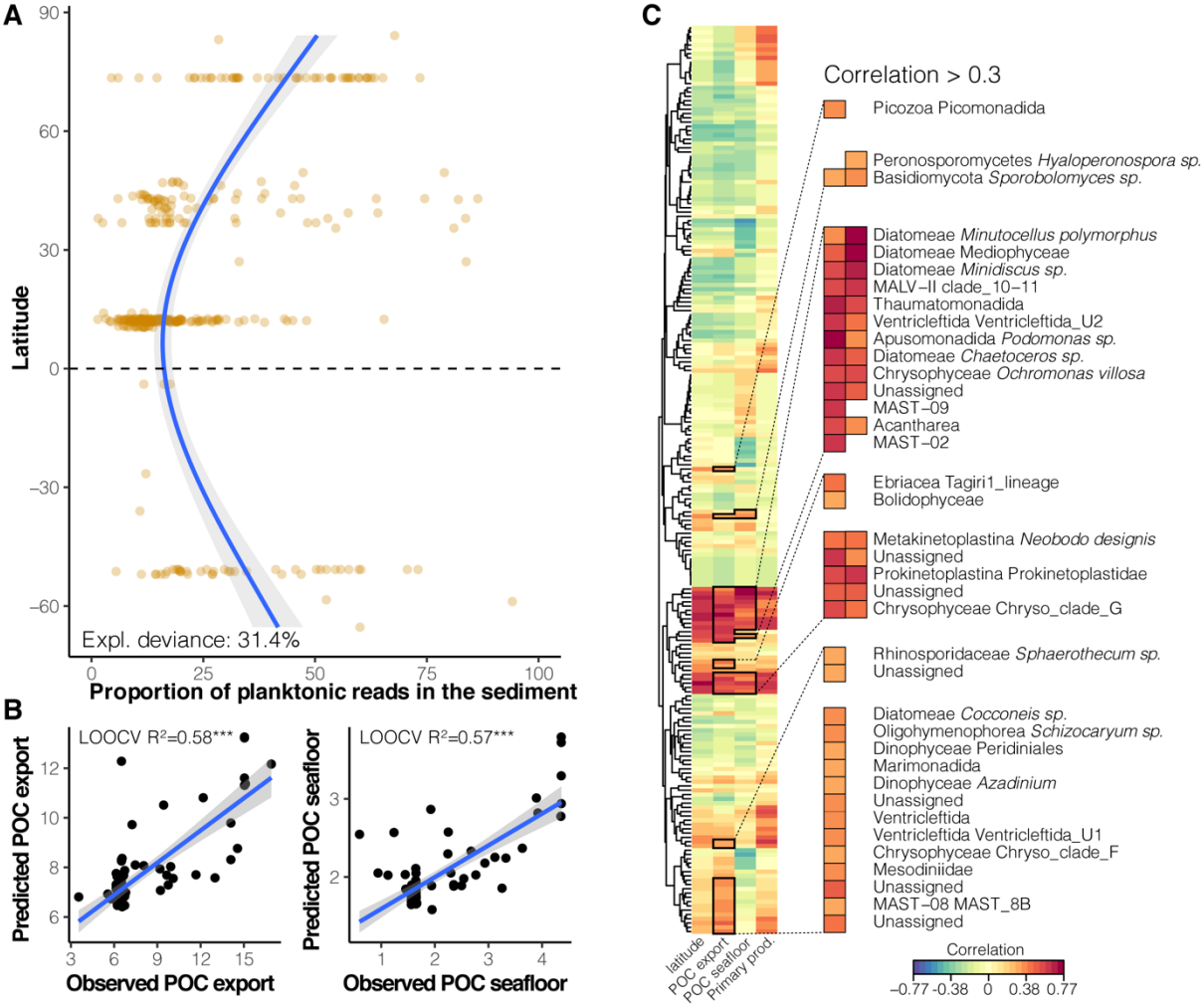

Figure 5.5 : ***Latitudinal abundance of plankton DNA reads in the deep-ocean floor, and eukaryotic taxa associated with particulate organic carbon export from the surface and it fraction reaching the seafloor.*** *A) Proportion of DNA reads representing sinking planktonic taxa in deep sediments as a function of latitude. The solid blue line represents a fitted generalized additive model (s=3), the shades displaying 95% confidence intervals. The explained variation of plankton proportion is indicated on the plot. See figures S11B and S11F for the taxonomic and functional breakdown of plankton DNA on the DOF as function of latitude. B) Prediction of POC export and POC reaching the seafloor from the composition of plankton DNA on the DOF (POC export*

and POC seafloor are expressed in g C_{org} /m^2/year). Random forest regressions were used in a Leave-One-Out Cross Validation (LOOCV) approach at the station scale. Correlation coefficients of linear models are reported on the plots. Blue lines represent linear models (both with p < 0.001) and shades represent 95% confidence intervals. C) Sinking planktonic ASVs associated with the POC export and POC reaching the seafloor identified with a sparse partial least square regression. Pelagic ASVs strongly associated (correlation > 0.3) with POC export and POC reaching the seafloor are detailed on the clustered heatmap (details in Table S6).

Comparison of the DNA-based log-transformed read abundance (**A**), the inferred size distribution (**C**), and the functional breakdown (**E**) of pelagic ASV sinking to the DOS with their nonsinking counterparts in pelagic realms (euphotic and aphotic datasets have been combined). Results of Wilcoxon tests in (A) and (C) are indicated as follows: ****$P < 0.0001$; not significant (ns), $P > 0.05$. Breakdown of taxonomic (**B**), inferred size classes (**D**), and functional (**F**) relative abundances of plankton DNA in DOS as a function of latitudinal bins.

We aggregated our data for each entire realm to investigate whether the most abundant sinking pelagic ASVs in the water column are also the most abundant among sinking pelagic ASVs detected in the sediment, providing insight into their overall taphonomy (fig. S11). Similar abundance profiles would indicate that the structure of sinking plankton assemblages is overall preserved in surface sediment, whereas dissimilar profiles would indicate that sinking assemblages are consumed or repackaged during their downward transfer in a nonrandom manner. For instance, the abundance profiles of sinking pelagic ASVs of copepods, dinoflagellates, diatoms, acanthareans, eupelagonemids, hydrozoans, and spumellarians in both the euphotic and aphotic zones are similar to their abundance profiles in the DOS (fig. S11), suggesting that their structure in pelagic ecosystems is preserved on the seafloor. Their profiles in the DOS better mirrored those in the aphotic zone (higher R^2 of linear models), suggesting that the transformation of sinking material occurs mostly in the upper oceanic layer. However, this was not the case for diatoms and collodarians, for which the abundance profiles in the euphotic and aphotic layer were similarly preserved in the DOS, likely because of their higher propensity to sink and form aggregates (Agusti et al. 2015; Biard et al. 2017). Although plankton DNA transfer to sediment has yet to be investigated [but see (Kirkpatrick et al. 2016; Barrenechea Angeles et al. 2020; Armbrecht et al. 2021a)], notably by accounting for physical processes (e.g., deep waters currents and vertical mixing) that interact with biological and ecological processes in the deposition of sinking material, our study reinforces the significance of the DOS as a DNA archive of upper-ocean biodiversity and ecology and a source of potential new proxies to document past environmental changes (de Schepper et al. 2019).

Last, we attempted to correlate our global DOS plankton biodiversity dataset to yearly average POC export from the surface and the fraction of it reaching the DOS, as estimated, respectively, by thorium-derived export measurements, modeled at a global scale (Henson et al. 2011), and by evaluating the efficiency of POC transfer through the water column based on sediment trap POC flux data, net primary production estimates, and sea surface temperatures (Lutz et al. 2007). Overall, the pelagic ASVs detected in the DOS represent 21.4% of the DNA reads obtained from sediment samples. The proportion of DNA reads of pelagic origin in the sediment follows an increasing trend from low to high latitudes (Fig. 5.5 A). This proportion also broadly approximates POC export from the surface ($R^2 = 0.23$, $P < 0.001$), despite the higher remineralization rates at productive high latitudes (Guidi et al. 2015). Furthermore, the composition of plankton DNA in the DOS can predict up to 58% of the variation in POC export from the surface and 57% of the POC reaching the seafloor using cross-validated random forest regressions (Fig. 5.5 B). We used a multivariate regression method to identify the sinking pelagic ASVs that best explain the variation of POC export and POC reaching the seafloor (Fig. 5.5 C). Not unexpectedly, diatoms and dinoflagellates (Billett et al. 1983; Agusti et al. 2015) were important contributors, but we also identified some previously overlooked taxa that are not usually considered to contribute to POC export, such as alveolate parasites (MALV-II), cercozoans, chrysophytes, and several unknown eukaryotes [see also (Guidi et al. 2016)]. Our time-integrated data from the DOS therefore highlight previously underappreciated taxa that may be keystone drivers of the biological carbon pump.

5.5.4. Toward a holistic view of ocean biodiversity and ecosystem processes

Our global molecular meta-dataset from the ocean surface to the DOS provides the first unified vision of eukaryotic biodiversity patterns across the three dimensions of the world ocean (Fig. 5.1). It shows that the DOS is an extremely rich and unique realm with a strong connection to the water masses above that is reflected in the pelagic DNA signature (Figs. 5.1, 5.4, and 5.5 and fig. S11). Although focused on smaller-sized organisms (eukaryotic microbes and meiofauna), these DNA-based results are broadly consistent with morphological evidence from larger animals for high deep-sea benthic a diversities and small-scale patchiness (high local species turnover) (Grassle and Morse-Porteous 1987; Rex and Etter 2010; McClain et al. 2011). The DOS appears to be much more diverse than oceanic waters (Fig. 5.1 and fig. S1) and is composed of communities of mostly unknown eukaryotes (Fig. 5.2) that display clear biogeographic patterns at global scales and considerable patchiness at local scales (Fig. 5.3). These patterns are likely driven by the flux of sinking organic aggregates and fecal pellets (Fig.

5.3, table S4, and fig. S8) (Smith et al. 2008; Steinberg and Landry 2017). Our data also show that the DNA-based plankton abundance profiles are broadly preserved in the DOS and that the transformation of sinking material appears to occur mostly between the euphotic and aphotic layers (fig. S11). The deposition of eukaryotic plankton is maximal at productive high latitudes (Fig. 5.5), and the plankton contribution to time-integrated sedimentary DNA broadly approximates the yearly average POC export from the surface. Moreover, the taxonomic composition of planktonic assemblages in the DOS is an even better predictor of the variation of POC export from the surface and the fraction of it reaching the seafloor (Fig. 5.5), indicating that biodiversity is key for ocean carbon export and burial. These DOS assemblages comprise not only taxa that are known to be important drivers of the biological carbon pump but also several taxonomic and functional groups that have been overlooked in what is arguably one of the most fundamental ecological processes of the world ocean.

Together, our results highlight the DOS as one of Earth's richest modern ecosystems and fossil archives. They underline the need for concerted international efforts to further understand DOS biodiversity and its ecological role in planetary biogeochemical cycles. Our study, together with recent evidence that plankton DNA signal can be preserved in subseafloor sediments (Kirkpatrick et al. 2016; de Schepper et al. 2019; Barrenechea Angeles et al. 2020; Armbrecht et al. 2021a), paves the way for using sedimentary planktonic DNA to complement the microfossil-based proxies currently used to reconstruct ancient oceans, including their biological carbon sequestration processes. We hope it will also provide the basis for a more informed and effective stewardship strategy for protecting unique and relatively pristine deep-ocean ecosystems as the exploitation of seabed resources gathers pace.

Supplementary Material is accessible online under:

https://www.science.org/doi/10.1126/sciadv.abj9309

Figs. S1 to S13

Tables S1, S4

Captions for Data S2, S3, S5, S6, S7, S8

Other Supplementary Materials for this manuscript include the following:

Data S2, S3, S5, S6, S7, S8

Chapter 6: Assigning the unassigned: a signature-based classification of rDNA metabarcodes reveal new foraminiferal lineages specific to the deep-sea.

Barrenechea Angeles, I., Nguyen Ngoc-Loi, Greco Mattia, Koh Siang Tan, and Pawlowski, J.

6.1. Project Description

In 2020 I had the opportunity to join the Abyss2020 expedition with the Koh Siang group, where I collected sediments for eDNA metabarcoding and foraminiferal specimens. The metabarcoding data was added to previous data from the same region collected in recent years by several members of the Pawlowski's laboratory.

This project addresses a problem already observed in previous studies, the amount of abundant unassigned sequences and the lack of solution to integrate them in ecological analyses. We propose a signature-based assignment to regroup the unidentified sequences and we compare the presence of new lineages and clades in other bathyal and shallow-water sediment collections. I performed the molecular analysis, and the phylogeny analysis were done by Ngoc-Loi Nguyen.

A version of this chapter is ready to be submitted to *Molecular Ecology*

6.2. Abstract

Environmental DNA metabarcoding reveals a vast genetic diversity of marine organisms. Yet, most of the metabarcoding data remain unassigned due to the paucity of reference databases. This is particularly true for the deep-sea meiofauna and microbiota, whose hidden diversity is largely unexplored. Here, we tackle this issue by classifying unknown deep-sea foraminifera metabarcodes based on unique DNA signatures in the hypervariable region of the 18S rRNA gene. In our study, we analyzed metabarcoding data obtained from 311 deep-sea sediment DNA samples collected in the CCFZ zone of the Eastern Pacific Ocean. Using the signatures, we were able to classify 802 unassigned metabarcodes (OTUs) into 61 novel lineages, which have been placed in 27 phylogenetic clades. The comparison of CCFZ metabarcodes with other foraminiferal metabarcoding datasets shows that most novel lineages are widely distributed in the deep-sea. Five lineages are also present in the shallow-water datasets, particularly from the Arctic Ocean, but the phylogenetic analysis of OTUs in these lineages separate the deep-sea and shallow-water metabarcodes in all but one case. While the signature-based classification does not solve the problem of gaps in reference databases, grouping the unassigned metabarcodes provide precious information about their distribution and ecology, which could be very useful in future applications of metabarcoding for environmental biomonitoring.

6.3. Introduction

The past decade has seen metabarcoding become a common tool to assess biodiversity, with the capacity to overcome the limitations of traditional morphology-based methods. Yet, the taxonomic assignment of metabarcoding data remains problematic mainly due to the paucity of reference databases (Gold et al., 2021; Hestetun et al., 2020). The problem is particularly important in the case of microbial surveys, which are dominated by unknown taxa, also called "microbial dark matter" (Rinke et al., 2013). Unassigned sequences also prevail among protist and meiofaunal communities (de Vargas et al., 2015; Forster et al., 2016; Logares et al., 2014). These sequences are commonly lumped into an assemblage of unassigned or unknown metabarcodes, which are generally overlooked in further analyses.

Different strategies have been proposed to overcome this problem. A recent study showed that simple taxonomic assignment approaches based on sequence similarity and composition outperformed more complex phylogenetic and probabilistic methods (Hleap et al., 2021). Nevertheless, the accuracy of taxonomic assignment based on percentage similarity in the case

of usually very short metabarcodes is generally low. Alternatively, a network approach was proposed to characterize unknown species and elucidate their relationships (Zamkovaya et al., 2021).

Here, we tackle this issue by classifying the unassigned metabarcodes into higher-level taxonomic units using discrete DNA signatures. Although the signature approach to detect and identify microorganisms has been proposed already some time ago (Albuquerque et al., 2009; Phillippy et al., 2007), its use in current prokaryotic taxonomy is relatively limited (Hugenholtz et al., 2021). Among eukaryotes, distinctive molecular characters are generally used to resolve the taxonomy of closely related species (Zielske and Haase, 2015) or to analyze geographic patterns (Ganser et al., 2021). A recent study demonstrated the usefulness of DNA signatures to facilitate the taxonomic identification of ciliated protists (Ganser et al., 2022).

In our study, we applied the DNA signature approach to the taxonomic diagnosis of deep-sea benthos. It has been shown that this diversity is huge and far exceeds that of organisms living in surface waters (Cordier et al., 2022). However, due to the limited reference database, most eukaryotic metabarcodes could be classified into large supergroups only. Similarly, in the metabarcoding studies of deep-sea metazoans, the majority of meiofaunal sequences could not be assigned to the lower taxonomic level (Sinniger et al., 2016). This is the main reason why microbial and meiofaunal metabarcoding studies usually report the community composition at a higher taxonomic level, which makes its ecological relevance very limited.

We focused on foraminifera, which comprises a significant fraction of deep-sea benthic diversity (Gooday, 2019; Gooday et al., 2021), and represents more than 50% of the total biomass in some areas (Gooday, 2019). It has been suggested that at least some deep-sea foraminiferal species are distributed globally based on DNA barcodes of isolated specimens (Pawlowski and Holzmann, 2014; Pawlowski and Lecroq, 2010). This has been confirmed by studies reporting several cosmopolitan foraminiferal amplicon sequence variants (ASVs) or operating taxonomic units (OTUs) in deep-sea metabarcoding data (Lecroq et al., 2011; Lejzerowicz et al., 2021). Yet, most of these globally distributed metabarcodes could not be assigned or have only been assigned at the class level. The proportion of unassigned sequences in the deep-sea foraminiferal datasets exceeds 50% (Lecroq et al., 2011; Lejzerowicz et al., 2021).

The material for this study comes from the Eastern Pacific Clarion-Clipperton Fracture Zone (CCFZ), an area of potential polymetallic nodule exploitation. The biological community of CCFZ was targeted by several biodiversity monitoring studies (Glover et al., 2018; Jones et al., 2021; Washburn et al., 2021). The foraminiferal assemblage of CCFZ was shown to be

dominated by monothalamous taxa, most of which remained morphologically and genetically unidentified (Goineau and Gooday, 2017; Gooday et al., 2021, 2017; Lejzerowicz et al., 2021). We performed a metabarcoding analysis on sediments across different areas of CCFZ from four concessions (BGR, IFREMER, UK-1 and OMS) and characterized the foraminiferal metabarcodes, focusing on those that were unassigned. We classified them into 61 new lineages, each defined by specific signatures in the hypervariable region of the 18S rRNA gene. We then compared the lineages from CCFZ with other deep-sea basins and shallow-water regions. The taxonomy of the new lineages and their potential use for environmental monitoring of deep-sea resources are discussed.

6.4. Material and methods

6.4.1. Sediment sample collection

In this study, 36 samples were added to those already included in Lejzerowicz et al. 2021 from the Clarion Clipperton Fracture Zone (CCFZ), which encompassed the license areas of BGR, IFREMER, UK-1 and OMS. The additional samples were collected in 2020 using boxcores during RESOURCE Cruise 01 (OMS license area). At each station, three replicates were taken with a sterile syringe that was inserted into the sediment and pushed into a plastic cup where the last few centimeters were discarded. Only the first 2-3 centimeters were placed into a tube with 10 ml of LifeGuard Preservation solution. Samples were frozen on board, shipped frozen to the University of Geneva, and stored at -20°C until their extraction.

6.4.2. Sediment DNA extraction, amplification, and sequencing

Sediments from the RESOURCE cruise were extracted using the manufacturer's guidelines of the DNeasy® PowerMax® Soil Kit (QIAGEN). To target foraminifera eDNA, the 37F hypervariable region of nuclear 18S rRNA gene (68-196 bp), was PCR amplified using specific primers (Pawlowski and Lecroq, 2010). To allow multiplexing of samples in a single library, the forward s14F1 5′-AAGGGCACCACAAGAACGC-3′ and reverse s15 5'-CCACCTATCACAYAATCATG -3' primers were tagged with unique 8nt added at 5' ends of the primers (Esling et al., 2015). Three PCR replicates were amplified and pooled for each surface sample before being quantified using high-resolution capillary electrophoresis (QIAxcel System, QIAGEN). The amplicons were pooled into one tube according to their equimolar concentration. The short amplicons (<100bp) were then removed from the pool using the High Pure PCR Product Purification kit (Roche). The library was prepared using

TruSeq® DNA PCR-Free Library Preparation Kit (Illumina), and its concentration was quantified using Kapa Library Quantification Kit for Illumina Platforms (Kapa Biosystems). Finally, the library was sequenced with a MiSeq instrument using paired-end sequencing for 300 cycles with a v2 kit.

6.4.3. Bioinformatic analysis

We combined the obtained sequence dataset with the published CCFZ and other deep-sea foraminifera datasets obtained from samples recovered between -4000 and -9000 meters of water depth from North Atlantic, Mid Atlantic, South Atlantic, Southern Ocean and Northwest Pacific (Cordier et al., 2019a; Lejzerowicz et al., 2021) (see table S1; fig. S1), and available from ENA under PRJEB44134, PRJNA554310 and PRJNA899048. We also added the shallow water foraminifera datasets from the Tyrrhenian Sea (Cavaliere et al., 2021), Adriatic Sea (Cordier et al., 2019b; Frontalini et al., 2018; Greco et al., 2022), Persian Gulf (Al-Enezi et al., 2022) and around Svalbard (Nguyen et al., 2022) (see fig. S1), available under PRJNA723313, PRJNA897836, PRJNA813562, PRJEB29469, and PRJNA768352.

The new datasets were demultiplexed using the module DTD from the SLIM software (Dufresne et al. 2019). The fastq files pairs from all datasets were combined and processed using the module DADA2 implemented in SLIM with pseudo-pool parameters. DADA2 (Callahan et al., 2016) filtered good quality reads, trimmed primers, merged overlapping paired-end reads and removed chimaeras. We then clustered the ASV sequences at 97% similarity and continued with a LULU curation (Frøslev et al., 2017) as recommended in Brandt et al. 2021. The clustering was done using the DECIPHER R package and the curation with the LULU R package with the default parameters. We referred to the clustered ASVs as OTUs.

To retain only foraminifera sequences, we removed manually sequences not having “GACAG”, a pattern located in the conservative region 37, adjacent to the foraminiferal-specific hypervariable region 37F (Pawlowski and Lecroq, 2010). Sequences lacking the “TAGTCCTTT” and “TAGTCCCTT” patterns at the end of the 37 conservative region were also deleted. The remaining sequences were then filtered by their size. We retained sequences with > 70bp and > 100 reads.

We used three probabilistic approaches to assign the sequences taxonomically: VSEARCH(Rognes et al., 2016) at 95% similarity, IDTAXA (Murali et al., 2018) at 60% of confidence and BLAST+ (Camacho et al., 2009) at 95% similarity and 100 - 99% of coverage. We used our local database of benthic foraminifera and the PFR2 (Morard et al., 2015) database

to identify planktonic foraminifera. Our local database contained 4602 reference sequences representing foraminiferal classes of Globothalamea, Tubothalamea and Monothalamea. The latter comprised the well-defined clades (e.g., Clade A, Pawlowski et al. 2013), the groups consisting of environmental sequences from previous metabarcoding studies obtained through cloning and Sanger sequencing (e.g. ENFOR1, Pawlowski et al. 2011), and/or poorly defined clades (e.g. Monothalamea X or undet. Monothalamea), mainly comprising so-called squatter sequences (Gooday et al., 2013; Moodley, 1990).

6.4.4. Pattern identification

We prepared a subset of the CCFZ dataset including 2,245 OTUs that could not be assigned by VSEARCH as well as those that VSEARCH assigned to ENFOR (environmental sequences) or Monothalamea X. All sequences with more than 2-3 deletions, insertions, or ambiguities in the conserved regions located before the highly variable region 37F were removed. Sequences having similar molecular signatures at the beginning or at the end of 37F region were regrouped into lineages. They were validated if the number of reads was superior to 5000 reads and the lineages comprised at least 2 OTUs. The retained lineages were compared with the annotations made previously. Lineages were not considered if the same signature was present in the database, except if they were assigned to an environmental sequence or a Monothalamea X. After these restrictive filters, only 693 OTUs were used to define the patterns. We produced an R script, available on GitHub (https://github.com/MatGreco90/ForamSignature), with the Biostrings package, which allowed identifying the patterns without a mismatch in CCFZ, deep-sea and shallow water datasets.

6.4.5. Phylogenetic analysis

The phylogenetic trees specific to new lineages were constructed - covering the entire Monothalamids class to assign taxonomy and resolve undescribed clades. In total, 693 OTUs of new lineages, 388 well-described monothalamid reference sequences, and two type material sequences of non-foraminiferal rhizarians (*Cercomonas longicauda* and *Gromia oviformis*) were used as outgroups in constructing the phylogenetic trees. We aligned our sequences using the E-INS-i iterative refinement method in MAFFT v7 (Katoh et al., 2019). Trees were built using the IQ-TREE maximum likelihood method (Nguyen et al., 2015; Trifinopoulos et al., 2016). Ultra-fast bootstrapping (Hoang et al., 2018) was used to generate branch support values with 1000 bootstrap replicates. Phylogenetic tree visualization and annotation were done using

the R package *ggtree* version 1.12.7 (Yu et al., 2017). Default alignment parameters were used to align and generate a phylogenetic tree.

6.5. Results

6.5.1. Sequence data

After the clustering, LULU curation, removal of non-foraminiferal sequences and a filter of rare ASV (< 100 reads) the CCFZ dataset contained 37,127,019 reads and 2,382 OTUs, the deep-sea dataset 48,559,807 corresponding to 4,148 OTUs and the shallow water dataset comprised 26,349,529 reads and 3,745 OTUs. Details of reads removals at each step and for each basin are detailed in **table S1.**

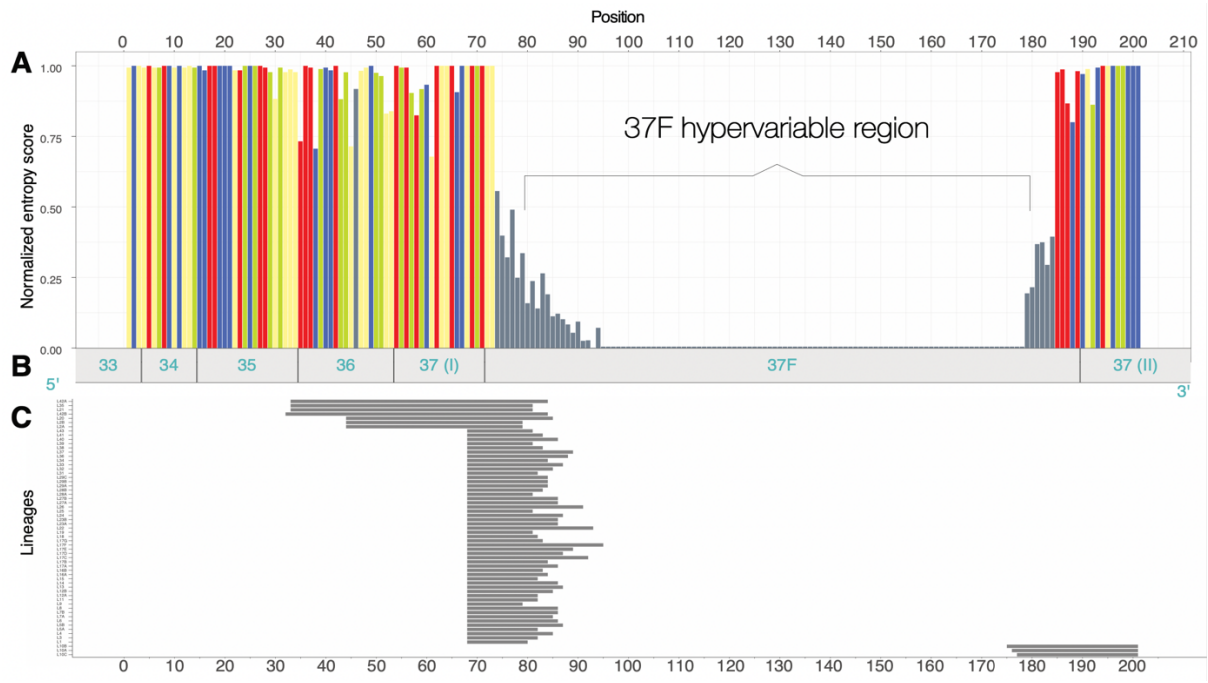

Figure 6.1: Positions of signatures in the foraminiferal 18S rRNA gene. A. entropy plot; B. foraminiferal regions from 33 to 37 after Pawlowski and Lecroq 2010*; C. Position and length of lineages.*

6.5.2. Taxonomic assignment

At first, the OTUs were assigned using the three standard methods, i.e., VSEARCH, BLAST and IDTAXA. All three methods recognized the main classes of foraminifera (Globothalamea, Tubothalamea and Monothalamea) but assigned less than 50% of OTUs. VSEARCH returned the greatest fractions of sequences (46.2%), followed by BLAST (24.1%) and IDTAXA (10.2%). The monothalamids, including environmental sequences (ENFOR) and

Monothalamea X (undetermined Monothalamea), were the most abundant class of foraminifera **(Fig. S2, more details in table S3)**. Globothalamids and tubothalamids were the minority in the three assignments. According to the VSEARCH assignment, monothalamids (including ENFOR and Monothalamea X (undet. Monothalamea)) represented 41.2 % of reads, globothalamids and tubothalamids about 5% of reads and unassigned OTUs accounted for 53%.

In the second part of this study, based on the alignment of 693 selected OTUs, we identified 61 different signatures corresponding to the unique pattern of nucleotide sequences. Each signature defined a lineage, named by the letter L and a number (e.g., L1, L43). A letter was added after the number (e.g., L2A, L2B) to differentiate similar patterns sharing most of the nucleotides. The length of signatures varied between 12 and 53 nucleotides (Table S4). Most of them (51) was located at the beginning of the 37F variable region, comprising the six conservative nucleotides "GACAGG" at the end of the 37 (I) helix (Fig. 1). Seven signatures started in the 35 or 36 regions and finished in the 37F variable region. We also used the end of 37F and 37 (II) regions to discriminate three lineages (Fig. 1).

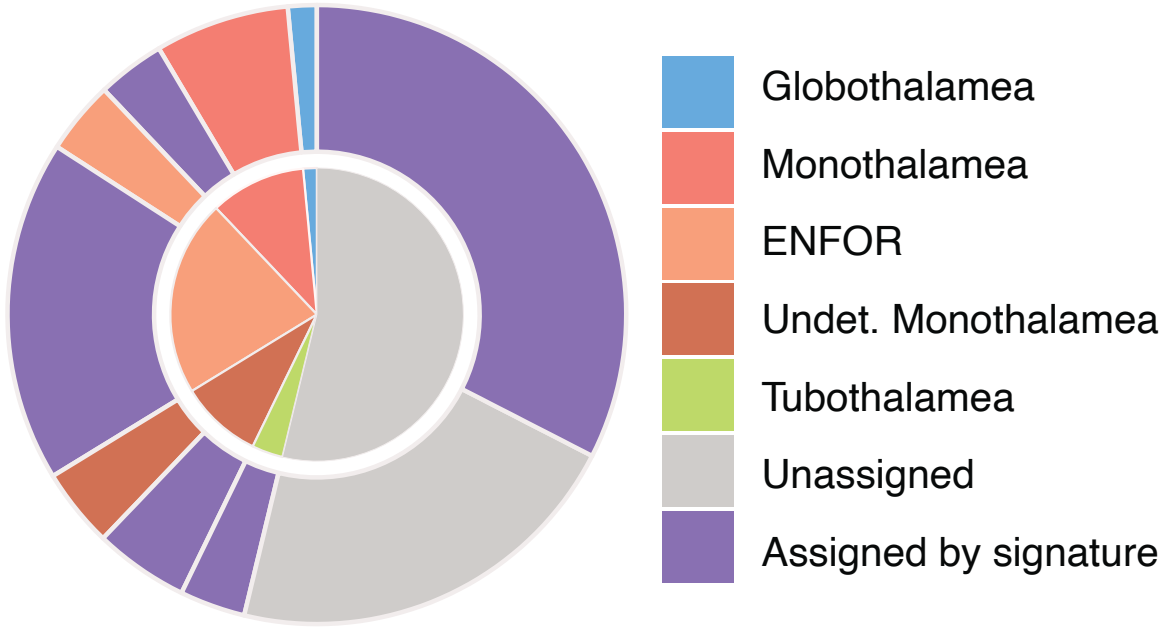

Figure 6.2: The nested pie chart shows the proportion of new lineages (in purple) before and after including the new lineages in unassigned, monothalamids, ENFOR and Undet. monothalamids. The inner pie chart represents the proportion using VSEARCH assignments and the outer ring a combined assignment using the new lineages.

With the R script, we could identify 109 additional OTUs containing the signatures without a mismatch, i.e., with 100% of similarity (see Table S5). In total, 802 OTUs (corresponding to

34% of total number of OTUs and 62% of total number of reads) were assigned to novel lineages. Compared to VSEARCH, our approach allowed us to reduce the number of unassigned OTUs from more than 50% to 21% (Fig. 2). The signatures were also found in many sequences already identified with VSEARCH at 95% similarity. The largest proportion of OTUs included in new lineages (82%) was found among the environmental ENFOR clades. We also found large proportion of OTUs assigned to novel lineages among the monothalamids (34%) and the undetermined monothalamids (Monothalamea X, 54%). One of the novel lineages was assigned to Tubothalamea, but this requires confirmation by single-cell sequencing. No signature was found among the globothalamids.

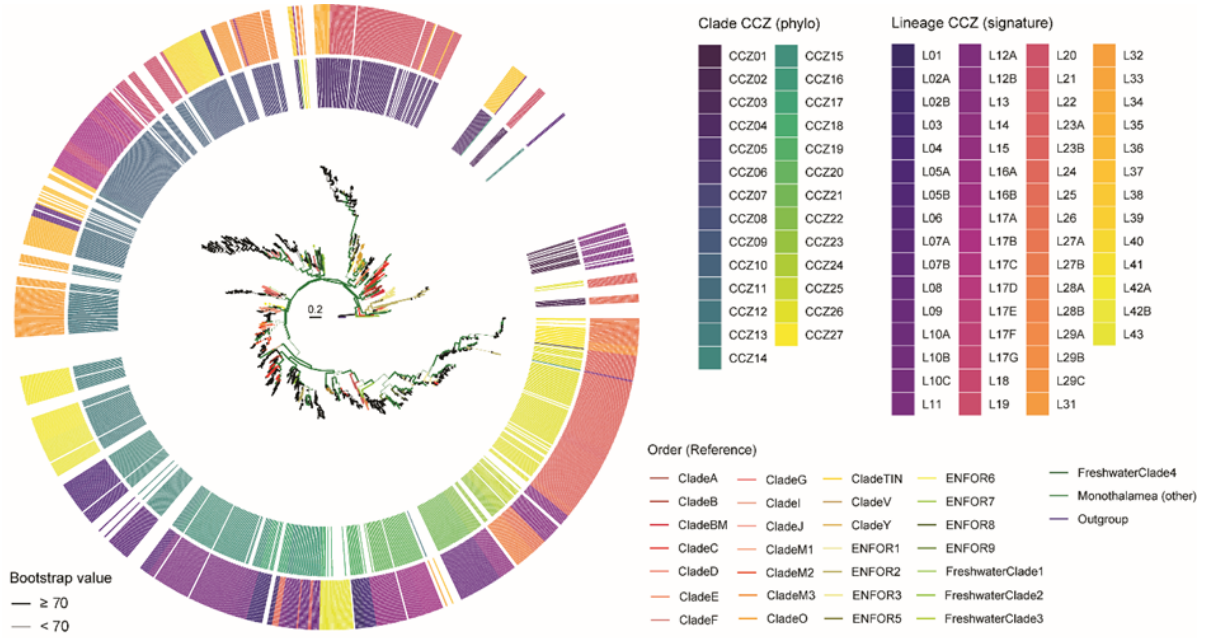

Figure 6.3: Phylogenetic diversity and novelty of foraminiferal OTUs based on signal identification. Phylogenetic analysis of selected OTUs revealed signatures (as new lineages) and representation of reference sequences of monothalamids from Clade A to Clade Y and some freshwater clades. Tree branches are coloured at the Order level. All sequences were aligned with MAFFT, and trees were constructed with IQ-TREE, based on the GTR+F0 model of evolution with 1000 bootstrap replicates. Bold branches indicate strong support (≥70% bootstrap support). Scale bars are in units of substitutions per site. The rings indicate clusters based on phylogenetic position (inner ring) and signatures (outer ring).

6.5.3. Phylogenetic placement of new lineages: definition of new clades

To evaluate the taxonomic assignment of the signature-based approach, we constructed a phylogenetic tree from the 693 OTUs containing the signature with reference monothalamid sequences. A simplified version of the tree is presented in Fig. 3 with a more detailed version

provided in Fig. S3. We regrouped the 43 lineages clustered together in 27 higher-rank groups (e.g., CCZ1) to provide the appropriate degree of phylogenetic specificity for each signature (Table S4). Most of the new lineages formed monophyletic groups. They belonged to the previously established clades of monothalamids (e.g., Clade C, Clade M, Clade I, Clade V) and environmental DNA-derived foraminiferal sequences (ENFOR clades).

The phylogenetic analysis indicated that signatures assigned lineages were more similar to each other than to those of distant ones. Most of the new lineages were placed on the tree at the specific clades, which indicated a general agreement between their signature assignment and phylogenetic positions. Interestingly, some new lineages were found in specific groups that are highly related to other CCFZ sequences from the database (i.e., L14, L19, L21, L23B, L28A, and L42A). The OTUs of one lineage (L17) form a group on their own, with no closest reference-related sequences.

6.5.4. Biogeography of new lineages

The comparison of metabarcoding datasets within CCFZ and with other deep-sea and shallow-water sites showed clear patterns of distribution of the newly defined lineages **(Fig. 4).** Within the CCFZ, the OMS and UK1 regions shared all the lineages whereas in BGR the lineage L29C was not found. The IFREMER area, located in the westernmost part of CFFZ, has the lowest number of lineages (49) shared with the eastern part of CCFZ. Compared to other deep-sea sites, most of the lineages (85%) were present in at least 4 out of 5 regions. Only five lineages were endemic to CCFZ (absent in all other areas). Besides those lineages, all others were present in the Northwest Pacific. In the Southern Ocean, we found 53 lineages, while 50 were found in all three regions of Atlantic Ocean. One lineage (L4) was present in the North and mid-Atlantic whereas two lineages (L8 and L17A) were found in the mid and south Atlantic. Lineages 28A and 29C only occurred in North Atlantic and the mid-Atlantic, respectively.

Compared to the deep-sea, 30 out of 61 lineages were also present in shallow water sites. 26 lineages were present in the Arctic fjords (Svalbard), while 10 were found in the two Mediterranean Sea sites. Only five lineages were present globally, including the Persian Gulf. Two of them (L21, L43) were most abundant and had in common with the other three cosmopolitan lineages a very short signature.

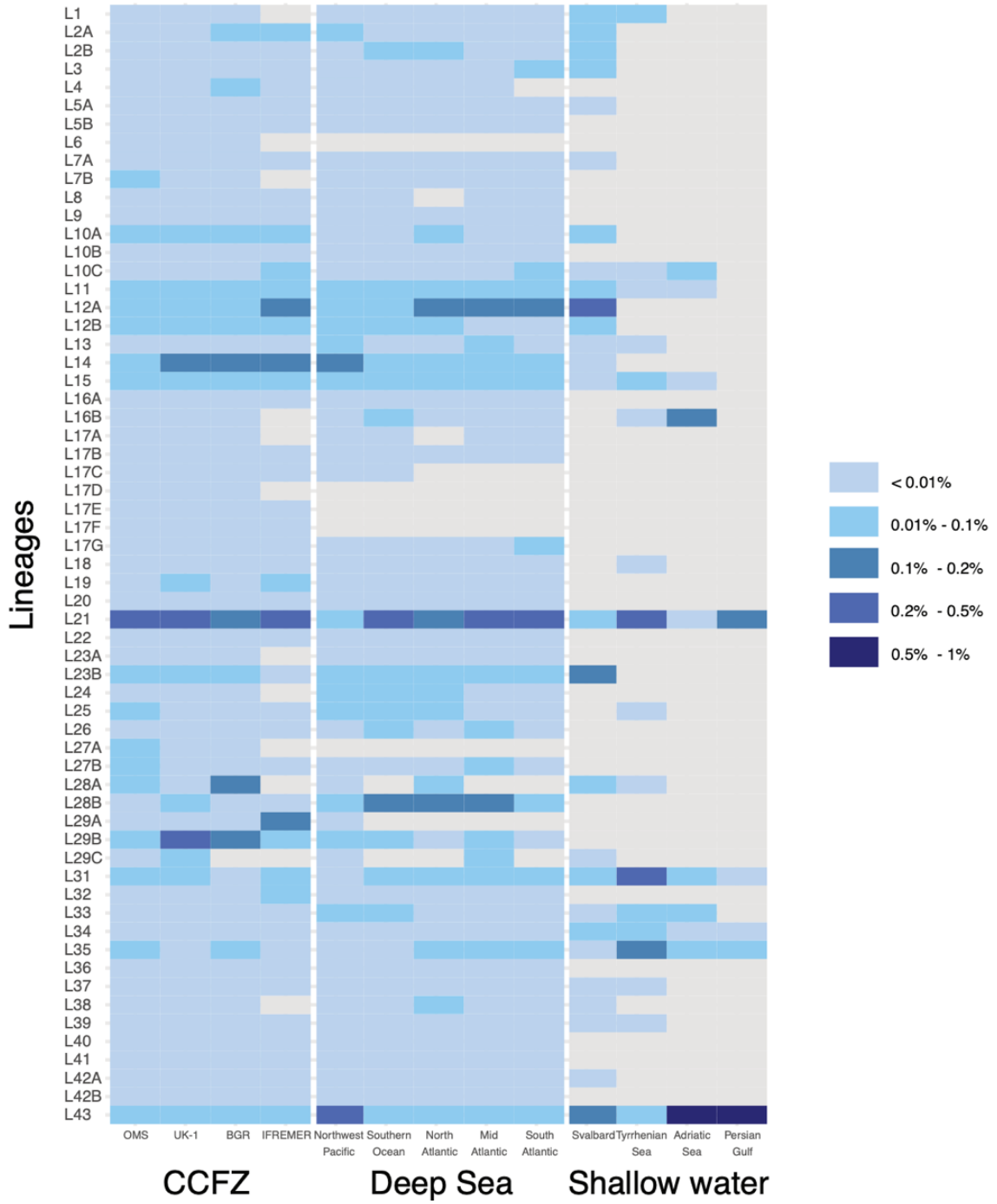

Figure 6.4: Heatmap showing the proportions of 43 lineages and signatures present in all regions from CCFZ, Deep-Sea (non-CCFZ), and Shallow Water (<200 m depth; see section 2.3). The calculations were made by taking only the reads assigned to one of the lineages. Unassigned and previously assigned reads were excluded. The blue gradient shows the relative abundance of lineages per area. The darker the colour, the more abundant the lineage is in each area. Grey areas indicate the absence of lineages.

To better understand the biogeography of the five cosmopolitan lineages (L21, L31, L34, L35 and L43), we analyzed the distribution of OTUs composing these lineages. The highest diversity in terms of the number of OTUs retrieved was observed in L21, which counted a total

of 162 OTUs. Most of them were characteristic of deep-sea sites (71), with 41 OTUs exclusive to CCFZ sites, while 29 were shared between them (Fig. 5). Within this lineage only a single OTU occurring in the shallow-water datasets was also observed in the deep-sea.

The lineages L31 and L34 presented an overall lower diversity (38 and 26 OTUs respectively), with the majority of the OTUs retrieved uniquely from shallow water samples. Along with L43, L31 and L34 were the only three lineages presenting OTUs with a distribution encompassing all the ecosystems analyzed. In particular, the overall diversity of L43 constituted 63 % of OTUs occurring in all the datasets. In contrast, L35 mainly presented OTUs with habitat-specific distributions with only 5 OTUs shared between CCFZ and deep-sea sites.

6.6. Discussion

Despite the advances in biological diversity assessment introduced by metabarcoding, taxonomically unassigned sequences remain a nightmare for researchers interested in organismal ecology. As shown by our study, about half of the deep-sea foraminifera metabarcodes could not be assigned. This proportion is even higher if we also consider as unassigned, the metabarcodes that were classified only at the highest taxonomic level (i.e., Foraminifera, Monothalamea, ENFOR). Indeed, the assignment at such a high level provides no information about the biology of organisms represented by given sequences, ASVs or OTUs, hampering any attempt of their ecological interpretation.

By using diagnostic 18S rDNA signatures, we were able to increase the number of assigned reads from 54% when using conventional bioinformatics tools (VSEARCH, IDTAXA and BLAST) to 80% using signature-based approach (Fig. 2). In total, 61 new foraminiferal lineages have been defined based on DNA signatures. Unsurprisingly, most of these lineages belong to the class Monothalamea, a paraphyletic assemblage of early evolved single-chambered foraminifera (Pawlowski et al., 2013). They are characterized by soft-walled or naked cells, which are generally overlooked in conventional foraminiferal surveys (Pawlowski et al., 2003). Our study has expanded our knowledge of this group and contributed to its classification.

Besides this taxonomic aspect, our approach has also contributed to our understanding of the ecology and geographic distribution of monothalamous foraminifera. This information is generally lost by lumping together all unassigned foraminiferal sequences. Some authors analyzed metabarcoding data at the level of ASV or OTU, for example, in the study of patchiness of deep-sea foraminifera (Lejzerowicz et al., 2014) or their distribution along the

gradient of depths (Cordier et al., 2019a). Yet, the ASV or OTUs represent a very low taxonomic level, corresponding to species or intraspecific variants. Inferring general patterns of distributions and ecological adaptations based on foraminiferal ASVs or OTUs might be difficult, especially given the presence of intragenomic polymorphism in this group (Weber and Pawlowski 2014). Our approach provides new information at an intermediary level, which groups the ASV/OTUs into higher taxonomic units that might be easier to correlate with environmental variables.

The outcomes of this approach are well illustrated by the results of our investigation on the distribution of deep-sea foraminifera. Previous studies suggested that some deep-sea species are globally distributed (Lecroq et al., 2011; Pawlowski et al., 2007). However, the species targeted by these studies (e.g., *Epistominella exigua*) represented genera that are widely distributed in the coastal environment, and the deep-sea species were considered as possessing special adaptations to this particular environment. Our study suggests that the numerous foraminiferal lineages are specifically deep-sea. For a specialist of deep-sea biodiversity, this might not come as a surprise. The Xenophyophorea are a highly diverse group of giant monothalamous foraminifera occurring exclusively on abyssal plains (Gooday et al., 2017). Nevertheless, according to our study, the taxonomic range of foraminiferal lineages adapted to the deep-sea might be much wider than expected.

Admittedly, the signature-based approach does not allow us to exactly determine the taxonomic status of the new lineages. We expect that at least some of them correspond to the genus or species level. This could be the case of lineages specific to CCFZ (L17D, E, F), characterized by a long signature. Our approach is based on the observation that the variability increases progressively at the end of 37 helix and at the beginning of 37F variable region (Esling et al., 2015; Pawlowski et al., 2014). Thus, the longer signatures might better define the lower taxonomic level. However, any inference of taxonomic status from a single variable region needs to be treated with caution, given the high variability of evolutionary rates in foraminiferal ribosomal genes (Pawlowski et al., 1997). This can be solved by increasing the number of sequences of each species to assess the divergence within a taxonomic group. Once a comprehensive molecular database for foraminifera is established, one would have to develop a further signature-based approach to make it useful for taxonomical and ecological studies.

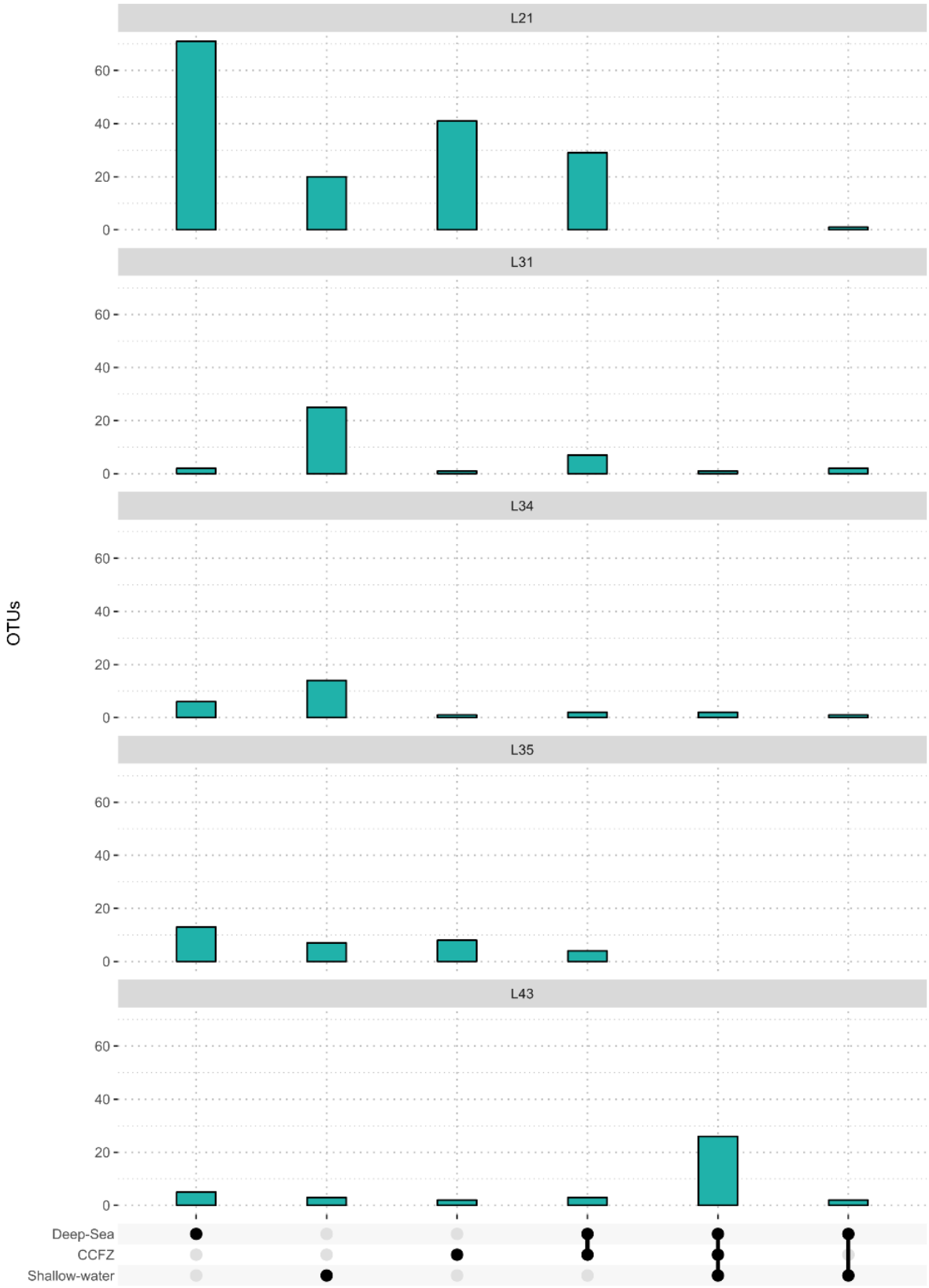

Figure 6.5: UpSet chart showing the shared OTUs between CCFZ, Deep-Sea and Shallow water lineages. All duplicate OTUs were removed and the number of OTUs is a conservative estimate per habitats.

A practical advantage of our approach is its technical simplicity and unambiguity. As the signature patterns are defined at 100% similarity, there is no place for any ambiguity regarding lineage identification. This aspect seems particularly important in the case of short (<100 nucleotides) metabarcodes, where one SNP equals 1% divergence. The shortcoming of such an approach is that the slightest variation in the signature, even one base change, prevents us from including a given OTU in the lineage. However, if we do not apply this rule, the signatures rapidly lose their specificity. Here, we preferred to create two or more lineages (e.g., A and B) that differ by an SNP, rather than accept one SNP change. Nevertheless, well-defined ambiguities could be accepted in the future, especially if their presence is confirmed by intra-genomic polymorphism analysis.

To conclude, we view our approach as an inclusive tool that allows expanding the information inferred from metabarcoding data to the currently "useless" unassigned metabarcodes. We do not view the signature-based classification as a panacea to fill the gaps in the reference database for particular habitats or taxa. There is no doubt that building a comprehensive reference database is essential for biodiversity surveys. Yet, in certain circumstances, this task might be unrealistic. We are convinced that our approach can be very useful in metabarcoding studies dealing with overlooked taxonomic groups, such as monothalamous foraminifera, and/or poorly explored habitats, such as in the deep-sea. We view it as a practical way to uncover hidden ecological information present in hitherto unassigned metabarcoding data. Our approach can be mainly useful in the case of environmental monitoring that targets particular groups of bioindicators or paleo-metabarcoding analysis of environmental changes. Its efficiency will certainly increase if the metabarcoding data are combined with single-cell high-throughput barcoding.

Supplementary figures:

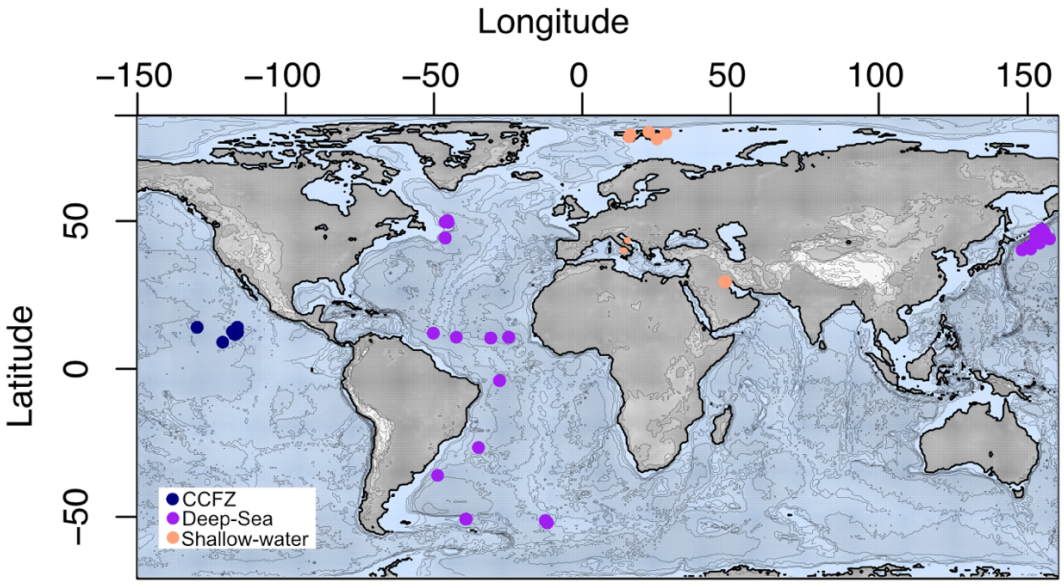

Figure S6.1 Location of basins from which the samples were collected. All the Clarion Clipperton Fracture Zone (CCFZ) sites are shown in blue circles, other the deep-sea sites in purple circles and the shallow water sites in orange circles.

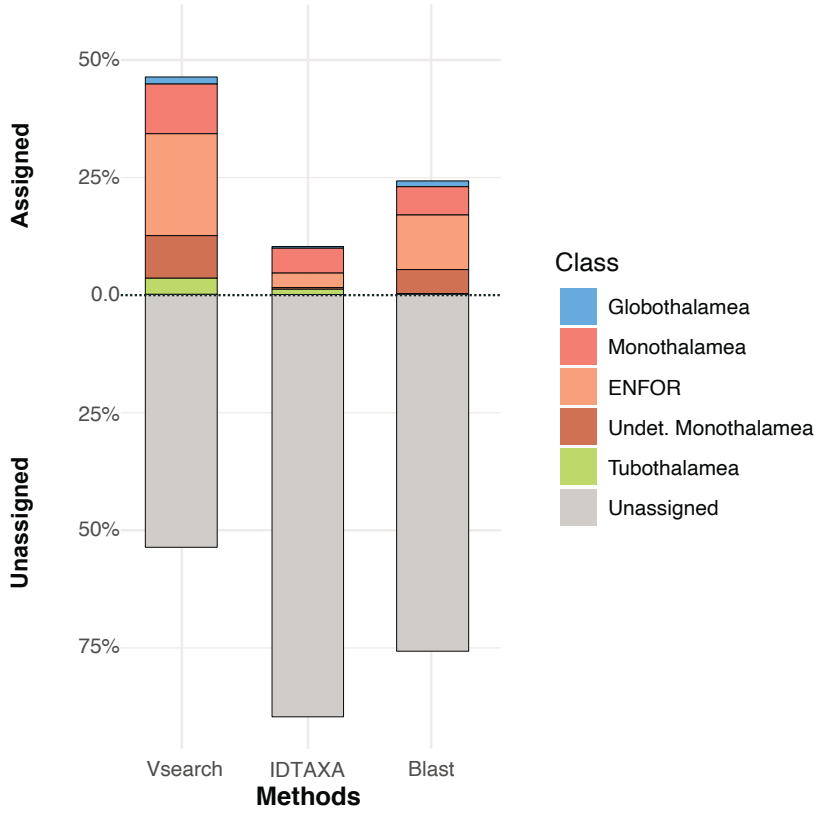

Figure S6.2 Taxonomic composition at class level and relative abundance of assigned and unassigned sequences using the three common methods: VSEARCH, IDTAXA and Blast. All monothalamids

sequences, including the environmental sequences (ENFOR) and sequences not regrouped in a clade that are grouped into undetermined monothalamea (Undet. Monothalamea) are coloured in shades of orange. More details in table S3.

Chapter 7: Assessing the Ecological Quality Status in the highly polluted Bagnoli area (Tyrrhenian Sea, Italy) using foraminiferal eDNA metabarcoding

Cavaliere, M., Barrenechea Angeles, I., Montresor M., Bucci, C., Brocani, L., Balassi, E., Margiotta, F., Francescangeli, F., Bouchet, V.M.P., Pawlowski, J. and Frontalini, F.

7.1. Project Description

At the end of 2019 with Jan Pawlowski, we joined the group of Frontalini and Montresor to collect samples in the bay of Bagnoli, in front of the old industrial area. The samples were collected following transect from inner side of the bay towards outer side. For this project, Frontalini and Montresor's group did the morphology and geochemistry part. I did the laboratory work and the analysis of the molecular data as well as some statistical analysis. These data allowed us to assess the ecological quality of Bagnoli Bay and to determine that the sites close to the former complex are still highly polluted.

A version of this chapter is published in *Science of The Total Environment* 790, 147871.

7.2. Abstract

Morphology-based benthic foraminifera indices are increasingly used worldwide for biomonitoring the ecological quality of marine sediment. Recent development of foraminiferal eDNA metabarcoding offers a reliable, time- and cost-effective alternative to the morphology-based foraminiferal biomonitoring. However, the practical applications of these new tools are still very limited. In the present study, we evaluate the response of benthic foraminifera and define the Ecological Quality Status (EcoQS) in the Bagnoli area (Tyrrhenian Sea, Italy) based on traditional morphology-based approach and eDNA metabarcoding. Geochemical data show that several sites in front of the former industrial plant contain higher concentrations of potentially toxic elements than the Effect Range Median and are characterized by the highest total organic carbon (TOC) content, whereas the distantly located sites can be considered relatively low- to unpolluted. Significant differences (i.e., diversity and assemblages' composition) in both morphological and molecular datasets are found between the relatively low- to unpolluted and the most polluted areas. Similarly, the selected ecological indices of both morphological and molecular datasets strikingly and congruently result in a clear separation following the environmental stress gradient. The molecular indices (i.e., g-exp (H'_{bc}), g-Foram AMBI and g-Foram AMBI-MOTUs) reliably identify poor to bad EcoQS in the polluted area in front of the former industrial plant. On the other hand, the Foram-AMBI based on morphology well identifies an overall trend but seems to overestimate the EcoQS if the traditional class boundaries are considered. The congruent and complementary trends between morphological and metabarcoding data observed in the case of Bagnoli site further support the application of foraminiferal metabarcoding in routine biomonitoring to assess the environmental impacts of heavily polluted marine areas.

7.3. Introduction

In recent years, the implementation of several marine legislations has emphasized the need to characterize the worldwide increasing degree of marine pollution of coastal environments (Lu et al., 2018). Coastal marine sediments act as the final sink of contamination (Salomons et al., 1987) posing a high anthropogenic stress of marine habitats and species living therein. In this context, several areas marked by strong anthropogenic impact were recognized in Italy and defined as Sites of National Interest (SINs), some of which are marine sites such as Bagnoli (NW sector of Naples, Italy) (Ausili et al. 2020).

Bagnoli was the former second largest steel factory of Italy (Trifuoggi et al. 2017); an area where intensive industrial activities widely operated until 1992. The former-industrial plant hosted steel, asbestos and concrete industries on a wide territory facing the sea. In 2000, the area has been declared a SIN and the first remediation programmes were planned. Today, Bagnoli is now a large brownfield area whose marine sediments are still intensively impacted as shown by several ecological and geochemical surveys (Romano et al. 2004, 2018; Sprovieri et al. 2020). Although the area has been largely investigated, the assessment of its ecological quality is crucial to highlight potential changes in the anthropogenic stress on the aquatic biota of the area. The evaluation of the impact is commonly carried out through geochemical analyses aimed at recognizing pollutants whose concentrations could trigger serious imbalances within the aquatic ecosystem. However, chemical analyses are not directly related to the ecological status of an ecosystem. The ecological quality is instead more linked to the biota living therein. Indeed, the detection of unforeseen impacts can be achieved through biological monitoring, (i.e., bioindicators). In this context, European Union has implemented the Water Framework Directive (WFD, 2000/60/EC) to protect, maintain and restore the environmental quality of water bodies and sustainable use of water and the Marine Strategy Framework Directive (MSFD, 2008/56/EC) to identify strategies to achieve a Good Environmental Status also through the consideration of biological elements. Among the numerous biological groups used to assess the health of marine ecosystems, benthic foraminifera, single-cell organisms with a test (i.e., shell), are increasingly applied as proxies of the environmental conditions (Alve 1995; Francescangeli et al. 2020) and, more recently, in environmental biomonitoring to evaluate the Ecological Quality Status (EcoQS) (e.g., (Alve et al. 2009, 2019; Bouchet et al. 2012, 2018, 2021; Francescangeli et al. 2016; el Kateb et al. 2020).

Morphology-based taxonomy is still the preferred approach used for evaluating changes in benthic foraminiferal species and assemblages' composition in response to stress conditions. Recently, environmental DNA (eDNA) metabarcoding has been successfully applied to characterize benthic foraminiferal community (Pawlowski et al. 2014a). This innovative and effective methodology has been applied for investigating the response of benthic foraminifera to different human-related activities and stress such as fish farms (Pawlowski et al. 2014a; He et al. 2019), oil and gas platforms (Laroche et al. 2016, 2018; Cordier et al. 2019c; Frontalini et al. 2020b). Compared to the classic morphology-based approach, eDNA metabarcoding allows to process rapidly and accurately larger number of samples and data (Pawlowski et al.

2014a). However, only a limited number of studies has simultaneously evaluated and compared the performance and congruence of the traditional morphology-based biomonitoring with the eDNA metabarcoding (Frontalini et al. 2020b). Furthermore, foraminiferal biotic indices are still to be implemented on eDNA-based datasets.

The aims of the present paper are therefore: 1) to provide an environmental characterization of Bagnoli area highlighting the spatial variations of the anthropogenic impact on the marine ecosystem; 2) to test foraminiferal biotic indices on eDNA datasets and 3) to assess and compare the EcoQS based on both molecular- and morphology-based benthic foraminiferal communities.

7.4. The study area

The Bagnoli area is located in the NW metropolitan territory of Naples (Campania Region, Southern Italy), along the coastline of the Pozzuoli Gulf (Tyrrhenian Sea) (Fig. 7.1a, b). The area of Bagnoli is part of the Campi Flegrei caldera, a large territory influenced by active volcanism and hydrothermal activity that modify the chemical composition of local groundwater (Celico et al., 2000).

The first industrial activities started at the beginning of the 20th century, and in the first decades of 1900, large plants of steel (ILVA-Italsider), asbestos (Eternit) and concrete (Cementir) were operating, reaching their development peak in the late 1960's. For stimulating productivity, the natural landscape of the territory has also been deeply modified. In 1920, two piers were built to permit the access to the factory of large ships carrying fossil coal, iron ores and limestone, while in 1935 the Island of Nisida was connected to the mainland by the construction of a long strip of land, changing the water circulation system in the gulf. High concentrations of heavy metals (i.e., Cu, Fe, Hg, Mn, Pb and Zn) and hydrocarbons, particularly polycyclic aromatic hydrocarbons (PAHs) occur in marine sediments in front of the former plants area (Damiani et al. 1987; Romano et al. 2009; Arienzo et al. 2017). Contamination is particularly high between the two long piers, where, in 1960-62, large amount of polluted materials was disposed to expand the plant operating area. The industrial production ceased in the 90s and the industrial facilities were completely dismantled in early 2000s (Romano et al., 2004; Ausili et al., 2020). In 2000, Bagnoli-Coroglio site was listed among the Italian SINs. These SINs are areas where a threat to human health occurs, and a recovery programme must be implemented. In the last years, the brownfield site has been the object of a governmental remediation project aimed at environmental restoration. Different remediation strategies have been proposed. In particular,

marine sediments from Bagnoli should be treated as special waste and physically removed from the site, although this technique is invasive and may seriously impact the environment.

Several studies have accurately addressed the response of benthic foraminifera to pollution in the Bagnoli area (Bergamin et al. 2003, 2005; Romano et al. 2008, 2009). On overall, these studies have evidenced from an absence to a very low abundance of living foraminifera near the plant, particularly between the two piers; given that the total assemblages were always considered (Bergamin et al. 2003; Romano et al. 2008). Additionally, several species have been either identified as tolerant (i.e., *H. germanica, Miliolinella subrotunda*, and *Quinqueloculina parvula*) or sensitive (*L. lobatula, A. mamilla and R. bradyi, Elphidium macellum* and *Miliolinella dilatata*) to pollution (Bergamin et al. 2003, 2005; Romano et al. 2008, 2009).

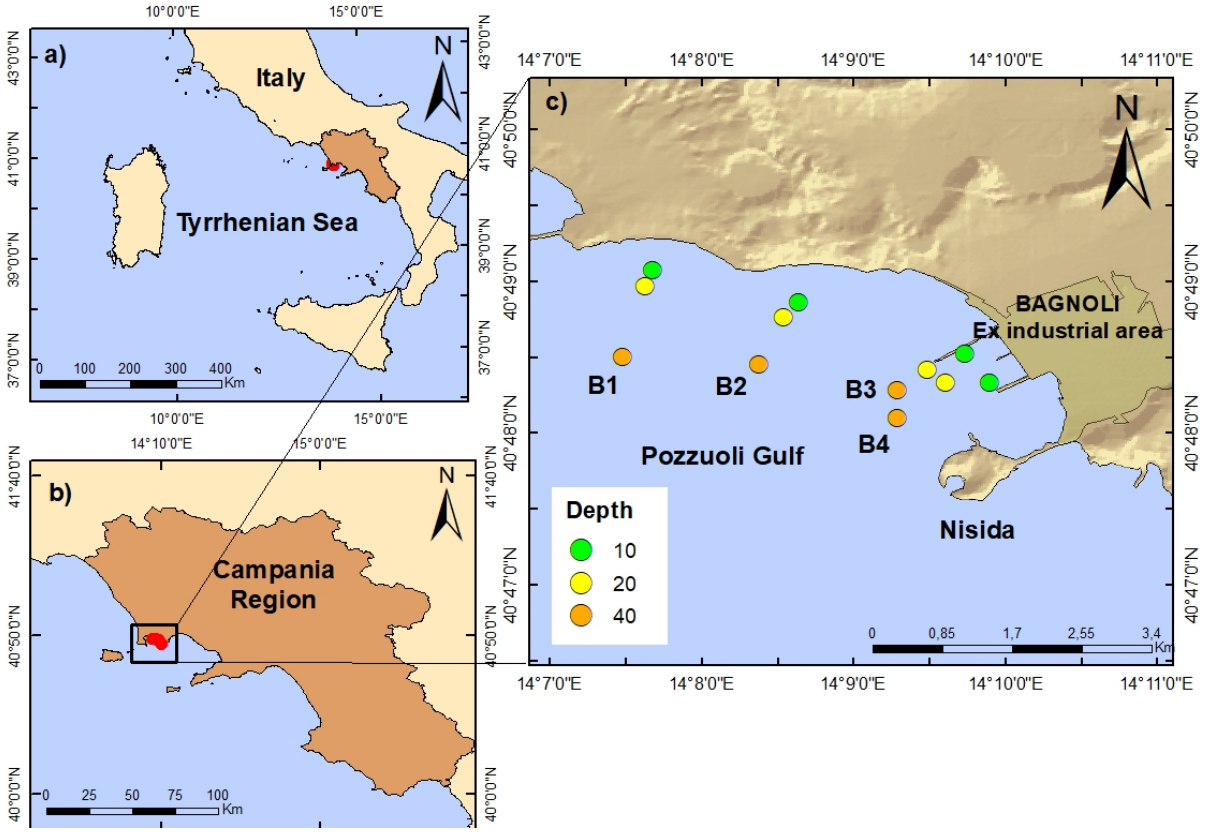

Figure 7.1: Study area: a) location of Campania Region in Southern Italy; b) location of Bagnoli area in the Campania region; c) sampling locations and relative water depth within the Pozzuoli Gulf. Geographical coordinates are referred to World Geodetic System 1984 (WGS-84).

Materials and Methods

7.4.1. Sampling and samples treatment

The superficial marine sediment samples were collected at 12 sites with a box corer along 4 transects (B1, B2, B3, and B4) at different water depths (10 m, 20 m, and 40 m) in the Bagnoli area (Fig. 7.1c). The transects B3 and B4 are placed in the axis of the two piers within the internal area so they would reflect the most impacted areas, whereas B1 and B2 are about 2.5 km away from the ex-industrial plant and would therefore represent a lower level of contamination.

Sediments were collected in three replicates for both morphological and molecular benthic foraminiferal analyses, by independent deployments of the box-corer. The sediments from the only first deployment were sub-sampled for geochemical and grain-size analyses. Only the uppermost part of the sediment (1 cm) was sampled and used for morphological, molecular, grain-size and geochemical analyses. Once collected, samples (50 cm^3) for foraminiferal morphological analyses were immediately stained with a rose Bengal solution (2 g/L) for the identification of living stained foraminifera. Samples (~10 g) for eDNA metabarcoding were immersed in 10 mL of Life Guard™ Soil Preservation Solution (Qiagen), stored on ice during transportation to the laboratory, and then frozen (−20°C) until processing. All manipulations have been carried out with sterile gloves and spoons. Samples were then delivered to the University of Geneva (Switzerland). Sediments for geochemical analyses (i.e., organic matter and trace element) were stored on ice, while another aliquot for grain-size was preserved at room temperature.

7.4.2. Sediment analysis

7.4.2.1. Grain-size

For grain-size analysis, samples (~50 g) were dried and weighted. They were then wet sieved on a 63-μm mesh sieve and a 2-mm mesh sieve, re-dried and weighted. The relative proportions of gravel, sand and mud were then calculated.

7.4.2.2. Organic matter characterization

Sediment samples were thawed out at room temperature and dried in an oven at 60°C. Later they were grinded and homogenized. Sub-samples were then weighed (~10–20 mg) in silver capsules and treated with 1 M HCl (three/four times) in order to completely remove carbonates.

Total Organic Carbon (TOC) and Total Nitrogen (TN) were analysed using a Thermo Electron Flash Elemental Analyzer (EA 1112) calibrated with acetanilide (C_8H_9ON, elemental composition: 71.09% carbon and 10.36% nitrogen) standards. The C/N ratios were calculated to investigate the origin of sedimentary organic matter (Meyers 1994).

7.4.2.3. Inorganic geochemistry

One aliquot of sediment was used for geochemical analysis by the certified laboratory Bureau Veritas (Canada). Sediments were digested in aqua regia (1:1:1 HNO_3:HCl:H_2O) in order to better understand the bioavailability. Chemical analyses were performed on the fine (<63 µm) fraction as the potential toxic elements (PTEs) (i.e., trace elements) are commonly adsorbed in the organic matter, which predominantly occurs in the finest sediments fractions (Baran et al., 2019). The concentrations of PTEs were analysed by Inductively Coupled Plasma-Mass Spectrometry (ICP-MS). Quality control involving the use of replicate and reference materials was also performed. To potentially identify anthropic contributions, selected pollution indicators were computed: Enrichment Factor (*EF*), Contamination Factor (C*f*), Modified Degree of Contamination index (mC_d) and Pollution Load Index (*PLI*).

The *EF* defines the contamination level by minimizing the grain-size effect through the normalization of one metal concentration respect to the concentration of a reference element. The *EF* was calculated for each of the nine selected PTEs (i.e., As, Cd, Co, Cr, Cu, Hg, Ni, Pb, and Zn) using Al as reference element. Aluminium is adopted by many authors in marine assessments (Francescangeli et al. 2020) since it is not normally influenced by anthropogenic processes. The *EF* was calculated as follow (1):

$$EF = \frac{[X_i]/[Al_i]}{[X_0]/[Al_0]} \quad (1)$$

where X_i and Al_i are the concentrations of the element X and Al, respectively in the sample i; X_0 and Al_0 are the local geochemical background concentrations of the element X and Al, respectively. The background values are based on different studies carried out nearby the Bagnoli area (Damiani et al. 1987; Cicchella et al. 2005; de Vivo and Lima 2008) (Table S1). The *EF* levels were assigned using the classes proposed by Müller (1979) (Table S2).

The C*f* represents the degree of sediment contamination by a single element and was calculated for the nine selected PTEs following the equation:

$$Cf = \frac{[element]i}{[element]background} \quad (2)$$

The mC_d represents the sum of all *Cf*s of the considered elements. It was obtained as follow (3):

$$mC_d = \frac{\sum_{i=1}^{i=n} C_f}{n} \quad (3)$$

where n is the number of analysed elements and *i*=ith element. The categories of mC_d, from Hakanson (1980), are reported in Table S2.

The *PLI* represents a synthetic parameter to evaluate the degree of contamination (Tomlinson et al., 1980). It was calculated following the equation (4):

$$PLI = \sqrt[n]{Cf_1\ Cf_2 \dots Cfn} \quad (4)$$

where n is the number of potentially contaminant elements, and C*f*i is the Cf of the element.

To define the environmental quality of marine sediment, chemical concentrations of PTEs have been compared with different Sediment Quality Guidelines (Burton, 2002 for a review) (Table S1). The Effect Range median (ERM) that expresses values above which effects are frequently observed (Long and Morgan, 1990; Long et al., 1995), and the Effect Range Low (ERL) indicating low concentrations of pollutants (O'Connor, 2004) (that is roughly related to toxicity) have been considered.

7.4.3. Benthic foraminifera

7.4.3.1. Morphological analysis

Benthic foraminiferal analyses were performed following the FOBIMO protocol (Schönfeld et al. 2012). Samples were gently washed through a 63 µm sieve in order to eliminate mud particles and excess stain. Residues were then oven dried at 40°C overnight. As the number of living specimens was low in the >125 µm fraction, the picking was also extended to the 63-125 µm fraction. Replicates were independently processed, and the resulting picked specimens were pooled before statistical analyses. The taxonomical classification was carried out through microscopic observation following comparisons with Cimerman and Langer (1991), Sgarella and Zei (1993)and (Milker and Schmiedl 2012). The World Register of Marine Species (Hayward et al. 2022) was followed for the systematic classification of the benthic foraminiferal species.

7.4.3.2. eDNA metabarcoding

The extraction of sedimentary eDNA was performed with DNeasy® PowerMax® Soil Kit (QIAGEN). The eDNA extracts were concentrated with 10.4 ml of ethanol and 200 µl of NaCl as in the manufacturer's guidelines and resuspended in 400 µl elution buffer. The eDNA extracts were then stored at -20°C. The 37f-41f region of the 18S rRNA gene was amplified

using the primer 14F1: 5'– AAGGGCACCACAAGAACGC –3' and S17: 5' – CGGTCACGTTCGTTGC – 3'. To enable the multiplexing of Polymerase Chain Reactions (PCR) products in sequencing libraries, we used tagged primers bearing 8 nucleotides attached at each primer's 5′-extremity (Esling et al. 2015a). PCR program is detailed in S1. Per each extracted sample, three PCR replicates and one PCR-negative control were amplified. Then, the PCR products were verified through agarose gel electrophoresis. The total volume of three PCR replicates were combined to be quantified using a high-resolution capillary electrophoresis (QIAxcel System, QIAGEN). The combined PCR products were pooled in equimolar concentrations and then purified using the Purification High Pure PCR Product Purification kit (Roche).

The library was prepared using TruSeq® DNA PCR-Free Library Preparation Kit (Illumina). It was then quantified using Kapa Library Quantification Kit for Illumina Platforms (Kapa Biosystems). Finally, the library was sequenced on a MiSeq instrument using paired-end sequencing for 2*250 cycles with kit v2.

The raw data was analysed using the SLIM v0.4 web application (Dufresne et al. 2019). The sequences were demultiplexed with no mismatch allowed in the tagged primers (module demultiplexer). The paired-end reads were merged using the VSEARCH toolkit (Rognes et al. 2016), with a minimum overlap of 16 base pairs and 5 mismatches allowed (module mergepair-vsearch). Potential chimeras were removed using the UCHIME *de novo* algorithm (Edgar et al. 2011). All sequences with ambiguous bases were removed (module chimera-vsearch). Then the sequences were clustered into Molecular Operational Taxonomic Units (MOTU) using the VSEARCH algorithm (Rognes et al. 2016) with 97% of similarity threshold (module out-vsearch). A MOTUs table was generated, and the representative sequences were compared to our local database using the VSEARCH toolkit and annotated if the minimum similarity between the reference and a sequence was 95%.

The raw data was submitted to the SRA public database under the accession number PRJNA723313.

7.4.3.3. Diversity and biotic indices

For both morphological and the eDNA dataset, the richness (S), the dominance (D), the Shannon Weaver (H) and the equitability (J) were calculated using Past 4.0 (Hammer et al. 2001).

Four biotic indices were calculated to evaluate the EcoQS based on the morphological and the eDNA data: 1) the quality index based on diversity exp(H'_{bc}) i.e. effective number of species (Bouchet et al. 2012 for details); 2) the Foram Stress Index (FSI) based on the relative percentages of two ecological groups according to their tolerance/sensitivity to organic matter enrichment (Dimiza et al. 2016 ; 3) the Tolerant Species index (% TS_{std}) based on the relative proportion of stress-tolerant taxa normalized for the grain size (<63 μm) (Barras et al. 2014; and 4) the Foram-AMBI, an adaptation of the macrofauna AMBI (Borja et al. 2000) to benthic foraminifera, where foraminiferal species are classified in five Ecological Groups (EGs) in relation to their response to organic matter enrichment (Alve et al. 2016; Jorissen et al. 2018; Bouchet et al. 2021). The categorizations to assess the EcoQS are reported in Table 1.

For the eDNA data, exp(H'_{bc}) and Foram-AMBI are called g-exp(H'_{bc}), and g-Foram-AMBI, respectively. The EcoQS class boundaries have not been defined yet for the g-exp(H'_{bc}). As molecular diversity is much higher than morphological one, class boundaries developed for morphological datasets (Schönfeld et al. 2012) cannot be applied. Therefore, an attempt has been made using the Ecological Quality Ratio (EQR) as suggested in the WFD (WFD, 2000/60/EC). The EQR is the ratio between the value of a biological metrics such as g-exp(H'bc) and the expected value under reference conditions (van de Bund and Solimini, 2007). The EQR varies therefore between 0 (i.e., bad EcoQS) and 1 (i.e., high EcoQS). Local-specific reference conditions are the anchor point to calculate the EQR, here we use the sites (B1-20 and B2-20) with lowest deviance from reference conditions to define the boundaries for the five classes of EcoQS (Table 1).

Since not all the MOTUs could be identified at species level and/or only 70 MOTUs are assigned to an ecological group, in the present study we compute the g-Foram-AMBI by assigning MOTUs to EGs following the method proposed by Bouchet et al. (2021). Specifically, the weighted averaging (WA) optimum and tolerance to TOC content were calculated for each MOTUs using the R Software Analogue package (Simpson Gavin et al. 2021). The estimated optimum provides an effective evaluation of the MOTUs environmental requirements. According to Birks et al. (1990), the WA optimum method is rapid, easy, and reliable tool to define the species-specific (i.e., MOTUs) indicative values. Afterwards, MOTUs assignment to EGs was done as follows: if a MOTU had an optimum in the TOC range 0-2%, it was assigned to EGI; in the range 2-2.5%, 2.5-3.4%, 3.4-4.1%, and above 4.1%, it was assigned to EGII, EGIII, EGIV and EGV, respectively (see more details in (Bouchet et al. 2021).

To compare the agreement/disagreement among biotic indices in assessing EcoQS, two EcoQS i.e. 'Acceptable' or 'Not acceptable' are considered following (Blanchet et al. 2008). The 'Acceptable' results from High or Good EcoQS and scores as 1, whereas 'Not acceptable' from Moderate, Poor or Bad EcoQS and scores as 0 (Blanchet et al. 2008; Bouchet and Sauriau 2008). Then, the scores are summed for any station that reflect the level of agreement/disagreement of the biotic indices and categorized (i.e., full agreement 0/7 or 7/7, partial agreement 1/7, 2/7, 5/7 or 6/7, and disagreement 4/7).

7.4.3.4. Statistical analyses and spatial distributional maps

Before performing statistical analyses, the three replicates of each station were pooled for both morphological and molecular datasets. For the morphological dataset, only species with relative abundance > 5% in at least one sample were considered for multivariate statistical purpose. The molecular dataset was filtered considering only MOTUs represented by > 100 reads, in at least one sample.

A chord Diagram was plotted to show the variations in the foraminiferal taxonomic composition among the stations for both morphological and molecular datasets. Chord Diagram was computing using the R software and the package circlize (Gu et al. 2014). A Mantel's test (999 permutations) was performed to test the significance of Spearman's correlation between the relative abundances of morphological foraminiferal species and eDNA MOTUs dissimilarity matrices (Bray-Curtis) using the R Package Vegan (Oksanen et al. 2007). In order to observe the capability of foraminiferal-based biotic indices to reflect the ecological quality, a Spearman's rho matrix of correlation was computed considering *PLI*, TOC, and biotic indices (for both morphological and molecular datasets), using the R package PerformanceAnalytics (Prentice et al. 2005). Prior to statistical analyses all biotic and abiotic data were log-normalized. A Q-mode cluster analysis (CA), based on environmental variables and geochemical indices (i.e., *EF*s, TOC, C/N, mud, *PLI* and mC_d) using the Ward's linkage method and the Euclidean distance, was carried out. Principal Component Analyses (PCAs) were also performed considering environmental variables and geochemical indices as primary variables and the foraminiferal relative abundance from both morphological datasets and molecular (unassigned and assigned MOTUs) as secondary variables. The PCA as well as the CA were performed using the software STATISTICA 13.5.

The distributional maps were generated using the software ArcMap 10.5 (Esri). For plotting the abiotic factors and the geochemical indices, the Inverse Distance Weighted (IDW) method

was used for interpolating the data. This method is fast and easy to compute and to interpret(Lu and Wong 2008). IDW interpolation method was also adopted for creating maps showing the spatial distribution of the biotic indices.

7.5. Results

7.5.1. Environmental characterization

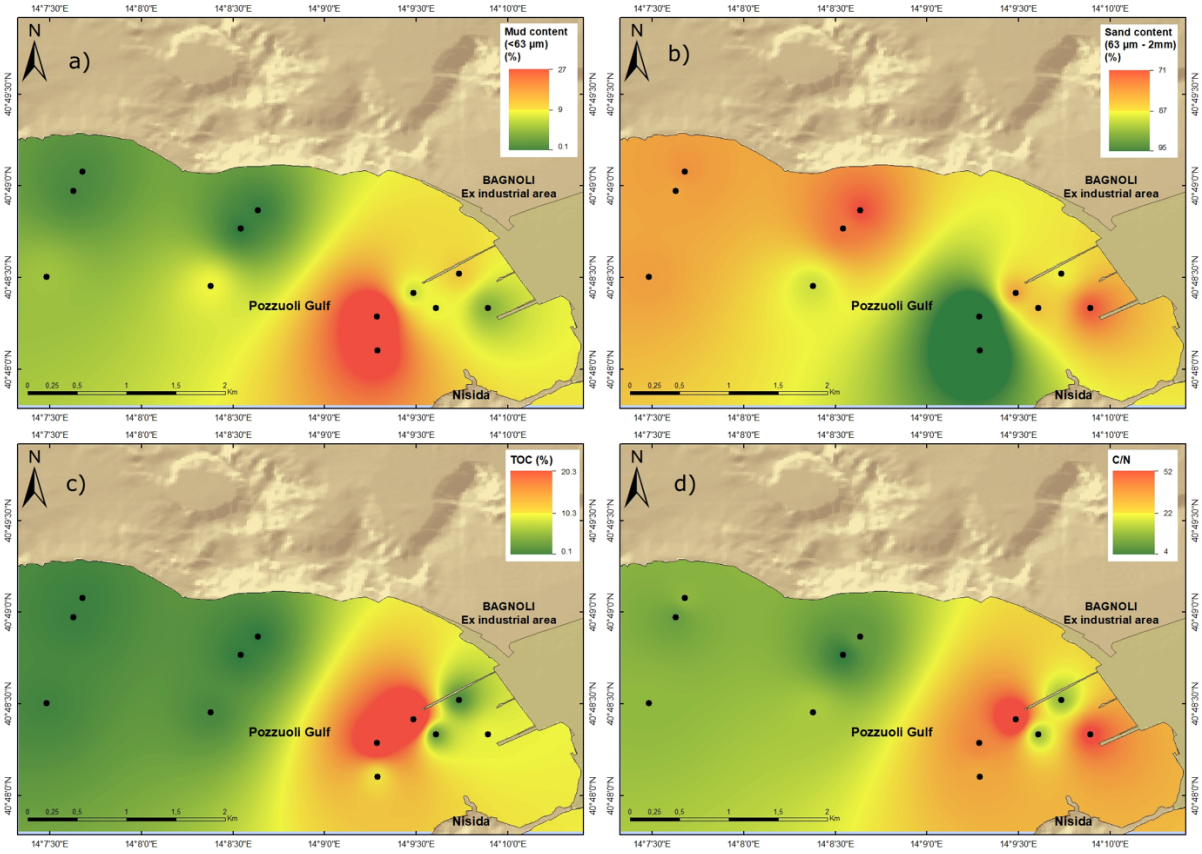

Figure 7.2: Spatial plots of: a) mud content (%); b) sand content (%); c) Total Organic Carbon (TOC) (%); and d) C/N ratio.

The grain-size analysis reveals that in the study area most of the sea bottom is composed by sand. The sandy fraction varies from 71.7 to 95.5%, whereas the mud content (<63 µm) ranges from 0.10 to 27.2% (Fig. 7.2a, b, Tables S3-4). A clear increase of the mud fraction is found seaward with all stations at 40 m exhibiting the highest values, particularly in transects B3 and B4 (Fig. 7.2a). This trend is more evident along transects B3 and B4 in front of the former industrial plant. Indeed, these transects show higher values of mud than transects B1 and B2. The TOC content ranges between 0.1 and 20.3%. The highest percentages of TOC are found in transect B3, particularly at 20 and 40 metres depth (Fig. 7.2c, Table S3). In the external area,

the TOC values strongly decrease compared to B3 stations and ranges from 0.1 to 0.7%. The C/N ratios varies from 5 to 52.5 with the highest values associated with the area in front of the ex-industrial plant (Fig. 7.2d).

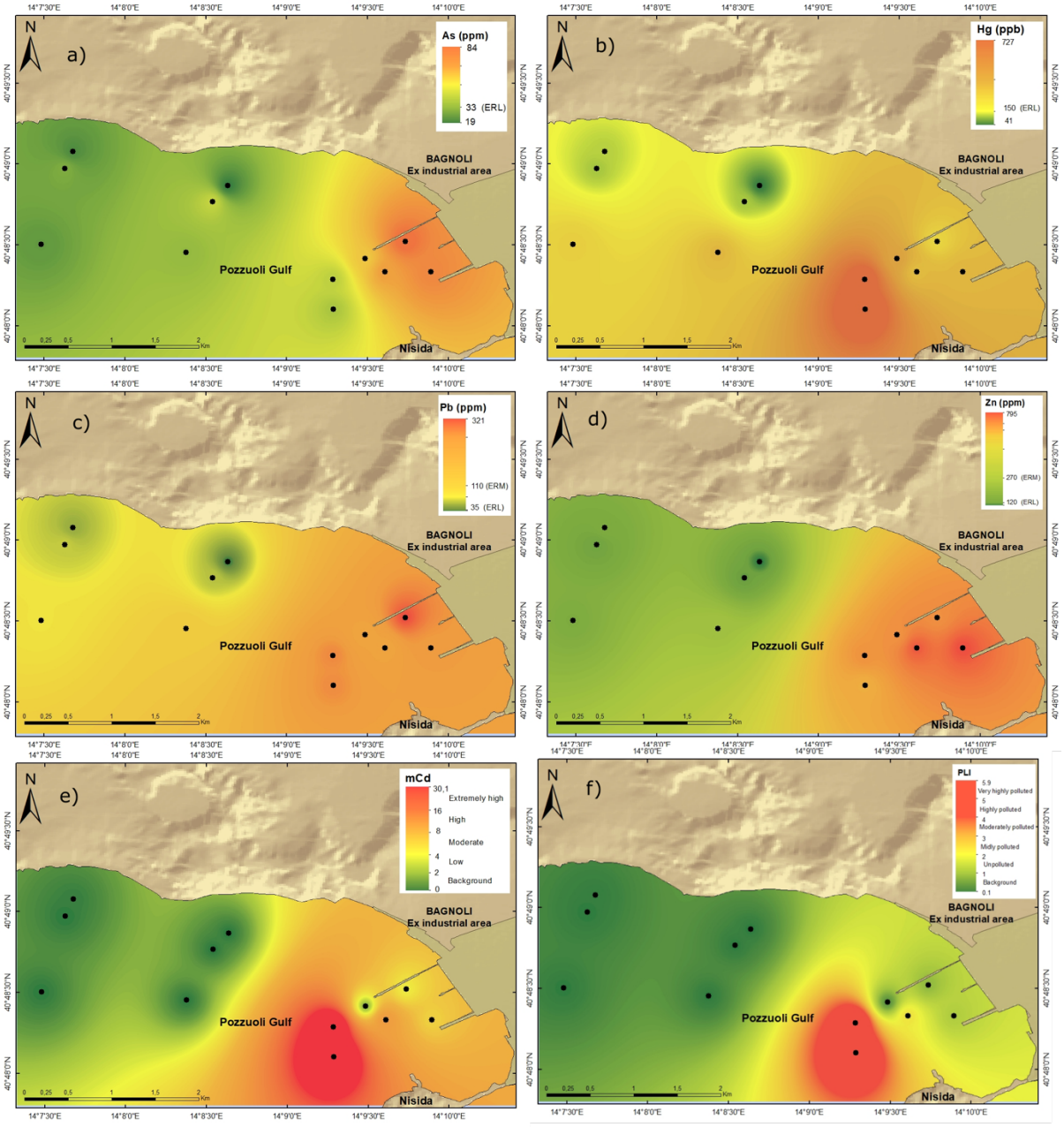

Figure 7.3: Spatial plots of: a) Enrichment Factor of As; b) Enrichment Factor of Hg c) Enrichment Factor of Pb; d) Enrichment Factor of Zn; e) Modified Contamination Degree (mC_d); and f) Pollution Load Index (PLI).

The concentrations of PTEs vary greatly in the study area (Fig. 7.3a-d, Table S3). In particular, Pb and Zn range between 31 and 322 mg/kg and 102 and 795 mg/kg, respectively. The highest

values of As (19–84 mg/kg) and Hg (41–728 ng/kg) are found at stations along transects B3 and B4. The *EF*s of Pb (0.55-7.27) and Zn (0.69-5.79) are constantly higher than 4 at stations in front of the former industrial area (Fig. 7.3c, d). A similar trend has been observed for the *EF*s of As (0.58-3.17) and Hg (0.18-3.07) (Fig. 7.3a, b).

In Bagnoli, the detected concentrations of selected elements (i.e., As, Hg, Pb and Zn) are above the ERM and the ERL in several stations. In particular, As and Hg concentrations are higher than the ERL at 8 stations (Table S1), whereas Pb shows concentrations up to 3 times higher than the ERM in the whole internal area (B3 and B4 transects; 6 stations) and above the ERL at the other 5 stations (Table S1). The concentration of Zn is two times higher than the ERM at 6 stations along the transects B3 and B4, and higher than the ERL at the other 4 stations (Table S1).

The mC_d shows very low values along transect B1 and B2, whereas the highest values are found along transects B3 and B4 (3.98 and 4.14, respectively) at 40 m depth (Fig. 7.3e). The *PLI* varies from 0.61 to 1.22 (Fig. 7.3f). All stations along transects B3 and B4 show *PLI* values greater than 1 and the highest values are associated to the deepest stations. Based on the Q-mode CA, two main clusters (I and II) and two subclusters (IIa and IIb) could be identified (Fig. S1). Cluster I is represented by stations along transects B1 and B2, whereas Cluster II groups stations belonging to transects B3 and B4. Sub-cluster IIa includes all shallower stations (i.e., 10 and 20 m water depth), whereas sub-cluster IIb only stations at 40 m.

7.5.2. Benthic foraminifera

7.5.2.1. Morphological dataset

On overall, 57 species (4 agglutinated, 6 porcelaneous and 47 hyaline) belonging to thirty-six genera are identified. The specific richness (S) values range between 11 and 30 (18.4, on average) with relatively higher values along external transects B1-B2 (19.7, on average) than along B3-B4 (17.1, on average). The H values (1.66-2.76) shows a clear difference between transects B1-B2 and B3-B4; in particular, the lowest values of H are recorded in stations along transects B3 and B4 (Table S5).

The most abundant (%) species are *Bolivina striatula* (14.2% on average), *Haynesina depressula* (12.3% on average), *Rectuvigerina phlegeri* (9.3% on average), *Cibicidoides refulgens* (9.0% on average), *Discorbinella berthelotiana* (7.2% on average), *Elphidium advenum* (5.6% on average), *Bolivina variabilis* (5.4% on average), *Bulimina elongata* (4.9%

on average), *Elphidium crispum* (4.9% on average), *Hopkinsina pacifica* (3.5% on average), *Bolivina spathulata* (2.7% on average)*, Haynesina germanica* (2.6% on average), *Gavelinopsis praegeri* (2.3% on average), *Rosalina bradyi* (2.2% on average), *Nonionoides turgida* (1.6% on average) and *Asterigerinata mamilla* (1.5% on average). Some taxa, namely *B. striatula, H. depressula, E. advenum, B. elongata, H. pacifica*, and *H. germanica* exhibit a higher abundance at stations along transects B3 and B4 compared to those of B1 and B2, whereas an opposite trend is found for *C. refulgens, D. berthelotiana, E. crispum, G. praegeri, R. bradyi* and *A. mamilla* (Fig. 7.4a).

7.5.2.2. eDNA dataset

The total number of raw sequences is 17,857,602, from which 5,112,227 are retained in the downstream analysis after stringent quality filtering. Overall, 12,400 MOTUs are produced by the VSEARCH clustering algorithm. This number has been reduced to 1134 by removing MOTUs represented by less than 100 reads and combining the MOTUs assigned to the same morphospecies. Among them, 1061 are unassigned whereas, 44 MOTUs are represented by the multichambered calcareous or agglutinated Globothalamea, 27 MOTUs by single-chambered, organic-walled or agglutinated Monothalamea and only two MOTUs by Tubothalamea. The number of reads per sample (i.e., the pooled replicates per station) varies from 320178 to 597844. The richness (S) ranges between 426 and 751 (567, on average), with higher values along external sites (583, on average) than along transects B3 and B4 (552, on average) (Table S5). The Shannon-Weaver (H; 3.1-4.8) shows a clear difference between transects B1-B2 (4.45) and B3-B4 (3.89). The highest values of D and the lowest values of H and J are associated with stations along transects B3 and B4 transects.

The assigned MOTUs are mostly represented by monothalamous taxa. The most common monothalamous species are *Bathysiphon* sp*., Micrometula* sp., *Vellaria* sp., *Bathysiphon, Psammophaga* sp., and *Cylindrogullmia alba,* whereas the most common MOTUs assigned to Globothalamea are *Bulimina elongata, Bulimina* sp*., Liebusella goesi, Reophax* sp., *Planorbulinella* sp., *Spiroplectammina* sp., *Buliminella elegantissima, Leptohalysis scotti, Epistominella* spp., *Cibicidoides and Rosalina* sp (Fig. 7.4b). The most abundant (>1 %) unassigned MOTUs, in order of decreasing reads, are 0, 1, 2, 4, 7, 5, 6, 8, 3, 39, 34, 33 and 15. The change in diversity indices values with respect to the transects B1-B2 and B3-B4 is also reflected in the taxonomic composition of foraminiferal assemblages in terms of MOTUs. Accordingly, the transects B1-B2 show higher relative abundance (5.7% B1-B2 *vs*. 4% B3-B4)

of MOTUs assigned to Textulariida, whereas those belonging to Monothalamea exhibit higher relative abundance (1.4% B1-B2 *vs*. 3.5% B3-B4) along transects B3-B4 located in front of the former industrial site (Fig. 7.4b). MOTUs assigned to Globothalamea do not exhibit significant changes across transect but a clear increasing in their abundance has been related to depth (Fig. 7.4c). Indeed, some MOTUs (i.e., *Bathysiphon* sp., *B. elongata*, *Micrometula* sp., *Reophax* sp., *Spiroplectammina* sp., *Vellaria* sp., *L. scottii*, *Psammophaga* sp., *C. alba, Eggerelloides,* and Monothalamea Clade A) show higher number of reads in stations at transects B3 and B4, whereas an opposite trend is found for Textulariida, Rotaliida, *L. goesi*, *Planorbulinella* sp., *Rosalina* sp., *B. elegantissima*, *Cibicidoides*, *Glabratellina* sp. (Fig. 7.4 c,d).

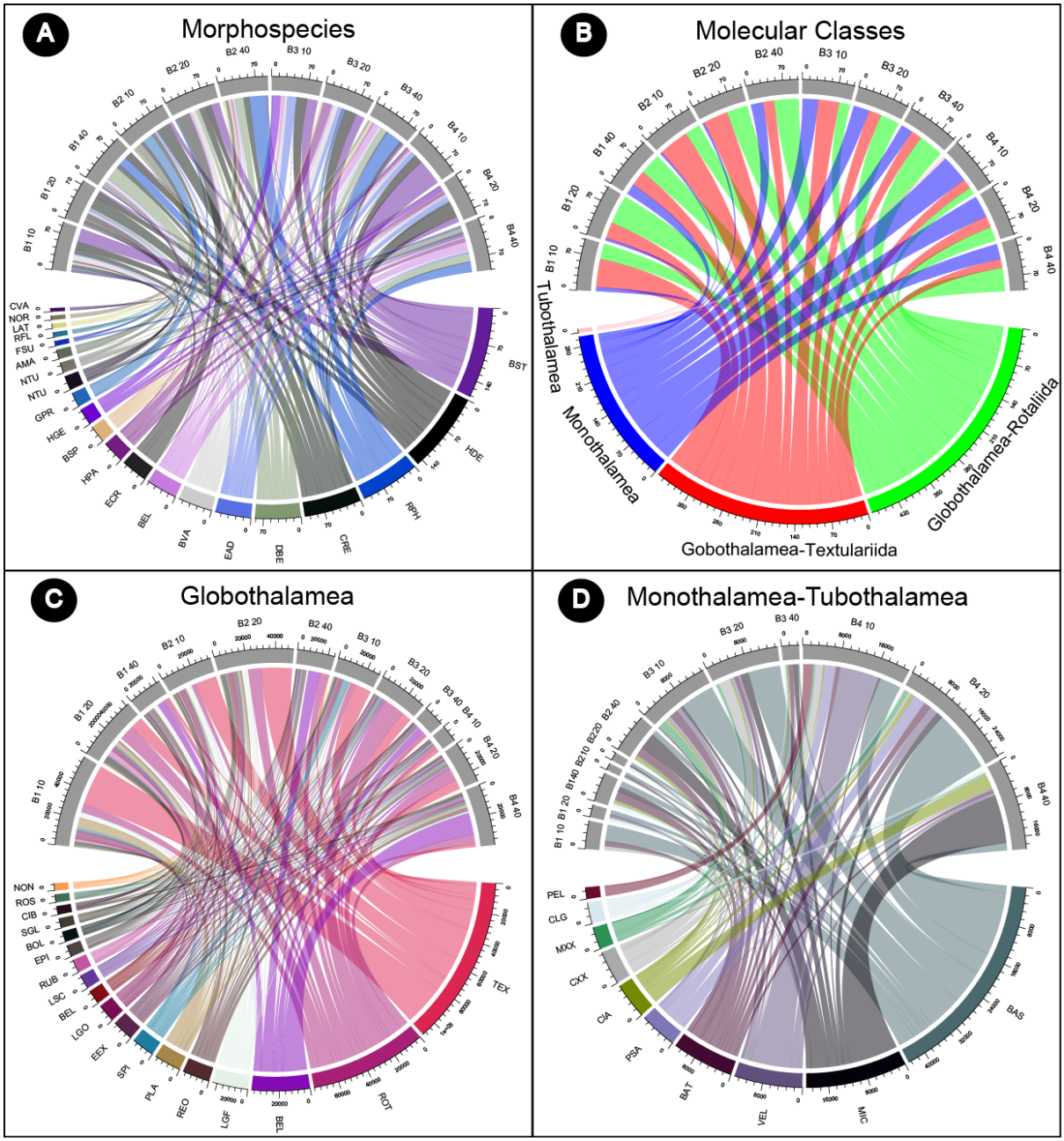

Figure 7.4: Chord diagram based on a) Morphospecies and the eDNA dataset: b) Molecular classes; b) Globothalamea; and d) Monothalamea-Tubothalamea; -The top part shows the stations, while the lower part refers to morphospecies and MOTUs. The complete list of abbreviations is reported in Table S9.

7.5.2.3. Relationship between environmental parameters and benthic foraminifera

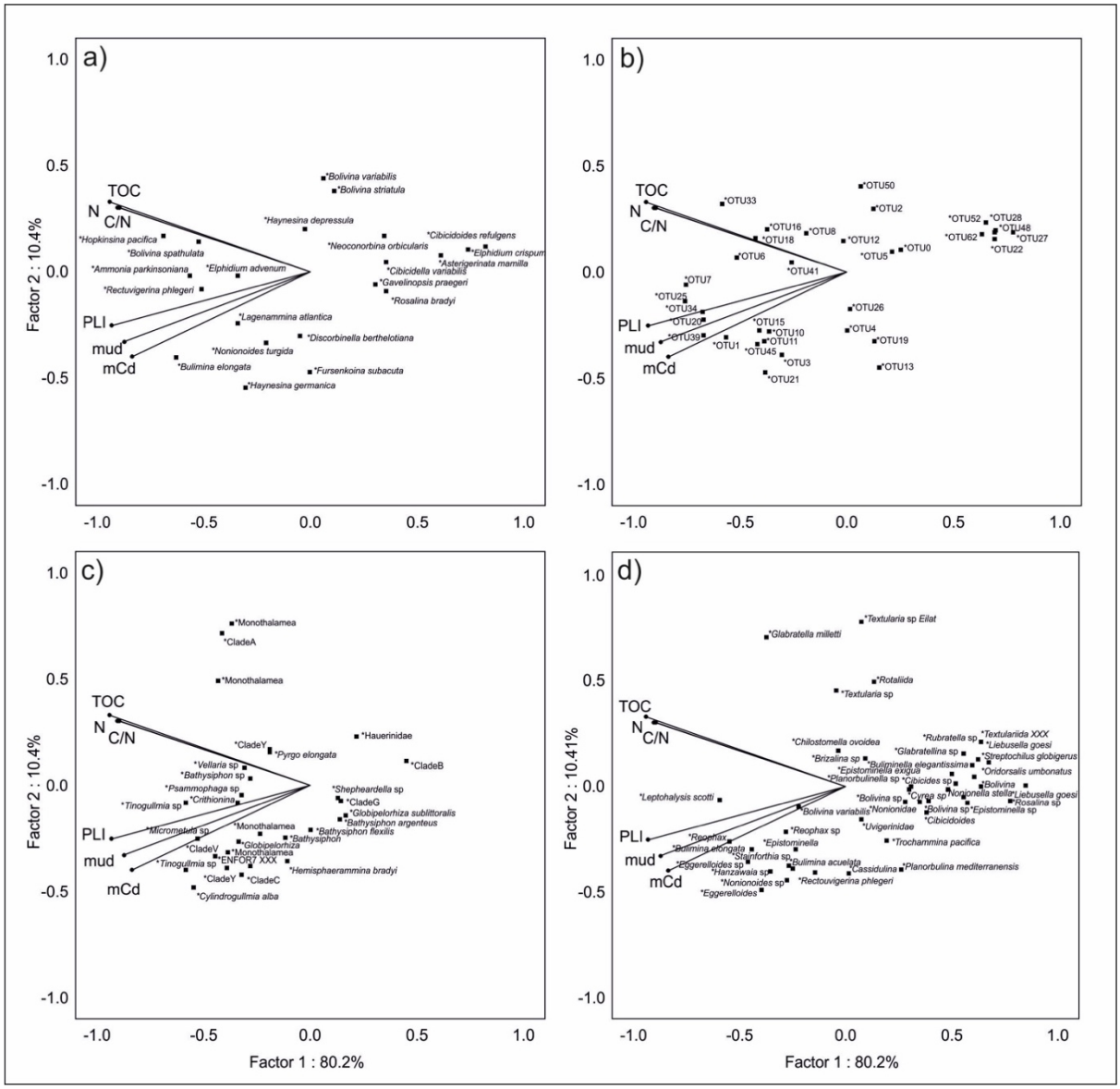

*Figure 7.5: PCA diagram based on environmental parameters and a) morphospecies, b) unassigned OTUs > 5%, c) assigned Monothalamea, and d) assigned Globothalamea. Secondary variables are marked with an *.*

Mantel's test shows a significant positive correlation (r = 0.7; *p* value < 0.001) between the molecular and morphological dissimilarity matrices. In the PCA, the first two axes explain ~90.6% of the total variance (Fig. 7.5). The environmental parameters and geochemical indices

are strongly related to axis 1. On the basis of the PCAs, the considered environmental parameters, namely *PLI*, mC_d, TOC, C/N, and mud are strongly related to the first axis that can be therefore interpreted as the environmental stress (ES) component.

By projecting the morphospecies (Fig. 7.5a) and the MOTUs (Fig. 7.5 b-d) on the PCA plane, an increase in the relative abundances of some morphospecies and MOTUs towards negative values of the first component can be observed. In particular, some morphospecies (i.e., *B. elongata, N. turgida, H. pacifica, B. spathulata, A. parkinsoniana, R. phlegeri* and *E. advenum*) are negatively related to this axis (Fig. 7.5a). An opposite trend is found for *E. crispum, C. refulgens, A. mammilla, C. variabilis, R. bradyi,* and *G. praegeri.* The unassigned MOTUs (i.e., 7, 25, 34, 20, 39, 1, 6 and 33) and MOTUs assigned to Monothalamea (i.e., *Tinogullmia* sp., *Micrometula* sp., *C. alba*, *Vellaria* sp., *Psammophaga* sp., *Bathysiphon* sp. and Clade V) and Globothalamea (i.e., *L. scottii*, *Reophax*, *B. elongata*, *Eggerelloides* sp., *Nonionoides* sp., and *Stainforthia* sp.) are negatively related to the first component (Fig. 7.5b-d). On the other hand, the unassigned MOTUs (i.e., 52, 62, 28, 48, 27 and 22) and MOTUs assigned to Monothalamea (i.e., Clade B, Clade G, *Globipelorhiza sublittoralis*, *Shepheardella* sp., and *Bathysiphon argenteus*), Tubothalamea (Hauerinidae) and Globothalamea (i.e., *L. goesi*, Textulariida, *O. umbonatus*, *Rosalina* sp., *Epistominella* sp., *Cibicidoides* and *Cibicides* sp.) are positively related with the first component (Fig. 7.5b-d).

7.5.2.4. Biotic indices

The exp(H'_{bc}) based on morphospecies varies from 26.7 (station B1-20) to 8.2 (station B1-10) (Table 2). EcoQS ranges between excellent and poor, without showing a clear trend in the study area. For FSI, $\%TS_{std}$ and Foram-AMBI the percentage of assigned species to EGs was higher than 80%. The FSI values vary from 2 to 9.1 with higher values at stations along transects B1-B2 (5.9 ± 2.0%) then the ones along B3-B4 transects (2.9 ± 0.6%). EcoQS varies from excellent and moderate in transects B1-B2 and from moderate to poor in B3-B4 (Table 2). The values of $\%TS_{std}$ vary from 9.6 to 84.5% with markedly higher values (75.4 ± 7.8%) along the transects B3 and B4 than the transects B1 and B2 (42 ± 21%) (Table 2). EcoQS varies from good to poor along the transects B1 and B2, while it is persistently bad along the transects B3 and B4. The Foram-AMBI ranges from 0.35 to 2.4 with a similar trend of TS_{std}. The relative EcoQS ranges between excellent and good in transects B1-B2 and good in B3-B4 (Table 2). As this result overestimates the EcoQS in the study area, the new boundary classes (Foram-AMBI*) proposed by Parent (2019) are here considered (Table 1). On the basis of the new boundary

classes definition, EcoQS ranges from excellent to good along the transects B1 and B2 and from good to moderate along transects B3 and B4 (Table 2).

The g-exp(H'$_{bc}$) (i.e., based on eDNA) varies from 21.8 (station B3-40) to 121.9 (station B2-10) with comparatively much higher values in transects B1-B2 (88.8 ± 27) than B3-B4 (53.1 ± 21) (Table 2). EcoQS varies from excellent and moderate in transects B1-B2 and from moderate to bad in transects B3-B4. Two g-Foram-AMBI were calculated: the first is based on only assigned MOTUs (g-Foram-AMBI), whereas the second one (g-Foram-AMBI-MOTUs) considers all MOTUs regardless their assignment (Tables S7-8). The g-Foram-AMBI and g-Foram-AMBI-MOTUs vary between 1.5 (station B2-10) and 5.6 (stations B3-20 and B4-10) and 0.4 (station B2-10) and 5.5 (station B3-40) (Table 2). For g-Foram-AMBI, EcoQS varies from good and bad in transects B1-B2 and is from poor to bad in transects B3-B4. For g-Foram-AMBI-MOTUs, EcoQS varies from excellent to moderate in transects B1-B2 and from moderate to bad in B3-B4 (Table 2). These two indices are significantly and positively correlated (r= 0.92, *p* value < 0.001). They show a clear trend towards higher values in all sites in front of the former industrial plant, where values are constantly higher than 3.7 and 4.9 for g-Foram-AMBI and g-Foram-AMBI-MOTUs, respectively.

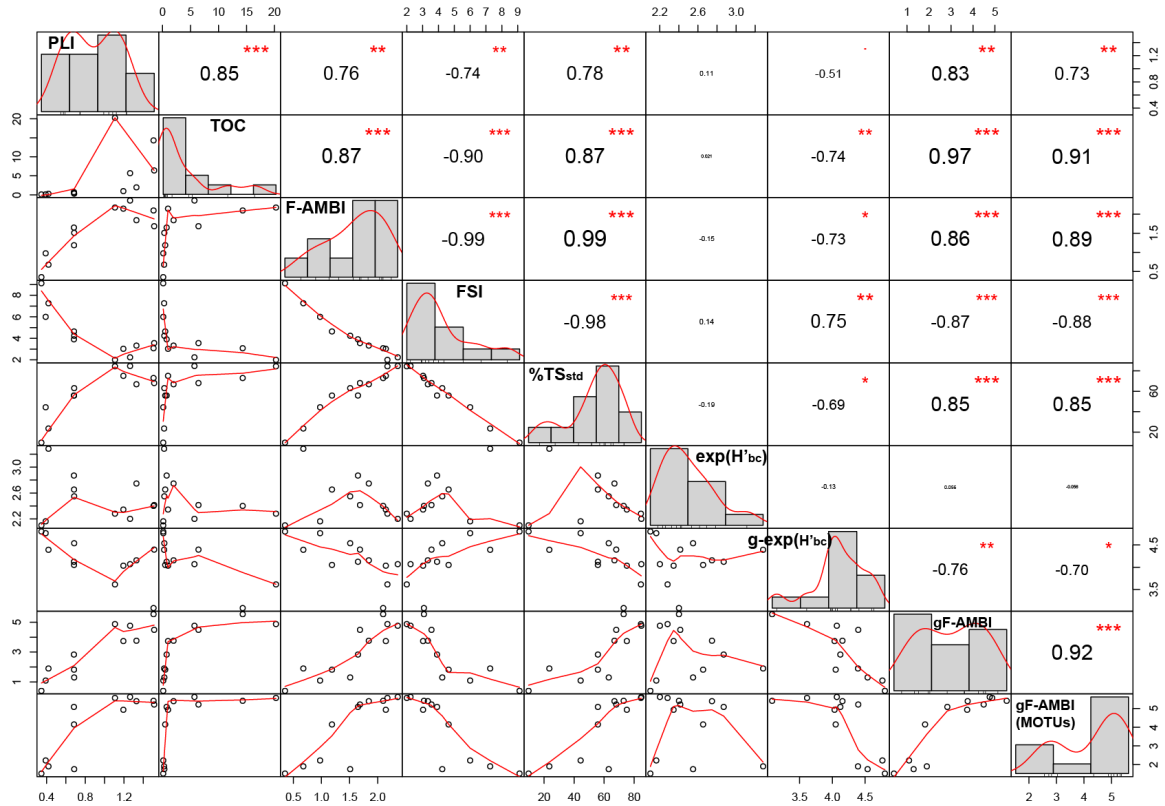

*Figure 7.6: Spearman's rho matrix of correlation between environmental parameters (i.e., PLI and TOC) and biotic indices (i.e., Foram-AMBI, FSI, TS_{std}, exp(H'bc), g-exp(H'bc), g-Foram-AMBI and g-Foram-AMBI MOTUs). The distribution (histogram) of each variable is shown on the diagonal. On the bottom of the diagonal: the bivariate scatter plots with a fitted line are displayed. On the top of the diagonal: the Rho values and the relative p-values are displayed. Significance level are reported as stars: 0.001, 0.01, 0.05, 0.1, 1 <=> "***", "**", "*", ".", " ".*

The Spearman's rho matrix of correlation highlights significant correlations between *PLI* and all ecological indices, except for exp(H'_{bc}) (Fig. 7.6). Similarly, TOC has a significant correlation with all different indices except for exp(H'_{bc}). In particular, the Foram-AMBI, $\%TS_{std}$, g-Foram-AMBI and g-Foram-AMBI-MOTUs are positively correlated with the *PLI* (r= 0.76, 0.78, 0.83, 0.78, respectively) and the TOC (r= 0.87, 0.87, 0.97, 0.91, respectively). On the other hand, FSI and g-exp(H'_{bc}) are negatively correlated to *PLI* (r= -0.74 and -0.51, respectively) and TOC (r= -0.90 and -0.74, respectively) (Fig. 7.6). Indeed, the ecological indices were plotted and correlated to the ES (Fig. 7.7 and Table 7.2). All ecological indices, except exp(H'_{bc}), are significantly correlated to the ES. The highest correlations were found between the ES and g-Foram-AMBI (r=-0.95) and g-Foram-AMBI-MOTUs (r=-0.86) (Table 2).

Indexes	B1_10	B1_20	B1_40	B2_10	B2_20	B2_40	B3_10	B3_20	B3_40	B4_10	B4_20	B4_40	Spearman's rho correlation to ES	*p*. level
exp(H'bc)	12.8	26.7	14.2	8.2	8.7	17.6	10.4	9.8	11.0	9.0	15.6	11.2	-0.05	0.880
Foram Stress Index	4.2	7.3	4.6	9.1	6.0	3.9	3.0	2.0	3.1	2.2	3.3	3.6	**0.83**	0.001
%TSstd	63.1	23.5	55.8	9.6	44.4	55.9	75.2	84.5	73.0	84.8	66.9	68.1	**-0.83**	0.001
Foram-AMBI	1.5	0.7	1.2	0.4	1.0	1.6	2.1	2.2	2.1	2.4	1.8	1.7	**-0.82**	0.001
Foram-AMBI*	1.5	0.7	1.2	0.4	1.0	1.6	2.1	2.2	2.1	2.4	1.8	1.7	**-0.82**	0.001
g-exp(H'bc)*	93.6	80.8	57.5	121.9	117.3	61.7	56.5	37.1	21.8	58.5	63.7	81.1	**0.69**	0.013
g-Foram-AMBI*	1.8	1.9	4.1	1.5	2.2	5.1	4.9	5.6	5.4	5.6	5.4	5.2	**-0.86**	0.000
g-Foram-AMBI-MOTUs*	1.3	1.9	1.8	0.4	1.1	2.8	3.7	4.9	5.5	4.8	3.8	4.5	**-0.95**	0.000
Sum of score/indexes	4/7	3/7	1/7	6/7	4/7	2/7	0/7	0/7	0/7	0/7	1/7	1/7	Full: 4/12 - Partial: 5/12 - Disagreement: 3/12	
Interpretation	disagreement	disagreement	partial	partial	disagreement	partial	full	full	full	full	partial	partial	Full: 33.3% - Partial: 41.7% - Disagreement: 25%	

Table 7-1 Ecological Quality Status calculated on the selected indices (see Table 1 for categorization). Spearman's rho correlation coefficients between indices and environmental stress (ES that is the scores of PCA1) are also reported. Bold values p. <0.05. The sum of the score agreement (Acceptable: High or Good EcoQS scored as 1 and Not acceptable: Moderate to Bad EcoQS scored as 0) is reported with the interpretation of the agreement (i.e., Full 0/7 or 7/7, Partial 1/7, 2/7, 5/7 or 6/7, and disagreement 4/7).

7.6. Discussion

7.6.1. Environmental characterization of the SIN of Bagnoli

On the basis of our geochemical analyses very high concentrations of several PTEs (i.e., As, Hg, Pb and Zn) have been identified in Bagnoli area, particularly in front of the former industrial plant. The spatial distribution of As, Hg, Pb and Zn matches well with previous

assessments in the area (Romano et al. 2009; Trifuoggi et al. 2017; Sprovieri et al. 2020). The distribution of *PLI* values indicates low level of pollution in the area corresponding to the transects B1 and B2 (Fig. 7.2). Values above 1 are constantly found along the transects B3 and B4, defining the area as polluted (Jahan and Strezov 2019). The mC_d distribution highlights a clear separation between the northern area (transects B1 and B2), where a low degree of contamination occurs and the stations along the B3 and B4 characterized by a low to high (B4 at 40 m) mC_d. The separation is also supported by the Q-mode CA. The pollution level appears to be influenced by the distance from the ex-industrial area and depth (i.e., increasing toward the deeper stations). This increase can be ascribed to the sediment characteristics such as mud and organic matter contents (Savvides et al. 1995). In fact, the highest values of mC_d and *PLI* recorded at 40 m of depth, between the two piers, are probably linked to the higher content of mud and organic matter that are capable to adsorb PTEs. The C/N ratio is commonly used as a proxy to distinguish the source of organic matter in marine environments (Gordon and Goni 2003), where values higher than 20 are likely associated to a terrestrial source of organic matter (Meyers 1994). In the study area, the C/N ratio is extremely high in the marine sediments, particularly in all stations along transects B3 and B4 40 m depth in front of Bagnoli ex-industrial area (B3 and B4 transects), suggesting a probable terrestrial carbon input and/or a significant contamination by hydrocarbons (with a highly unbalanced C/N ratios).

Following the PCA score plot representing the ES (Fig. 7.7a), it is clear that the most stressful conditions occur in the area in front of the former industrial plant (i.e., B3 and B4 transects) and the stations that result as the most polluted are B3-20 m, and B3-40 m and B4-40 m. On the other hand, the lowest ES is found in the northern area (B1 and B2) with the highest score values at shallower stations (i.e., 10 and 20 m). The marine sediments along the transects B3 and B4 are affected by the continental source of pollution from the former plant area. Indeed, the trace elements are mobilized from the Bagnoli plain (i.e., ex-plant soils) and transported westward toward the deeper sea sediments along an axe that is parallel to the northern pier (Trifuoggi et al. 2017; Sprovieri et al. 2020). The morphological features of the sea bottom coupled with the marine currents could be responsible of the sedimentary dispersal mechanism from the inshore-contaminated area toward the deeper environments (Trifuoggi et al. 2017). This ultimately promotes the accumulation of PTEs at 40 m water depth stations at transects B3 and B4. Although the northern area (transects B1 and B2) has been identified as relatively low to not polluted, minor contamination input may affect the marine sediments due to a secondary current circulation from the former plant toward NW (Sprovieri et al. 2020).

According to this model, sediments of the transects B1 and B2 at 40 m of depth could be also affected by the same source of contamination of transects B3 and B4, but at lower extent.

7.6.2. Foraminiferal metabarcoding vs. morphology

7.6.2.1. Congruent trend of diversity metrics

The clear separation between B1-B2 transects (low- to unpolluted) and B3-B4 ones (highly polluted) based on environmental parameters (Fig. S1) is also reflected in benthic foraminiferal assemblages of both the molecular and morphological datasets. Accordingly, lower values of diversity (i.e., H and S) are recorded in front of the former industrial site that results as the most polluted area. A decrease in morphospecies foraminiferal diversity is commonly associated to stress conditions (i.e., unstable physico-chemical parameters of water, oxygen deficiency, pollution) (Alve 1995). Similarly, a lowering of MOTUs diversity has been associated to deteriorating environmental quality in response to fish farming (Pawlowski et al. 2014a; Pochon et al. 2015), gas and oil drilling activities (Laroche et al. 2016, 2018) and gas platforms (Cordier et al. 2019c; Frontalini et al. 2020b). In our study, the decrease in diversity values identified in the morphological dataset is also recognized in the molecular one that it is also more evident. In particular, the more pronounced reduction of diversity (i.e., exp (H'_{bc})) in the molecular dataset can be ascribed to the higher number of MOTUs compared to the number of morphospecies in the morphological one. In fact, the molecular dataset includes soft-shelled (i.e., naked monothalamids) species that are commonly neglected in the traditional morphological approach as well as small-size taxa (i.e., <63 μm) that are mostly overlooked (Pawlowski et al. 2014a).

7.6.2.2. Foraminifera as bioindicators

Similar to previous studies in the area (Bergamin et al. 2003, 2005; Romano et al. 2008, 2009) a limited number of living specimens has been found and most of them in the finer fraction (i.e., 63-125 μm). This can be likely ascribed to the prevailing high pollution conditions. It is worth to mention that differently from the previous studies (based on total assemblages), the present investigation is only based on living specimens, providing reliable information about local ecological conditions at the moment of the sampling.

The relative abundances of some morphospecies and assigned and unassigned MOTUs vary in relation to the pollution gradient. In fact, some morphospecies (i.e., *B. elongata, N. turgida,*

H. pacifica, B. spathulata, A. parkinsoniana, R. phlegeri, and *E. advenum*) appear to be tolerant to environmental stress. All these taxa except *A. parkinsoniana* and *E. advenum* have been identified as opportunist (i.e., from first to third order) (Jorissen et al. 2018). On the other hand, some taxa (i.e., *E. crispum, C. refulgens, A. mammilla, C. variabilis,* and *R. bradyi*) show a more sensitive behaviour in accordance with their assignment to the sensitive ecological group in Jorissen et al. (2018)

Monothalamea appears to be more abundant at the more polluted stations, whereas an opposite trend is observed for Globothalamea, particularly for Textulariida (Fig. 7.4b). Some Monothalamea (i.e., *Tinogullmia* sp., *Micrometula* sp., *C. alba*, *Vellaria* sp., *Psammophaga* sp., *Bathysiphon* sp. and Clade V) and Globothalamea (i.e., *L. scottii*, *Reophax*, *B. elongata*, *Eggerelloides* sp., *Nonionoides* sp., and *Stainforthia* sp.) are positively related to the stress gradient and have been therefore identified as tolerant species (Fig. 7.5). The MOTU assigned to *Leptohalysis scotti* has been found to be well adapted to impacted conditions (i.e., enriched sediment) of fish farming (Pawlowski et al. 2014a; Pochon et al. 2015). The same MOTU has been revealed to be related to finer sediment fractions (i.e., silt, clay and mud) (Frontalini et al., 2020) where likely oxygen levels are lower. Moreover, even in the traditional morphological approach, *L. scotti* has been included in the first- (Jorissen et al. 2018), second- (Bouchet et al., 2018b) and third-order opportunist (Parent et al. 2021). Similarly, the MOTU assigned to *Vellaria* sp. has been reported to be well adapted to impacted conditions (Pawlowski et al., 2014a; Pochon et al., 2015) and high concentration of Hg (Frontalini et al., 2018). Higher abundance of the MOTU assigned to *Psammophaga* sp. has been documented near fish farms (i.e., Pawlowski et al., 2014) and associated to moderately organic-enriched sediment (Pochon et al., 2015). Higher read numbers of the MOTU assigned to *Bathysiphon* sp. have been found close to an oil drilling site (Laroche et al., 2016). The MOTU assigned to *Micrometula* sp. has been associated to lower organic enrichment and with a preference to oxic conditions (Pawlowski et al., 2014a; Pochon et al., 2015). Similarly, the morphospecies *Micrometula hyalostriata* has been found in sites with low TOC and high bottom-water oxygen concentrations and this species has been therefore identified as sensitive (Bouchet et al., 2018b). On the contrary, in the morphological approach, *Micrometula* spp. including *M. hyalostriata* and *Micrometula* sp. has been assigned to the first-order opportunists (group V) (Fossile et al., 2021). The morphospecies *Bulimina elongata* has also been identified as an opportunistic taxon (Jorissen et al., 2018). The morphospecies *Cylindrogullmia alba* is reported

in environments with high to low TOC and bottom-water oxygen concentrations and identified as indifferent species (Bouchet et al., 2018b).

Other MOTUs belonging to Monothalamea (i.e., Clade B, Clade G, *G. sublittoralis*, *Shepheardella* sp., and *B. argenteus*), Tubothalamea (Hauerinidae) and Globothalamea (i.e., *L. goesi*, *O. umbonatus*, *Rosalina* sp., *Epistominella* sp., *Cibicidoides* and *Cibicides* sp.) show a more sensitive behavior. Most of these taxa have been rarely found in previous eDNA metabarcoding studies and a comparison of their ecology and driving factors is therefore limited. Interestingly, the MOTU assigned to *Epistominella* spp. has been found to be more prevalent at the high-flow (sandy) site (Pochon et al., 2015) so it matches well with its higher abundance at B1 and B2 transects that present, comparatively, coarser grain-size and lower organic enrichment. On the other hand, the MOTU assigned to *B. argenteus* has been found as more common at moderately impacted low-flow sites (Pochon et al., 2015). Given that, the ecological behavior of these MOTUs has been compared to the morphospecies and it is also worth to mention that most of the morphospecies belonging to *Rosalina* and *Cibicides* genera have been regarded as sensitive species and therefore assigned to group I (Jorissen et al., 2018). *Liebusella goesi* has been recognized as tolerant taxon (Bouchet et al., 2018b).

Despite the different taxonomic composition of the metabarcoding and morphology assemblages, some congruent trends can be identified (Fig. 7.5). This is supported by the significant positive correlation (i.e., Mantel's test) between the molecular and morphological datasets suggesting a clear and similar shift in the two dissimilarity matrices (i.e., foraminiferal turnover) in the study area. In particular, *B. elongata* is placed in the same PCA position (i.e., negative values of PCA1) suggesting a clear match in its ecological behaviour in the morphological and molecular dataset. *Nonionoides turgida* (morphological dataset) and *Nonionoides* (molecular dataset) show a very similar trend as well as at some extent *R. phlegeri*. Moreover, *R. bradyi*, *C. refulgens*, *C. variabilis* (morphological dataset) and *Cibicides*, *Cibicidoides*, and *Rosalina* sp. (molecular dataset) show a very congruent pattern towards positive value of PCA1.

7.6.2.3. Assessing the Ecological Quality Status of Bagnoli based on benthic foraminiferal biotic indices

All the biotic indices computed for both morphological and molecular datasets show congruent trends with the ES (Fig. 7.7a and Table 7.2). In fact, lower EcoQS values are found in the area in front of the former industrial site (i.e., stations along B3 and B4 transects) (Fig. 7b-e). All

these indices except exp(H'bc) exhibit a significant Spearman's rho correlation to *PLI*, TOC and the ES (Fig. 7.6 and Table 2). The diversity-based index exp(H'_{bc}) has been successfully applied to assess the EcoQS in different marine settings (Bouchet et al. 2012, 2018; Francescangeli et al. 2016; Dijkstra et al. 2017; S. dos S. de Jesus et al. 2020). However, it also provided controversial results in previous environmental characterizations (el Kateb et al. 2020). On the contrary, the g-exp(H'_{bc}) based on foraminiferal eDNA shows a significant negative correlation to *PLI*, TOC and ES. This can be ascribed by the higher number of taxa (i.e., MOTUs) and "specimens" (i.e., reads) compared to the morphological dataset. Indeed, naked and small-size foraminifera that are commonly overlooked in the dry-picked morphology-based assessment are retained in the eDNA approach (Pawlowski et al., 2014). The application of metabarcoding provides a more holistic view of foraminiferal diversity (He et al., 2019: Frontalini et al., 2020).

The recent advances in the foraminiferal morphospecies' assignments to ecological groups (EGs) (Alve et al., 2016; Bouchet et al., 2021; Jorissen et al., 2018) significantly facilitate the implementation of the sensitive-based indices (FSI, $\%TS_{std}$, Foram-AMBI). The Foram-AMBI may indeed suffer of a lacking of foraminiferal assignments to EGs invalidating its proper effectiveness, when the percentage of assigned species is lower than 80% (Borja and Muxika 2005). For instance, in the Gulf of Gabes (Tunsia) (El Kateb et al., 2020) and in the Gulf of Manfredonia (Italy) (Fossile et al. 2021), the assessment of the species was often under this threshold. On the contrary, in our study, the percentage of unassigned species for FSI, $\%TS_{std}$ and Foram-AMBI is always > 80%, providing reliable ecological quality assessments.

For evaluating the EcoQS, one of the key aspects is not only the choice of the metrics as well as the definition of proper class boundaries (i.e., categories) for each metrics. Boundary setting is one of the most critical step in the design of assessment methods as it defines the target values for environmental management (Birk et al., 2012).The EcoQS based on g-exp(H'_{bc}) fits very well with the ES in the area (Fig. 7.7) as well as with the categorization for the definition of the EcoQS. The class boundaries of Foram-AMBI have been derived from the AMBI, a well-known biotic index based on macrofauna (Borja et al., 2003), Table 1). However, if the same macrofaunal categories are applied, the EcoQS of the study area results from excellent (i.e., in the area away from the industrial plant) to good (in correspondence of the former plant) (Fig. 7.7). Similar controversial results have been reported in the assessment of the EcoQS in the Gulf of Manfredonia (Italy) (Fossile et al., 2021). It is evident that a redefinition of boundaries is urgently required in order to optimize the application of Foram-AMBI in

biomonitoring. In light of it, we follow the new categorization of Parent (2019) (Tables 1 and 2). On the basis of the new categorization, the EcoQS varies from moderate to high (Fig. 7.7). In particular, the area in front of the former industrial plant exhibits an EcoQS ranging from good to moderate that matches much better with those computed from $\%TS_{std}$ as well as of the molecular dataset (i.e., g-exp(H'_{bc}), g-Foram-AMBI and g-Foram-AMBI-MOTUs). The g-Foram-AMBI and g-Foram-AMBI-MOTUs evidence critical ecological conditions (poor-bad) in the stations with the highest degrees of pollution (Fig. 7.7) that also match well with the EcoQS derived from g-exp(H'_{bc}). The comparison of EcoQS (acceptable or not acceptable) resulting from the different biotic indices suggests the lowest percentages of agreement for Foram-AMBI with the traditional classes and all other ecological indices, whereas the highest agreement has been found between g-Foram-AMBI and all the other indices. When, the Foram-AMBI with the traditional classes is removed, EcoQS of the stations results with full agreement in 4 out of 12 stations (about 33%) and partial agreement in 5 out of 12 stations (ca. 42%) (Table 2).

7.1. Conclusions

The innovative foraminiferal eDNA metabarcoding has been used to assess the EcoQS in the highly polluted SIN of Bagnoli and compared to the more traditional morphology-based biomonitoring. Geochemical and statistical analyses identify the area, in front of the former industrial plant, as the most polluted. Significant changes in both morphological and molecular datasets are associated to environmental stress (i.e., high trace element and TOC levels) and are well spatially defined (i.e., in front *vs*. away from the former industrial plant). Similarly, the selected ecological indices inferred from both morphological and molecular datasets strikingly and congruently resulted in a clear separation following the environmental stress gradient. The molecular indices (i.e., g-exp(H'_{bc}), g-Foram-AMBI and g-Foram-AMBI-MOTUs) reliably identify moderate to bad EcoQS in the polluted area and good to high EcoQS in the area more distant to the former plant, particularly those sites at shallower water depth (i.e., 10 and 20 m). The morphology-based Foram-AMBI seems to well register the overall trends of EcoQS but at the same time it seems to overestimate the EcoQS if the traditional class boundaries are considered. The congruent and complementary trend between morphological and molecular biotic indices further support the application of foraminiferal metabarcoding in routine biomonitoring to assess the environmental impacts of heavily polluted marine areas.

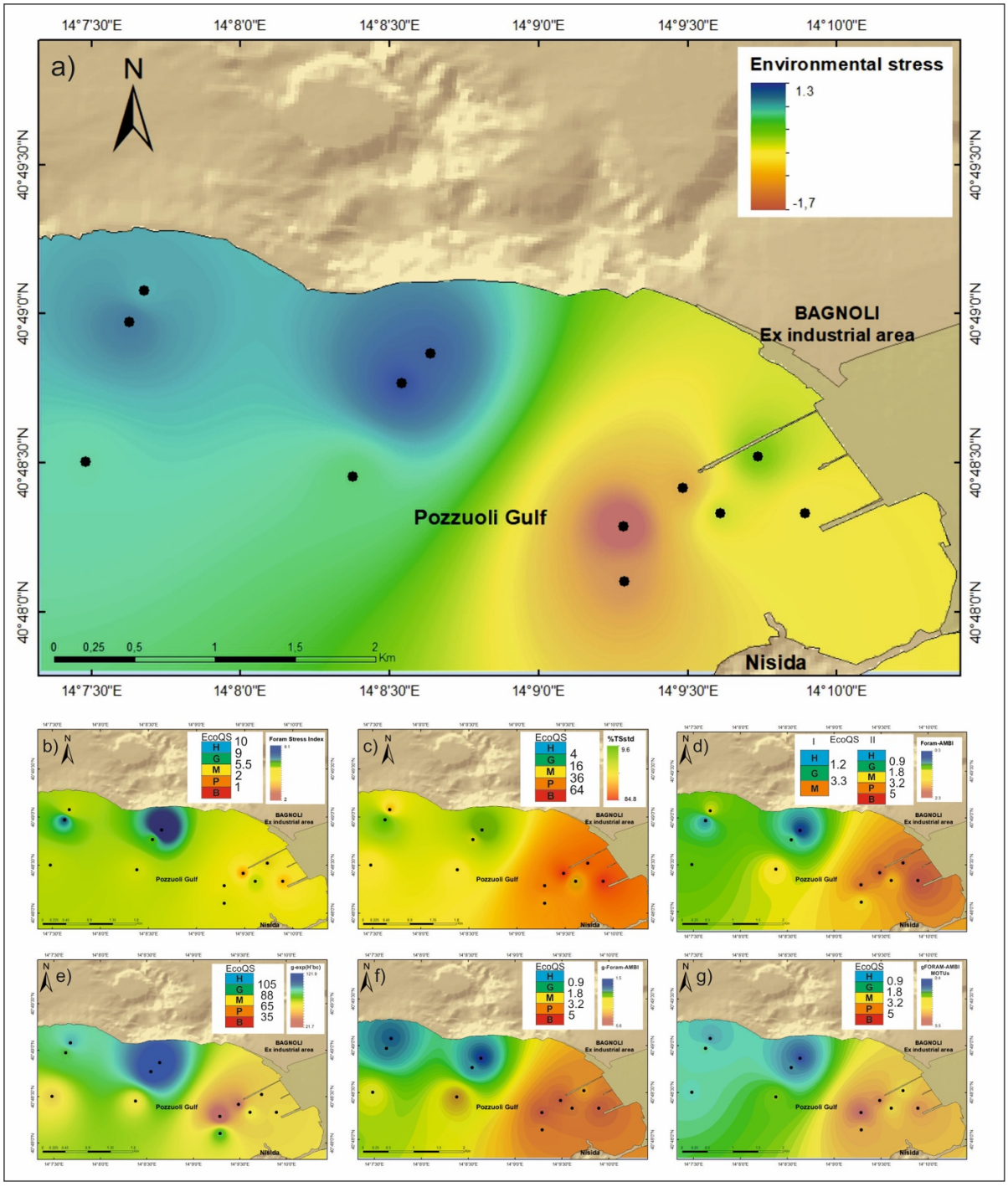

Figure 7.7: Spatial plots of a) Environmental stress (ES) based on PCA first axis scores, b) Foram Stress Index, c) %TSstd, d) Foram-AMBI, e) g-exp(H'$_{bc}$), f) g-Foram-AMBI and d) g-Foram-AMBI MOTUs. The Ecological Quality status categorizations are also reported following Table 1. Two categorizations I and II are provided for Foram-AMBI based on Borja et al. (2003) and Parent (2019), respectively.

Table 7-2: Indices categorization to assess the Ecological Quality status: %TSstd (Barras et al., 2014;

Index	Ecological Quality Status				
	High	**Good**	**Moderate**	**Poor**	**Bad**
exp(H'bc)	<20	15-20	10-15	5-10	<5
Foram Stress Index	9-10	5.5-9	2-5.5	2-1	<1
%TSstd	<4	4-16	16-36	36-64	>64
Foram-AMBI	<1.2	1.2-3.3	3.3-4.3	4.3-5.5	>5.5
Foram-AMBI*and g-Foram-AMBI*	<0.9	0.9-1.8	1.8-3.2	3.2-5	>5
g-exp(H'bc)*	>105	105-88	88-65	65-35	<35

*Parent et al., 2021), exp(H'bc) (Bouchet et al., 2012), Foram Stress Index (Dimiza et al., 2016), Foram-AMBI (Borja et al., 2003) *Foram-AMBI and *g-Foram-AMBI (Parent, 2019) and *g-exp(H'bc) (present paper).*

Supplementary materials are accessible online under:

S1. Polymerase Chain Reactions program and reagents.

Table S1. PTEs concentrations (ppm), references values: Effect Range Median (ERM); Effect Range Low (ERL), background (ppm), Pollution Load Index (PLI).

Table S2. Categories of different geochemical indices: Enrichment factor (*EF*) after Müller (1979), Contamination index (mC_d) after mC_d Hakanson (1980) and Pollution Load Index (*PLI*).

Table S3. Raw data of C/N, TN (%), TOC (%) and mud (%).

Table S4. Basic statistic of grain-size (%), organic matter content (%) and PTEs concentrations (ppm).

Table S5. Benthic foraminiferal morphological data.

Table S6. Benthic foraminiferal molecular data.

Table S7. Assigned MOTUs and Ecological Group.

Table S8. Unassigned MOTUs and Ecological Group.

Table S9. Abbreviation of morphospecies and MOTUs as reported in Figure 7.4.

Figure S1. Q-mode cluster analysis based on the selected environmental parameters.

Chapter 8: Encapsulated in sediments: eDNA deciphers the ecosystem history of the one of the most polluted European marine sites

Barrenechea Angeles, I., Romero-Martínez, M.L., Cavaliere, M., Varrella, S., Francescangeli, F., Piredda, R., Mazzocchi, M.G., Montresor, M., Schirone, A., Delbono, I., Margiotta, F., Corinaldesi, C., Chiavarini, S., Montereali M.R., Rimauro, J., Parrella, L. Musco, L., Dell'Anno, A., Tangherlini, M.[3], Pawlowski, J., and Frontalini, F.

8.1. Project Description

This project is part of an environmental restoration plan in the Bay of Bagnoli-Coroli. A core sample was taken in front of the former industrial site. The core analyzed here records the last 180 years, i.e., before industrialization up to the present day. Each team processed the core with its own expertise, ranging from geochemistry to metabarcoding of various groups such as prokaryotes, eukaryotes, metazoans, and foraminifera. For this project, I performed laboratory work and analysis of molecular data as well as some statistical analysis. Our data allowed to reconstruct the variations of the biodiversity under anthropogenic stress.

A version of this chapter is in revision in *Environment International*

8.2. Abstract

The Anthropocene is characterized by dramatic ecosystem changes in the environment driven by human activities. The impact of these activities can be assessed by different geochemical and paleontological proxies. However, each of these proxies provides only a fragmentary insight into the effects of anthropogenic impacts. It is highly challenging to reconstruct, with a holistic view, the state of the ecosystems from the preindustrial period to the present day, encompassing all biological components, from prokaryotes to multicellular eukaryotes. Here, we used sedimentary ancient DNA (*sedaDNA*) archives encompassing all trophic levels of biodiversity to reconstruct the two century-natural history in Bagnoli-Coroglio (Gulf of Pozzuoli, Tyrrhenian Sea) one of the most polluted marine-coastal sites in Europe. The site was characterized by seagrass meadows and high eukaryotic diversity until the beginning of the 20th century. Then, the ecosystem changed totally, with seagrasses and associated fauna as well as diverse groups of planktonic and benthic protists being replaced by low diversity biota dominated by dinophyceans and infaunal metazoan species. The *sedaDNA* analysis revealed a five-phase evolution of the area, where changes appear as the result of a multi-level cascade effect of impacts associated with industrial activities, urbanization, water circulation and land-use changes. The *sedaDNA* allowed to infer reference conditions that must be considered when restoration actions are implemented.

8.3. Introduction

The increasing human impacts and the concurrent threat on the environmental resources at both global (i.e., climate change) and local (e.g., pollution, overexploitation) scales have led to defining the current Epoch as the Anthropocene (Lewis and Maslin 2015). The deterioration of environmental conditions is widespread in marine environments, particularly in coastal ecosystems that represent the main global economic assets in terms of ecological services (Costanza et al. 1997; Keyes et al. 2021). The impact on coastal marine ecosystems has pervasive consequences by leading to species extinction (Barnosky et al. 2011; Dirzo et al. 2014), impairing the ecological functions (Obura et al. 2021), and ecosystem goods (Worm et al. 2006). In the light of it, the monitoring of marine biodiversity changes at different spatio-temporal scales is becoming a priority tool for planning conservation strategies at the international level (Dornelas et al. 2014; Danovaro et al. 2020b).

The emerging omics technologies (e.g., metagenomics) have recently offered a new perspective for understanding the impact of anthropogenic activities on biodiversity (Carraro et al. 2020). The environmental DNA (eDNA) has revolutionized the detection of species and found applications in biodiversity assessment and biomonitoring (Kelly et al. 2014; Cordier et al. 2018). Specifically, the application of eDNA metabarcoding improves the capability to simultaneously capture various biological components encompassing multiple trophic levels (Djurhuus et al. 2020). Indeed, the analysis of ancient eDNA preserved in sediment (*sedaDNA*) enables the reconstruction of the present-day to past biological communities and paleo-environmental conditions (Corinaldesi et al. 2011; Monchamp et al. 2017; Keck et al. 2020; Fordham et al. 2020; Siano et al. 2021). This allows for the comprehension of the effects of human activities on biodiversity over time and the definition of reference conditions (e.g., baseline or pre-industrial) (Jonkers et al. 2019). The emerging opportunity of *sedaDNA* has, however, scarcely applied to recent marine sedimentary records and has been mainly focused on a single marker that altogether prevents a holistic comprehension of community shifts or diversity changes and hampers the observation of their potential interactions.

In this work, we investigated changes in biodiversity and community composition spanning from prokaryotes to multicellular organisms, through a metabarcoding approach based on multi-markers, in relation to different historical anthropogenic impacts that occurred in one of the most polluted European (Sprovieri et al. 2020), likely worldwide, marine sites of Bagnoli-Coroglio (Gulf of Pozzuoli, Tyrrhenian Sea, Italy; Supplementary Fig.1). The bay and

hinterland area of Bagnoli-Coroglio have a long-lasting legacy of environmental impacts emerging from a complex interplay of anthropogenic activities (i.e., urbanization, industry, polluted-sediment disposal, land-use, coastline, and water-circulation changes) (Bertocci et al. 2019). The first industrial activities started to develop in the first decades of 1900 and reached their production peak in the late 1960s-early 1970s (Cavaliere et al. 2021). The site hosted the second-largest Italian steel factory (Ilva/Italsider), active till the early '90s (Trifuoggi et al., 2017), as well as asbestos (Eternit) and concrete (Cementir) factories. Due to the concentrations and hazardousness of pollutants, and their impacts on the ecosystem and human health, the area was declared a contaminated Site of National Interest (SIN) in 2000 (Law 388/2000). Today, Bagnoli-Coroglio is a large brownfield area, with marine sediments and benthic assemblages still intensively impacted by metals and polycyclic aromatic hydrocarbons (PAHs) (Gambi et al. 2020; Ausili et al. 2020; Tangherlini et al. 2020).

Here, we provide evidence that human activities have profoundly altered marine biodiversity and community composition, triggering a cascade effect on the ecosystem and its functioning during the last two centuries. We unveil the strong interconnection among multiple functional groups and their changes driven by the impacts of human-related activities. Our *sedaDNA* multimarker data reveal consistently a five-phase variation of the paleo-community that perfectly matches with the anthropogenically-induced environmental changes as supported by the historical archives and geochronology.

8.4. Methods

8.4.1. Core collection and processing.

The AB01 sediment core (110 cm in length) was collected at 55 m water depth in the Bay of Bagnoli (Gulf of Pozzuoli, Tyrrhenian Sea; 40° 48.150' N, 14° 08.913' E) (Supplementary Fig. 8.1) on December 5th, 2018, with an SW-104 corer Carma®, equipped with a liner of 10.4 cm in diameter. After collection, the core was transported to the laboratory, sliced through extruder equipment at 1 cm resolution discarding the outermost centimeter in contact with the liner, to avoid smearing effects downcore. Sediment layers were sliced using sterile spatulas, photographed, weighed and the color was scored following Munsell Soil Chart. For the first 41 cm and starting from the top layer (0-1 cm= layer #1), odd layers were used for grain size and chemical analyses, while even layers were used for biological analyses and organic matter characterization. Samples for DNA extraction for metabarcoding of eukaryotes and prokaryotes

were collected by cutting the central portion of each layer with a disposable, sterile petri dish (diameter 5 cm) and placed in 5 ml cryovials (2 for eukaryotes, 2 for prokaryotes), frozen in liquid nitrogen and stored at -80°C.

8.4.2. Grain size analyses.

Each core sediment sub-sample (odd layers) was washed with deionized water, then disaggregated using 1% sodium metaphosphate, and analyzed by a laser particle size analyzer (Helos KF SympaTEC). Grain size results from the top core to 40-41 cm sediment layer are reported as a volume percentage of sand, silt, and clay against the correspondent age (Fig. S8.6), accordingly to the sediment core dating described below.

8.4.3. Radiometric analyses and geochronology.

Gamma spectrometry analyses were performed on 20 g geometry of slightly ground sediment samples to determine ^{210}Pb and ^{226}Ra activities. Calibration, quality checks, and measurement procedures have been described elsewhere (Delbono et al. 2016). The excess ^{210}Pb ($^{210}Pb_{ex}$) activity is calculated as the difference between the total ^{210}Pb and the fraction in equilibrium with the parent radionuclide ^{226}Ra. Mass depth (g cm^{-2}) was used to account for the compaction of the sediment layers. The Constant Rate of Supply (CRS) model (Appleby and Oldfield 1978) was applied to determine the age of each layer of the sediment core, assuming a CRS of unsupported ^{210}Pb. The ^{210}Pb dating was validated and extended over several centuries using information taken from ^{226}Ra activities. In fact, the ejected sediments by Vesuvius volcanic activities are marked by a high ^{226}Ra concentration (Voltaggio et al. 2004), and its recent historical activity (1631-1944) is well-known and continuous (Scandone et al. 2008) giving a series of well-recognized time-markers: in the AB01 sediment core, ^{226}Ra peaks due to Vesuvius eruptions in 1944, 1906 and 1822 were identified (Armiento et al. 2022).

8.4.4. Chemical analyses.

Sediment samples were transferred to the laboratory in PEHD containers at a temperature of 4°C, where the phases of pre-treatment and analysis for metals and polycyclic aromatic hydrocarbons (PAHs) were performed. A detailed description of analytical procedures is reported in (Armiento et al. 2020). Briefly, after extraction by microwave-assisted acid digestion of the samples, metals (Cd, Cu, Cr, Hg, Pb, and Zn) analytical determinations were carried out, according to the EPA 6020b method, by inductively coupled plasma-mass

spectrometry (ICP-MS, Agilent 7800) except for Hg that was analyzed by atomic absorption spectrometer (AMA-254) according to EPA 7473 method. PAHs were extracted by an Accelerated Solvent Extractor (ASE 200 Dionex) and the analytical determination was carried out according to EPA 8270D method with the Agilent 7890A-5975C GC-MS system.

8.4.5. Organic matter

Sediment samples were dried in an oven (60°C) and subsequently ground and homogenized. Sub-samples were then weighed (~10–20 mg) in silver capsules and treated with increasing HCl concentration solutions (v/v 8%, 18%, and 25%) to remove the inorganic carbon fraction (carbonates) completely. Total organic carbon (TOC) and total nitrogen (TN) were analyzed using a Thermo Electron Flash Elemental Analyzer (EA 1112).

8.4.6. Prokaryotic metabarcoding

Total DNA was extracted in two replicates from 1 g of selected sediment layer (i.e., 1-2, 5-6, 9-10, 13-14, 17-18, 25-26, 27-28, 53-54) by using the DNeasy PowerSoil Kit (Qiagen) following manufacturer's instructions. The amount of DNA isolated was estimated by the absorbance at 260 nm and the purity by 260/280 and 260/230 nm ratios, by a NanoDrop spectrophotometer (ND-1000 UV-Vis Spectrophotometer; NanoDrop Technologies, Wilmington, DE, USA). Libraries were prepared from each replicate and sequencing was performed with all libraries on a single Illumina MiSeq flow cell by LGC Genomics GmbH (Berlin, Germany) using the 16S rRNA primers targeting V4-V5 regions (Parada et al. 2016). The raw reads were filtered low-quality base calls (<30 Phred score) using the QIIME II (Bolyen et al. 2019). Primer sequences used in the first PCR were discarded using Cutadapt (Martin 2011). Sequence reads were error-corrected, merged and amplicon sequence variants (ASVs) were identified using DADA2 (Callahan et al., 2016). ASVs were taxonomically classified using VSEARCH against the SILVA database (release 138) created trimming to the region amplified by the primers pairs used in the PCR step (Rognes et al. 2016) Taxonomic information was used to remove eukaryotic, chloroplast and mitochondrial-related sequences before subsequent analyses. To reduce analytical biases due to different sequencing depths among samples, the ASV table was randomly subsampled to the same number of sequences (13600). This normalized table was subsequently converted to a phyloseq class object in R to analyze and visualize the alpha and beta diversity (McMurdie and Holmes 2013). The raw data

is available from the Sequence Read Archive public database under the accession PRJNA822883.

8.4.7. Eukaryotes, foraminiferal and metazoan metabarcoding

For each sediment layer, 5 g of sediment was extracted using the DNeasy® PowerMax® Soil kit (QIAGEN©, Düsseldorf, Germany) following the manufacturer's instructions. All eDNA extractions and manipulations were performed to avoid cross-contamination between samples or modern eDNA. Three genetic markers, commonly used in biodiversity surveys (Frontalini et al. 2020b; Lanzén et al. 2021; Pawlowski et al. 2021a), were used to obtain a wide range of marine organisms. Eukaryotes, foraminifera, and metazoan were targeted by amplifying the V9 region of 18S rRNA with 1389F - 1510R primers (Amaral-Zettler et al. 2009), 37F region of 18S rRNA with s14F1- s15 primers (Pawlowski and Lecroq 2010), and mitochondrial cytochrome oxidase 1 gene with modified COI primers (Leray et al. 2013),respectively. Eight nucleotides were attached to the 5' extremity of primers to allow the multiplexing of samples (Esling et al., 2015). Primer sequences and PCR programs are detailed in Supplementary Table 5. Per sediment sample, three replicates and one control were performed and were checked on 1.5% agarose gels. The triplicates were merged, and the amount of DNA was quantified using high-resolution capillary electrophoresis with the QIAxcel system (QIAGEN©, Düsseldorf, Germany). The PCR products were equimolarly pooled into one 1.5 ml tube per marker and were purified using the High Pure PCR Product Purification kit (Roche Molecular Systems©, Basel, Switzerland). The final DNA concentration of pools was quantified using QubitTM HS dsDNA (Invitrogen, Thermo Fisher Scientific, Massachusetts, U.S.) kit. The libraries were prepared using the TruSeq® DNA PCR-Free Library Preparation Kit (Illumina©, California, U.S.) and then quantified by qPCR using the Kapa library quantification kit (Kapa Biosystems, Massachusetts, U.S.) for Illumina platform. Paired-end sequencing was performed on a MiSeq instrument (Illumina©, California, U.S.). We used the v2 kit with 300 cycles for V9 and foraminifera amplicons and the v3 with 500 cycles for COI amplicons. The raw data (i.e., forams, COI and V9) are available from the Sequence Read Archive public database under PRJNA824838. The bioinformatic analysis was performed using SLIM software (Dufresne et al. 2019). The raw sequencing data were demultiplexed with the module DTD. We performed DADA2 (Callahan et al., 2016) with the pseudo-pool option to filter the low-quality sequences, trim the primers, merge the paired-end reads and remove the chimeras. This dada2 module produced ASV sequences and ASV count tables. We applied a LULU (Frøslev et al., 2017)

curation on all markers datasets to filter PCR and sequencing errors or intra-individual variability as recommended in Brandt et al. (2021). Taxonomic assignments for the representative ASV-LULU V9 sequences were performed using standalone blast in the blast + suite (Altschul et al., 1990; Camacho et al., 2009) against PR2 database version 4.14.0 (Guillou et al., 2013) integrated with some private references sequences from diatoms strains. Assignments with a similarity of <90% and query coverage >90 bp were removed and assignments at the species level were manually checked. Taxonomic assignments for ASV-LULU foraminifera sequences were done using Vsearch (Rognes et al., 2016) at 95% similarity against our local database. The ASV-LULU COI sequences were taxonomically assigned using blast+ against the NCBI database with 95% similarity and a query coverage of >230 bp.

8.4.8. Benthic foraminifera

Benthic foraminifera were analyzed from 20 samples in the sediment fraction > 125 µm, where possible at least 300 specimens were picked. The taxonomical identification followed Cimerman and Langer (1991), Sgarella and Moncharmont Zei (1993) and Milker and Schmiedl (2012).

8.4.9. Foram-AMBI

For evaluating the Pale-EcoQS, the Foram-AMBI (Bouchet et al., 2012), an adaptation of the macrofauna AMBI (Borja et al., 2000) to benthic foraminifera, was calculated.

8.4.10. Metazoan AMBI

The AMBI values were calculated using AMBI Index Software v6.0 (Borja et al. 2012 ; http://ambi.azti.es; Species List v. Dec2020) with the following reference conditions: the highest observed diversity and richness and lowest AMBI values across all the samples.

8.4.11. Statistics

For each group, the species relative abundances were grouped using a constrained hierarchical clustering analysis (HCA) along the core depth (i.e., years). A similarity tree was produced using the Euclidian distance. Coniss (Grimm 1987) was used as the clustering method. The analysis was performed using the package vegan (Oksanen et al. 2007) and rioja (Juggins 2020) in RStudio. Based on the prokaryotic taxa encountered in the different sediment layers, the microgAMBI was calculated following Aylagas et al. (2017). Putative indicator taxa of the

ecological quality were identified based on their relationships with the pollution index as the probability of toxicity m-ERM-q (Long et al. 2000; Kowalska et al. 2018) by linear model regression analysis (lm) performed in R (R Core Team, 2013).

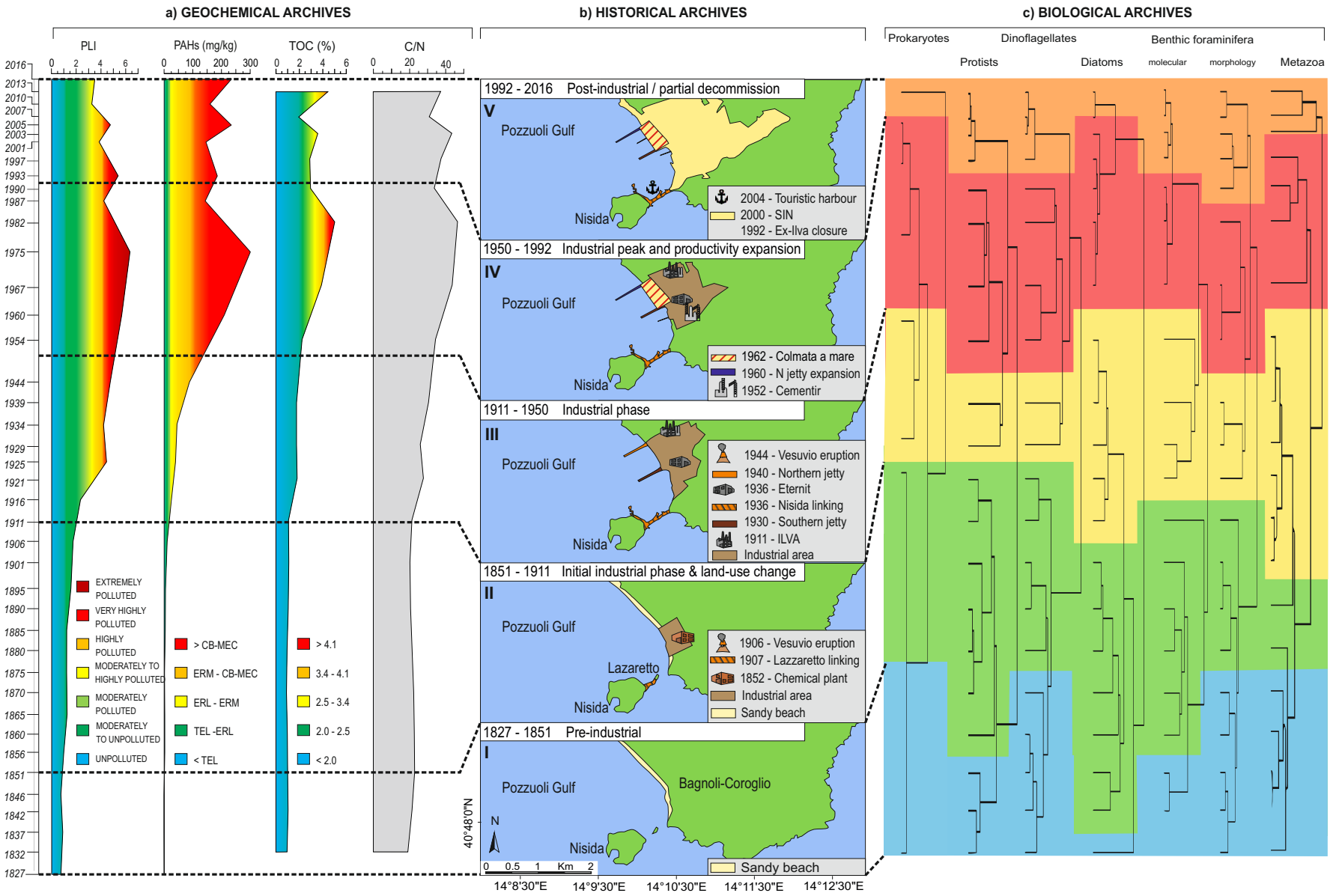
a) GEOCHEMICAL ARCHIVES
b) HISTORICAL ARCHIVES
c) BIOLOGICAL ARCHIVES
PLI
PAHs (mg/kg)
TOC (%)
C/N
Prokaryotes
Dinoflagellates
Benthic foraminifera
Protists
Diatoms
molecular
morphology
Metazoa
1992 - 2016 Post-industrial / partial decommission
V
Pozzuoli Gulf
Nisida
2004 - Touristic harbour
2000 - SIN
1992 - Ex-Ilva closure
1950 - 1992 Industrial peak and productivity expansion
IV
1962 - Colmata a mare
1960 - N jetty expansion
1952 - Cementir
1911 - 1950 Industrial phase
III
1944 - Vesuvio eruption
1940 - Northern jetty
1936 - Eternit
1936 - Nisida linking
1930 - Southern jetty
1911 - ILVA
Industrial area
1851 - 1911 Initial industrial phase & land-use change
II
Lazaretto
1906 - Vesuvio eruption
1907 - Lazzaretto linking
1852 - Chemical plant
Industrial area
Sandy beach
1827 - 1851 Pre-industrial
I
Bagnoli-Coroglio
EXTREMELY POLLUTED
VERY HIGHLY POLLUTED
HIGHLY POLLUTED
MODERATELY TO HIGHLY POLLUTED
MODERATELY POLLUTED
MODERATELY TO UNPOLLUTED
UNPOLLUTED
> CB-MEC
ERM - CB-MEC
ERL - ERM
TEL -ERL
< TEL
> 4.1
3.4 - 4.1
2.5 - 3.4
2.0 - 2.5
< 2.0
N
0 0.5 1 Km 2
40°48'0"N
14°8'30"E
14°9'30"E
14°10'30"E
14°11'30"E
14°12'30"E

Figure 8.1: ***Geochemical, historical, and biological archives.*** *Plot of selected geochemical proxies (PLI: Pollution Load Index; PAHs: Polycyclic Aromatic Hydrocarbons, TOC: Total Organic Carbon; C/N: carbon-nitrogen ratio). Classification of PLI after Zhang et al. (2011) with modification, sediment quality guidelines for PAHs (TEL: threshold effect level; ERL: effects range low; ERM: effect range median; CB-MEC: Consensus Based - midrange effect concentration) after Burton (2002) and TOC after Bouchet et al. (2018). Historical archives with significant human-induced changes and activities. Hierarchical Cluster Analyses (HCAs) based on prokaryotes, protists, dinoflagellates, diatoms, benthic foraminifera (molecular and morphological) and metazoa. The colour-shaded areas over HCAs denote the stepwise temporal phases along the AB01 core in the Bagnoli-Coroglio SIN (Gulf of Pozzuoli, Tyrrhenian Sea), namely I: Pre-industrial (1827-1851; blue), II: Initial industrial phase & land-use change (1851-1911; green), III: Industrial phase (1911-1950; yellow), IV: Industrial peak and productivity expansion (1950-1992; red) and V: Post-industrial/partial decommission (1992-2016; orange).*

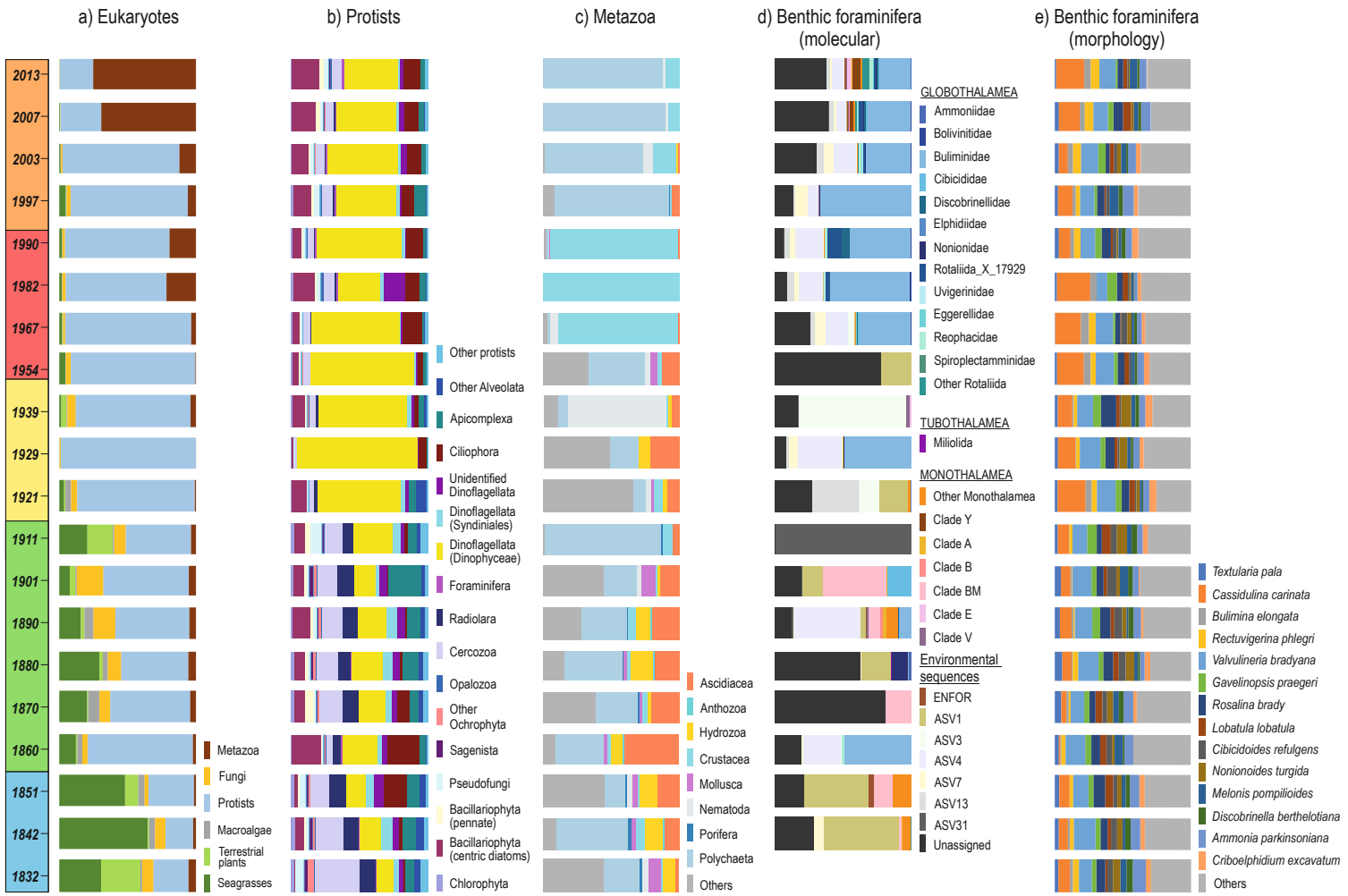

Figure 8.2:Paleo-community composition shift. Eukaryotes, protists, metazoa, and benthic foraminifera (molecular and morphological) composition changes along the AB01 core in the Bagnoli-Coroglio SIN (Gulf of Pozzuoli, Tyrrhenian Sea) inferred from multimarker sedaDNA data from 1827 to 2013. The different colours over the ages refer to the main phases (I: blue; II: green; III: yellow; IV: red and V: orange) as identified in Figure 1.

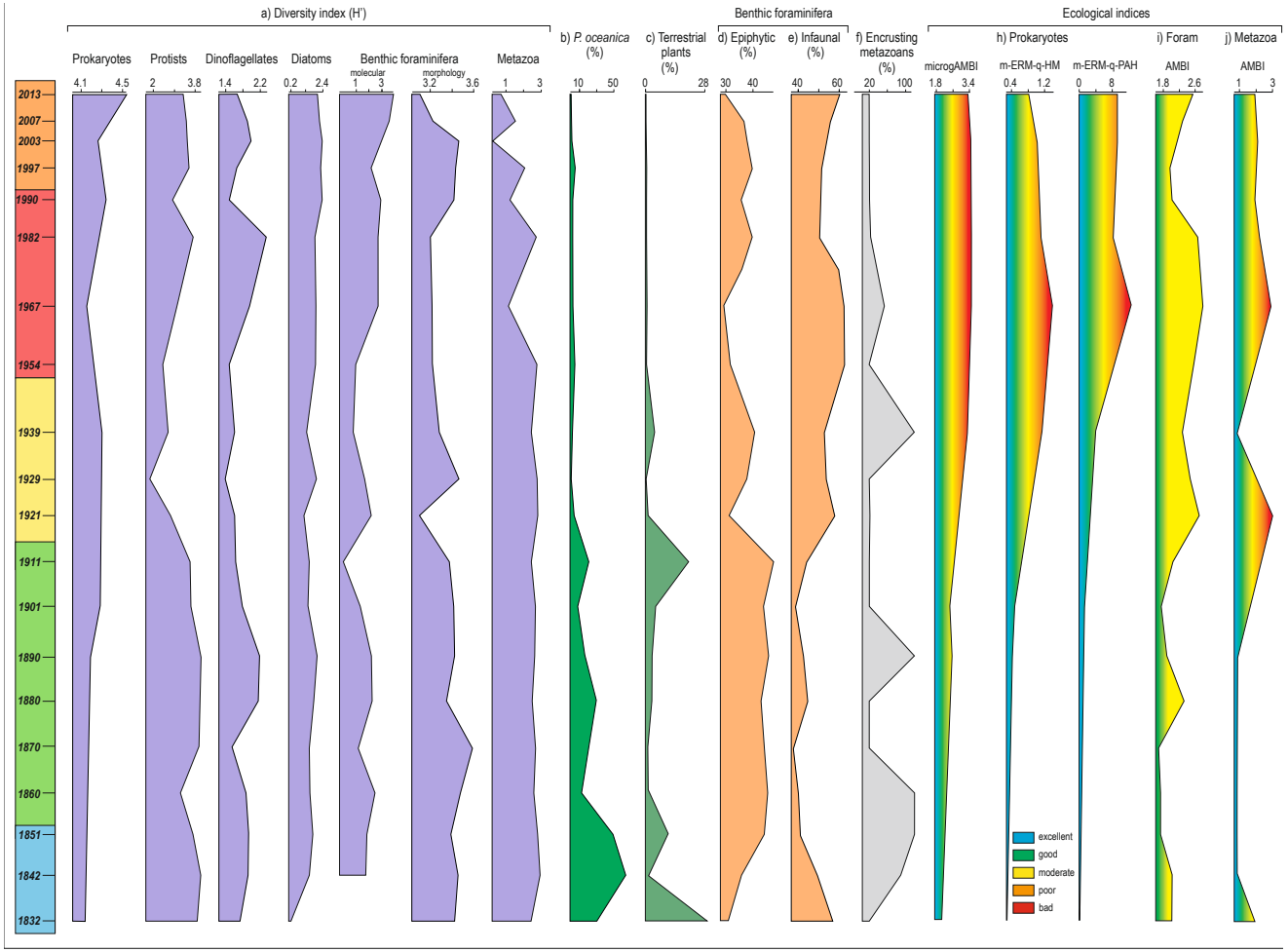

Figure 8.3:Changes in diversity and ecological indices. . Diversity (H') plots of prokaryotes, protists, dinoflagellates, diatoms, benthic foraminifera (molecular and morphological) and metazoa coupled with relative abundance of Posidonia oceanica, land plants, epifaunal and epiphytic foraminifera and encrusting metazoa percentages. Ecological indices for prokaryotes (microgAMBI, m-ERM-q-HM, and m-ERM-q-PAHs), foraminifera (Foram AMBI) and metazoa (AMBI) are also plotted. The different colour over the ages refer to the main phases (I: blue; II: green; III: yellow; IV: red and V: orange) as identified in Figure 1.

8.5. Results

8.5.1. The geochemical signature of the pollution history at the Bagnoli-Coroglio SIN

The analysis of the AB01 sediment core provides a high-resolution (4.6 yr cm^{-1}), continuous, complete, and well-constrained (i.e., ^{210}Pb and ^{226}Ra) record for the reconstruction of paleo-environmental changes in the Bagnoli-Coroglio SIN in the last ~190 years (Fig. 8.1 and Supplementary Table 1) (Armiento et al. submitted).

The geochemical profiles highlight a progressive increase in the concentrations of inorganic (i.e., trace elements as underlined by the Pollution Load Index: PLI) and organic (i.e., Polycyclic Aromatic Hydrocarbons: PAHs) pollutants as well as in the quantity (i.e., TOC) and quality (i.e., C/N) of the organic matter (OM) at the very beginning of the 20th century (i.e., 1901) and a marked peak in the period 1954-1992 (Fig. 8.1, Supplementary Tables 2-3). The PLI record reveals moderately polluted conditions since 1921 that turned from highly to extremely highly polluted in 1944 up to 1990. A similar pattern is observed in the total concentration of PAHs whose origin is mostly pyrogenic (i.e., Ant/Ant+Phe > 0.1, Fig. S8.2). The TOC trend shows a progressive increase since 1916, with a peak between 1954 and 1992. Based on the C/N values that are persistently >19, a terrigenous origin of the OM has been inferred (Meyers et al 1994). However, the significant increase of C/N in the 1960-1987 interval suggests a higher input of carbon of terrestrial origin, possibly resulting from the "*Colmata a Mare*" that corresponds to the disposal of contaminated soil between the two piers in 1962-1967 to enlarge the industrial area and/or from contamination by hydrocarbons (i.e., PAHs with a highly unbalanced C/N ratio).

8.5.2. Community composition shifts from past to present

The Hierarchical Clustering Analysis (HCAs) based on community changes in multimarker *sedaDNA* datasets and foraminiferal morphospecies consistently identified five main groups corresponding to different time intervals related to the development and exploitation of the area based on historical archives (Fig. 8.1). These stepwise temporal phases were defined as pre-industrial (1827-1851), initial industrial phase and land-use change (1851-1911), developing industrial phase (1911-1950), maximal industrial expansion (1950-1992), and post-industrial/partial decommission (1992-2016) (Fig. 8.1). The detailed analysis of community changes for different taxa along the vertical profile of the sediment is presented below.

The prokaryotic metabarcoding analysis showed that the Amplicon Sequence Variants (ASVs) belonging to *Bathyarchaeia* accounted on average for 13.5% of the prokaryotic assemblage, followed by *Anaerolineaceae* (on average, 7.9%), and *Aminicenantales* (on average 4.8%). However, the relative abundance of these main prokaryotic taxa changed in relation to the analyzed sediment layer (Fig. S8.3). In particular, *Bathyarchaeia* accounted for one-fifth of the assemblage in the deepest sediment layer (the oldest one), but its relative abundance progressively decreased towards the top of the sediment core (up to ca. 3% in the surface layer). At the same time, ASVs belonging to *Thermoanaerobaculaceae* and *Flavobacteriaceae* accounted for a very low fraction in the oldest sediment layers, while they were present to a major extent in 2013 (Fig. S8.3). ASVs related to *Desulfosarcinaceae* showed a hump shape pattern along with the sediment profile with the highest contribution in sediment layers dated from 1967 to 2003.

The eukaryotic metabarcoding data revealed clear temporal faunal turnover when considering the major groups of multi- and unicellular organisms (Fig. 8.2). The eukaryotic assemblage is dominated by seagrasses (*Posidonia oceanica*) in the oldest layers (1832-1851). A clear DNA signature of *P. oceanica* persisted until the beginning of the industrial phase (up to 1911). A similar temporal trend was also detected for terrestrial plants, whose signature almost disappeared after 1911 (Fig. 8.2). The plant DNA signatures were progressively replaced by those of microbial eukaryotes (protists) from 1860 to 2003 and metazoans from 2007 to 2013.

The dramatic shift in the paleo-community composition observed at the beginning of the industrial period (i.e., after 1911) was also detected within protists (Fig. 8.2). Before 1911, the protist assemblage was highly beta-diverse, represented by free-living dinoflagellates (Dinophyceae), parasitic dinoflagellates (Syndiniales), Apicomplexa, ciliates, diatoms, and Rhizaria (Radiolaria, Foraminifera, and Cercozoa). A remarkable decrease in the abundance of cercozoans and radiolarians (dominated by planktonic spumellarians and nassellarians), as well as Apicomplexa, was observed after 1911 (Fig. 8.2). From 1921, the protist community was dominated by Dinophyceae, mainly including genera known to produce resting stages (e.g., *Protoperidinium*, *Scrippsiella* and other taxa belonging to the family Thoracosphaeraceae, *Gymnodinium*) (Fig. S8.4). Among other groups, diatoms represented a rather constant percentage of the protist assemblage. They were markedly dominated by planktonic taxa that produce resting stages, among which the genus *Chaetoceros* was retrieved all along with the core, while *Biddulphia*, *Skeletonema* and *Thalassiosira* became abundant after the main

industrialization (i.e., after 1954) (Fig. S8.5). From 1967, an increasing abundance of ciliates was also detected.

A community composition shift after 1911 was also observed in benthic foraminifera analyzed with both molecular and morphological approaches (Fig. 8.2). Before 1911, the assemblages inferred by *SedaDNA* were dominated by monothalamids and representatives of taxonomically non-identified environmental clades. In 1911, a small number of reads corresponding almost to only one ASV was identified. Since 1967, buliminids and other calcareous globothalaminids were continuously present. In the morphological dataset, the most abundant species were *Valvulineria bradyana*, *Cassidulina carinata*, *Rosalina bradyi*, *Ammonia parkinsoniana, Gavelinopsis praegeri,* and *Lobatula lobatula*. The abundance of *C. carinata* showed a marked increase between 1911 and 1921 (from 8 to 20%) that was associated with a lower abundance of *A. parkinsoniana* (1880-1997), *Textularia pala* (1911-1990), *L. lobatula* (1921-2001) and *Cibicides refulgens* (1901-2013), *Nonionella turgida* (1954-2013). *Bulimina elongata* exhibited a significant increase after 1954 (Fig. 8.2, Supplementary Table 4).

Similar changes were also observed in the metazoan community, although their timing was slightly different. Encrusting metazoans (ascidians, sponges) were constantly observed until 1911, with the only exception in 1939. After 1911, the encrusting and colonial ascidians were replaced by epibenthic and solitarian ascidiaceans that also disappeared after 1954 (Fig. 8.2). The sequences of mainly planktonic Hydrozoa decreased throughout the core and almost disappeared in 1939. The most recent layers (i.e., 1997-2013) were mainly dominated by burial polychaetes (Annelida), whereas Crustacea were dominant in the phase of heavy industrialization (1967-1990).

8.5.3. Changes in diversity and ecological indices

The faunal turnovers were accompanied by changes in the alpha-diversity and ecological indices (Fig. 8.3 and Supplementary Table 4). From 1832 to 1921, relatively higher diversity values were identified for protists and dinoflagellates, and morphospecies of foraminifera. A similar trend between 1954 and 1967 was observed for metazoan diversity, while the opposite was instead found for prokaryotes, diatoms, and foraminifera. *Posidonia oceanica*, coming from the nearby meadows, was consistently identified from 1832 to 1911 when it substantially disappeared along with terrestrial plants. This dramatic event was associated with a clear shift in foraminiferal assemblage dominated by epiphytic species and replaced by infaunal ones after 1911 (Fig. 8.2 and Supplementary Table 4). Encrusting metazoans follow a similar trend

showing higher percentages in the pre-industrial and partly in the early industrial phases, except in 1939.

The microgAMBI values increased progressively from the oldest sediment layer to the recent ones with the highest values (3.5) corresponding to the period from 1967 to 1990 and then slightly decreased. These values allowed us to classify ecological quality status (EcoQS) conditions from good/moderate in 1901 to mostly poor in 1967-1990 (Fig. 8.3 and Supplementary Table 4). A similar trend was found for the probability of toxicity (i.e., m-ERM-q) with a marked peak in 1967. Foram AMBI showed a clear increasing trend from 1911-1922 to 1982-1990 and a shift from good-moderate to moderate EcoQS. A similar trend was also found for gAMBI values based on macroinvertebrate molecular data with two relevant peaks pointing to bad EcoQS in 1921 and 1967. Overall, the gAMBI mostly classified the EcoQS as excellent up to 1901 and moderate-to-bad afterwards (Fig. 8.3 and Supplementary Table 4).

8.6. Discussion

The fluctuations of geochemical proxies, the *sedaDNA* records, and the alpha-diversity and ecological indices observed in the AB01 sediment core over two centuries reflect the ecosystem changes in the contamination history of the area resulting from anthropogenic activities. The HCAs multimarker *sedaDNA* records consistently reveal a five-phase evolution of the area that perfectly matches with the anthropogenically induced environmental changes as supported by historical archives and other sources (Figs. 8.1 and 8.2). The environmental changes have not only resulted from contamination of the site but also from the geomorphological changes resulting from the expansion of the production site (i.e., reclaiming the area between the two piers "*Colmata a Mare")* and those that modified water circulation and sediment deposition (e.g., Lazzaretto and Nisida linking, piers' construction) (Fig. 8.1). Taken together, these anthropogenic activities led to a concurrent modification of the biological communities from prokaryotes and unicellular eukaryotes to higher plants and metazoans, impacting their diversity and ecological indices.

8.6.1. Phase I (1827-1851)

Until the Industrial Revolution, the Bagnoli-Coroglio plain was a marshy back-barrier as supported by geomorphological evidence and the presence of palustrine sediments (Ascione et al. 2021) (Fig. 8.1-b). Agriculture was the main activity and the whole area was characterized by a few residential constructions and some touristic resorts along the beach, where thermal

springs were just discovered. The lowest values of both organic and inorganic contaminants are recorded in this phase together with the highest beta-diversity values of different biological communities (i.e., protists, dinoflagellates, foraminifera, metazoa). All ecological indices (e.g., AMBIs and m-ERM-q) inferred excellent and good conditions. This finding indicates pristine environmental conditions, the ecosystem baseline with insignificant anthropogenic impacts. The most evident signature in our records is the very high relative abundance in this phase of higher plants and more importantly of the seagrass *P. oceanica* (Fig. 8.2).

8.6.2. Phase II (1851-1911).

From the mid-1800 to the beginning of the 20th century, a long sequence of anthropogenic interventions has deeply modified the natural landscape of the area. The initial projects for settling the industrial district at Bagnoli plain date back to 1852, when the glasswork Melchiorre - Bournique and the chemicals plant Lefevre went into operation (Cirillo et al. 2022) (Fig. 8.1-b). In 1907, the tuff reef of Lazaretto was linked to the Island of Nisida (Fig. 8.1-b). In 1905, the Ilva/Italsider company set up the construction of one of the largest Italian steel plants that started its production in 1911.

The environmental parameters and paleo-community records do not show any substantial change during phase II still reflecting good environmental conditions with minor variations imputable to early urbanization and land-use changes. These data match well the information deriving from the historical archives and the mapping of benthic community where *P. oceanica* meadows were indeed recorded between 10 and 30 m depth up to the first decade of the pre-industrial period (1910) in front of the Bagnoli-Coroglio industrial area (Porzio et al. 2020). However, the signature of *P. oceanica* decreased in this period paralleled by a reduction of terrestrial plants, whose signature almost disappeared after 1911 (Figs. 8.2 and 8.3). This was probably due to a change in land use of the area where orchards were progressively replaced by buildings for workers (ca. 2000 people in 1910 and 4000 in 1918 at Ilva/Italsider) in the expanding industrial area (Selvaggio, 2015). This phase, as compared to the subsequent ones, was characterized by higher abundances of parasitic taxa, mainly Syndiniales (parasites of dinoflagellates, radiolarians, ciliates) and Apicomplexa (parasites of metazoans) (Fig. 8.2), which may be related to the presence of specific hosts and further indicates a marked change of the community structure in the area.

8.6.3. Phase III (1911-1950)

This period marks the beginning of the industrial phase. In 1927, a factory for the production of blast furnace cement was built, and three years later, a pier was constructed for loading finished products and receiving coal and iron ores from vessels (Fig. 8.1-b). In 1936, the Island of Nisida was completely linked to the mainland by a 700-m-long strip of land that modified water circulation dynamics and the sedimentological pattern in the Gulf of Pozzuoli (Fig. 8.1-b). In 1936, a plant for the production of asbestos (Eternit) started its activity and in 1940 a northern jetty was built. The enlargement of the factory continued till 1943, when the Second World War interrupted plant operations up to 1946. In this period, a marked contamination started with an overall increase of mostly inorganic chemicals (i.e., PLI) pointing to moderately-to-highly polluted conditions. PAHs started to increase with concentrations higher than the Effect Range-Low (ER-L) and, at the end of the Phase III, even higher than Effect Range-Median (ER-M). The extreme concentrations of PAHs (up to 87,000 μg g^{-1} d.w.) and their predominantly pyrogenic origin were also reported along with short cores in the beaches surrounding the former industrial plant (Passaro et al. 2020).

These changes are accompanied by radical paleo-community shifts in *sedaDNA* archives, namely the disappearance of *P. oceanica*, and important turnovers in the communities of protists and metazoans. *Posidonia oceanica* is an ecosystem engineer species endemic to the Mediterranean Sea that forms dense meadows in areas characterized by clean and clear waters (Marbà et al. 2014). The disappearance of this habitat-forming species was accompanied by a drastic decrease of epiphytic (i.e., plant-dwelling) foraminiferal taxa as well as of encrusting animals such as ascidians (*Botryllius* sp.) and sponges (Fig. 8.2). It can be reasonably hypothesized that the disappearance of *P. oceanica* was related to the construction of the connection between Lazaretto and the mainland in 1906 that altered the hydrodynamics of the area markedly and led to the deposition of finer sediments (silt and clay) leading to higher water turbidity and ultimately to reduced photosynthetic activity. However, grain-size data of AB01 sediment core show only minor variations in the sediment fine fractions (silt and clay) after 1906 (Fig. S8.6). According to Porzio et al. (2020), the massive hydrocarbons (PAHs) released in this area could also contribute to their disappearance. The changes in water circulation could also explain a remarkable decrease of planktonic radiolarians, while the modification of the seabed could have led to the replacement of encrusting and colonial ascidian by epibenthic and solitarian ascidiaceans after 1911. In phase III, considerable changes are also observed in the protist community, with a decrease in diversity (Fig. 8.3) and a shift in its composition. Lower

relative abundances of parasitic dinoflagellates (Syndiniales), Apicomplexa, Radiolaria, and Cercozoa were compensated by a marked dominance of Dinophyceae, mainly including taxa known to produce resting cysts (Figs. 8.2, 8.3 and Fig. S8.4).

Many of these changes reflect a worsening of the environmental conditions triggered by the beginning of industrial activities. Overall, a reduction in diversity of eukaryotes, protists and foraminifera is associated with the beginning of the industrial activities in 1911 and becomes particularly evident during the maximum expansion of the industrial phase (i.e., 1920-1980) (Fig. 8.3). The increase in Foram-AMBI and microgAMBI, proxies for estimating the EcoQS, is associated with the increase of OM (i.e., TOC and C/N) and reliably follows its patterns. Indeed, this shift in the quantity (i.e., TOC) and possibly the quality (i.e., C/N) of OM strongly matches with the decrease of epiphytic foraminiferal species that are affected by the enhanced OM availability (i.e., higher TOC) with a more refractory origin (i.e., higher C/N). On the other hand, the concurrent increase of infaunal (i.e., organisms living within the sediment) taxa perfectly reflects the enhanced OM availability at the seafloor.

8.6.4. Phase IV (1950-1992)

In 1952, the Cementir started its production of concrete and the '60s marked the production peak of steel reaching up to $2x10^6$ tons yr^{-1} and involved ca. 5400 workers at Ilva/Italsider. This period defines the peak of the industrial phase and is mainly reflected by the highest level of contamination with highly-to-extremely highly polluted conditions based on PLI (Fig. 8.1-a). Between 1962 and 1964, a large amount of contaminated soil from the industrial area was used to enlarge the industrial area by filling the gap between the two piers and creating the so-called "*Colmata a mare*". In the '70s with 7698 workers at Ilva/Italsider, the steel production started to decrease, in 1985 the Eternit ceased its asbestos production and the steel factory permanently closed in 1992 (Fig. 8.1-b). The highest level of contamination (i.e., PLI and PAHs) was recorded between 1960 and 1987. The ecological indices associated with prokaryotes, namely m-ERM-q-HM and -PAHs, clearly evidence an increasing trend and a marked peak in 1967 (Fig. 8.3). The Phase IV is also characterized by the worst EcoQS equally indicated by AMBI of prokaryotes, foraminifera, and metazoans, with a marked peak in 1967, in response to the enhanced OM availability. Despite the increase in C/N ratio suggesting a terrestrial origin of OM, the absence of higher plants record based on *sedaDNA* suggests an overall decrease over time of terrestrial inputs that implies a strong contribution of PAHs to the increase of C/N values (Meyers, 1994; Rumolo et al., 2011). In Phase IV, the protist community is still characterized

by the marked dominance of cyst forming dinoflagellates as in Phase III (Fig. 8.2). The remarkable abundance of these free-living heterotrophic and mixotrophic dinoflagellates can reflect a change in the trophic structure of the community. Alternatively, it could indicate a specific response of these unicellular organisms to the high concentration of pollutants in the sediments and the overlying water column. The transition between free-living motile stages to non-motile resting cysts occurs in the deeper layers of the water column, close to the sediments (Brosnahan et al. 2017) and it has been shown that high concentrations of heavy metals and PAHs are related to an increase of cyst production (Horner et al. 2011; Triki et al. 2017). Phase IV was characterized by a marked reduction of metazoan diversity starting from 1954 (Fig. 8.3). The delayed response of metazoans compared to planktonic protists might be associated with the higher sensitivity behaviour of the latter to water chemical and physical variations that allow them to respond to changes rapidly. It might be speculated that the time lag response of metazoans is also linked to their higher position in the marine trophic web.

8.6.5. Phase V (1992-2016)

After the dismantling of the site started between 1994 and 2000 the Bagnoli-Coroglio area was included in the SIN list by the Italian Government, highlighting the threat to human health and the need to apply a remediation program (Ausili et al. 2020; Morroni et al. 2020).In the early 2000s, the first remediation projects were planned but the complete remediation has not been completed (Fig. 8.1-b). This post-industrial phase shows an overall reduction of contamination, though PLI and PAHs still remain high, and a concurrent increase of diversity indices values. Metazoans dominated in the most recent layers (2013 and 2007) and were mainly represented by polychaetes that are currently abundant in the sandy bottom of the Bay of Bagnoli-Coroglio (Fasciglione et al. 2016; Morroni et al. 2020).

8.7. Conclusion

The present study provides the first holistic insight into the ecosystem changes that occurred in one of the most polluted marine coastal areas during the last two centuries. In the absence of long-term ecological studies of coastal communities, paleogenomics offers a unique opportunity to reconstruct past changes in biodiversity by linking them to environmental change and increased anthropogenic pressure. The investigation of *sedaDNA* allows us to retrospectively identify a five-stage stepwise evolution in the prokaryotic and eukaryotic communities, demonstrating how the industrial pollution and land-use changes have deeply

modified their natural equilibrium. The *sedaDNA* approach also enables to disentangle the multi-level cascade effects of industrialization, bringing to light the pre-industrial reference conditions and showing that after 20 years of decommissioning, the good ecological conditions have not fully recovered yet. As shown by our study, paleogenomics serves not only as a tool for recording past ecosystems but might also help in assessing the recovery of degraded coastal marine ecosystems.

Supplementary materials

Layer	Year	u(Year)
cm	a	a
0-1	2016	1
1-2	2013	1
2-3	2010	1
3-4	2007	1
4-5	2005	1
5-6	2003	1
6-7	2001	1
7-8	1997	2
8-9	1993	2
9-10	1990	2
10-11	1987	2
11-12	1982	2
12-13	1975	3
13-14	1967	4
14-15	1960	4
15-16	1954	5
16-17	1944	3
17-18	1939	3
18-19	1934	3
19-20	1929	3
20-21	1925	3
21-22	1921	3

Layer	Year	u(Year)
cm	a	a
19-20	1929	3
20-21	1925	3
21-22	1921	3
22-23	1916	3
23-24	1911	3
24-25	1906	3
25-26	1901	3
26-27	1895	3
27-28	1890	3
28-29	1885	3
29-30	1880	4
30-31	1875	4
31-32	1870	4
32-33	1865	4
33-34	1860	4
34-35	1856	4
35-36	1851	4
36-37	1846	4
37-38	1842	3
38-39	1837	3
39-40	1832	3
40-41	1827	3

Table 8 S.1 : AB01 sediment core chronology. For every layer, ages and age uncertainties are presented.

Age	PLI	PAHs
2016	3.47	235.96
2010	3.27	160.62
2005	4.76	235.31
2001	3.83	146.77
1993	5.32	186.61
1987	4.18	144.01
1975	6.35	301.61
1960	5.61	210.94

1944	4.74	89.29
1934	4.17	45.78
1925	4.44	40.89
1916	2.38	26.40
1906	1.73	11.33
1895	1.57	5.90
1885	1.22	4.86
1875	1.21	2.29
1865	1.27	3.23
1856	1.03	3.32
1846	0.79	0.58
1837	0.91	0.79
1827	0.79	0.93

PLI class	
0-1	unppolluted
1-2	moderately to unpolluted
2-3	moderately polluted
3-4	moderately to highly polluted
4-5	highly polluted
5-6	very highly polluted
6-7	extremely polluted

PAHs limits and threshold		**mg/kg**	<TEL
SQGs	Threshold Effect Limit (TEL)	0.87	TEL-ERL
	Effects Range Low (ERL)	3.5	ERL-ERM
	Effects Range Median (ERM)	23.58	ERM-CB
	Consensus Based-Midrange Effect Concentration (CB_ME)	100	>CB

	Cd	Cr	Cu	Hg	Pb	Zn

SITE SPECIFIC BACKGROUND by Armiento et al. 2022	ppm	ppm	ppm	ppm	ppm	ppm
	0.29	23	26.9	0.45	85	80

Table 8 S.2: AB01 sediment core a) Pollution Load Index (PLI) values and classes; b) Polycyclic Aromatic Hydrocarbons (PAHs) and thresholds; c) site specific background concentration of trace elements.

Age	TN	TOC	C/N
	%	%	
2013	0.12	4.41	37.2
2007	0.07	2.09	30.9
2003	0.08	3.61	43.4
1997	0.08	2.97	37.4
1990	0.09	3.02	33.7
1982	0.11	4.95	46.5
1967	0.09	3.89	43.6
1954	0.07	2.35	34.4
1939	0.06	1.92	30.4
1929	0.07	1.91	25.9
1921	0.07	1.96	27.6
1911	0.06	1.27	21.2
1901	0.06	1.30	20.1
1890	0.06	1.25	20.5
1880	0.05	1.16	21.3
1870	0.05	1.12	22.2
1860	0.05	1.19	22.5
1851	0.05	1.21	22.7
1842	0.06	1.24	21.3
1832	0.06	1.16	19.0

TOC
%
>4.1
3.4-4.1
2.5-3.4
2.0-2.5
<2.0

Table 8 S.3: AB01 sediment core Total Nitrogen (TN), Total Organic Carbon (TOC) and C/N values.

Age	Foraminifera	Terrestrial plants	*Posidonia oceanica*	Metazoa Mean AMBI

	Foram-AMBI	Infaunal %	Epiphytic%			
2013	2.5	60.1	30.28	0.00	0.06	1.5
2007	2.3	55.5	38.33	0.01	0.52	n.a.
2003	2.1	50.3	40.22	0.11	0.99	1.7
1997	2.0	51.3	36.18	0.31	4.81	n.a.
1990	2.0	50.3	40.22	0.24	1.98	1.5
1982	2.6	59.5	36.36	0.25	1.90	1.9
1967	2.7	62.2	29.67	0.45	2.18	2.9
1954	2.5	62.4	32.07	0.24	4.44	n.a.
1939	2.2	52.6	41.11	3.75	1.55	0.0
1929	2.4	53.5	38.16	0.06	0.19	n.a.
1921	2.6	57.6	31.60	0.92	3.59	3.0
1911	2.0	43.9	48.13	18.61	20.57	n.a.
1901	1.8	38.5	44.39	4.19	7.70	n.a.
1890	1.9	42.4	46.29	2.66	15.82	0.0
1880	2.3	44.4	43.52	2.56	29.34	n.a.
1870	1.7	37.4	44.68	0.79	20.45	n.a.
1860	1.8	39.9	45.96	0.96	12.17	0.0
1851	1.8	40.9	44.68	9.64	48.14	0.0
1842	2.0	49.2	36.27	1.02	64.68	0.0
1832	2.0	56.7	31.34	28.74	30.61	1.5

Age	m-ERM-Q-PAH	m-ERM-Q-HM	microgAMBI
2013	6.48	0.79	3.28
2013	6.48	0.79	3.38
2003	6.42	1.01	3.54
2003	6.42	1.01	3.40
1990	5.63	1.10	3.50
1990	5.63	1.10	3.50
1967	8.99	1.39	3.59
1967	8.99	1.39	3.40
1939	2.37	1.12	3.28
1939	2.37	1.12	3.32
1901	0.29	0.43	2.80
1901	0.29	0.43	2.11
1890	0.16	0.37	2.47
1890	0.16	0.37	2.65

Table 8 S.4: Diversity and ecological indices.

Genetic markers used for this study

Organism	region	primers name	forward sequence (5'-3')	reverse sequence (5'-3')	ref

Prokaryotes	V4-V5	515F-Y 926R	GTGYCAGCMGCCGCGGTAA	CCGYCAATTYMTTTRAGTTT	Parada et al., 2016
Eukaryotes	nc 18S V9	1389F - 1510R	TTGTACACACCGCCC	CCTTCYGCAGGTTCACCTAC	Amaral-Zettler, et al., 2009
Foraminifera	nc 18S 37F	s14F1 - s15R	AAGGGCACCACAAGAAGC	CCACCTATCACAYAATCATG	Pawlowski and Lecroq, 2010
Metazoan	mt COI	mlCOIintF - dgHCO2198	GGWACWGGWTGAACWGTWTAYCCYCC	TAAACTTCAGGGTGACCAAARAAYCA	Leray et al., 2013

Program PCR :

V9

94°C	3'	
94°	30"	30X
57°	60"	
72°	90"	
72°C	10"	

COI

95°C	5'	
94°C	10"	16 cycles
62-47°C	30"	
72°C	1'	
95°C	10"	35 cycles
46°C	30"	
72°C	1'	
72°C	2'	

Foraminifera

94°C	5'	
94°	30’	50 cycles
52°	30’	
72°	30’	
72°	2'	

Table 8 S.5: Genetic markers and PCR programs used for eukaryotes, metazoan and foraminifera

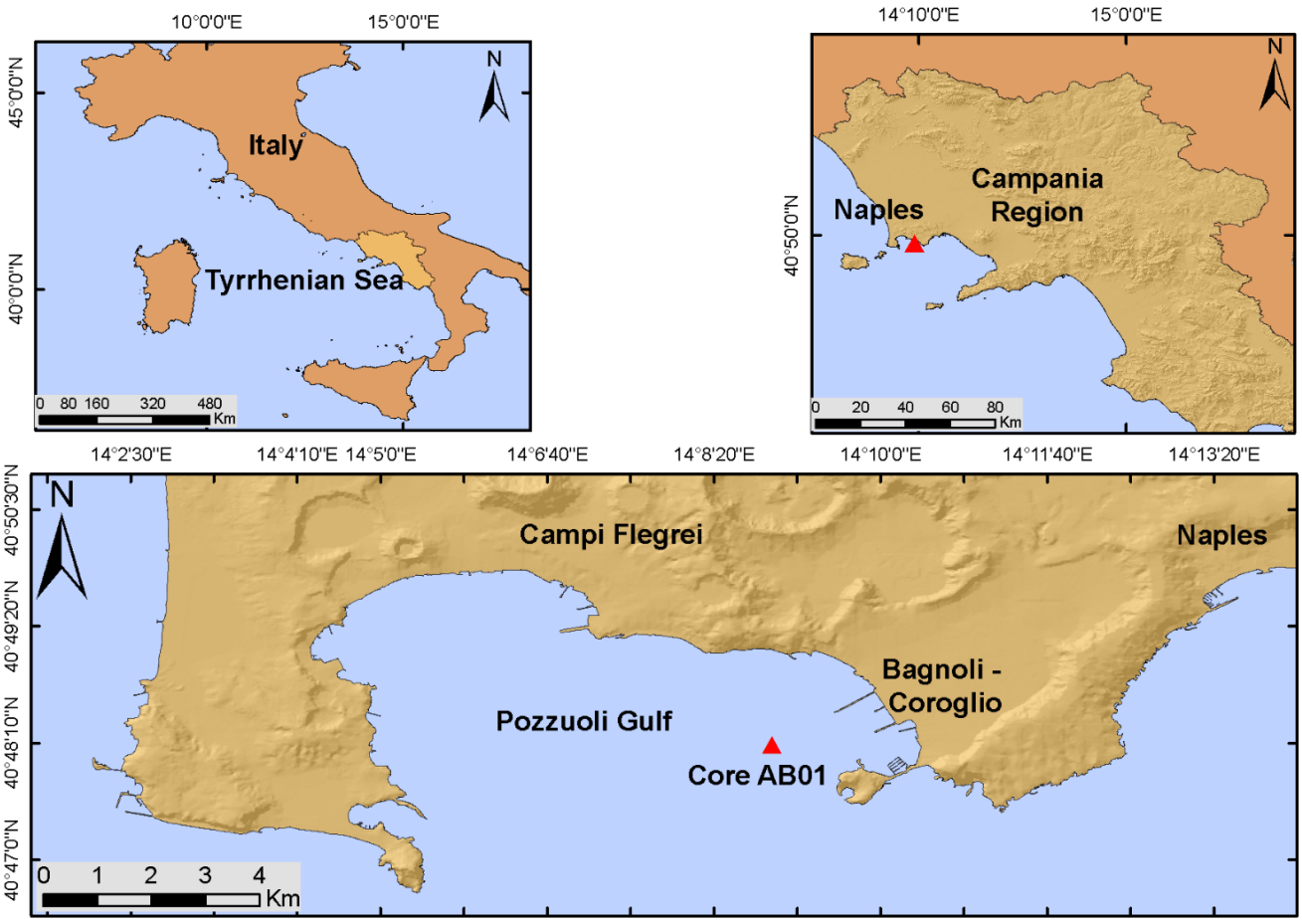

Figure S8. 1: Location of the Bagnoli-Coroglio SIN in the Pozzuoli Gulf (Tyrrhenian Sea) and position of the AB01 sediment core (55m water depth).

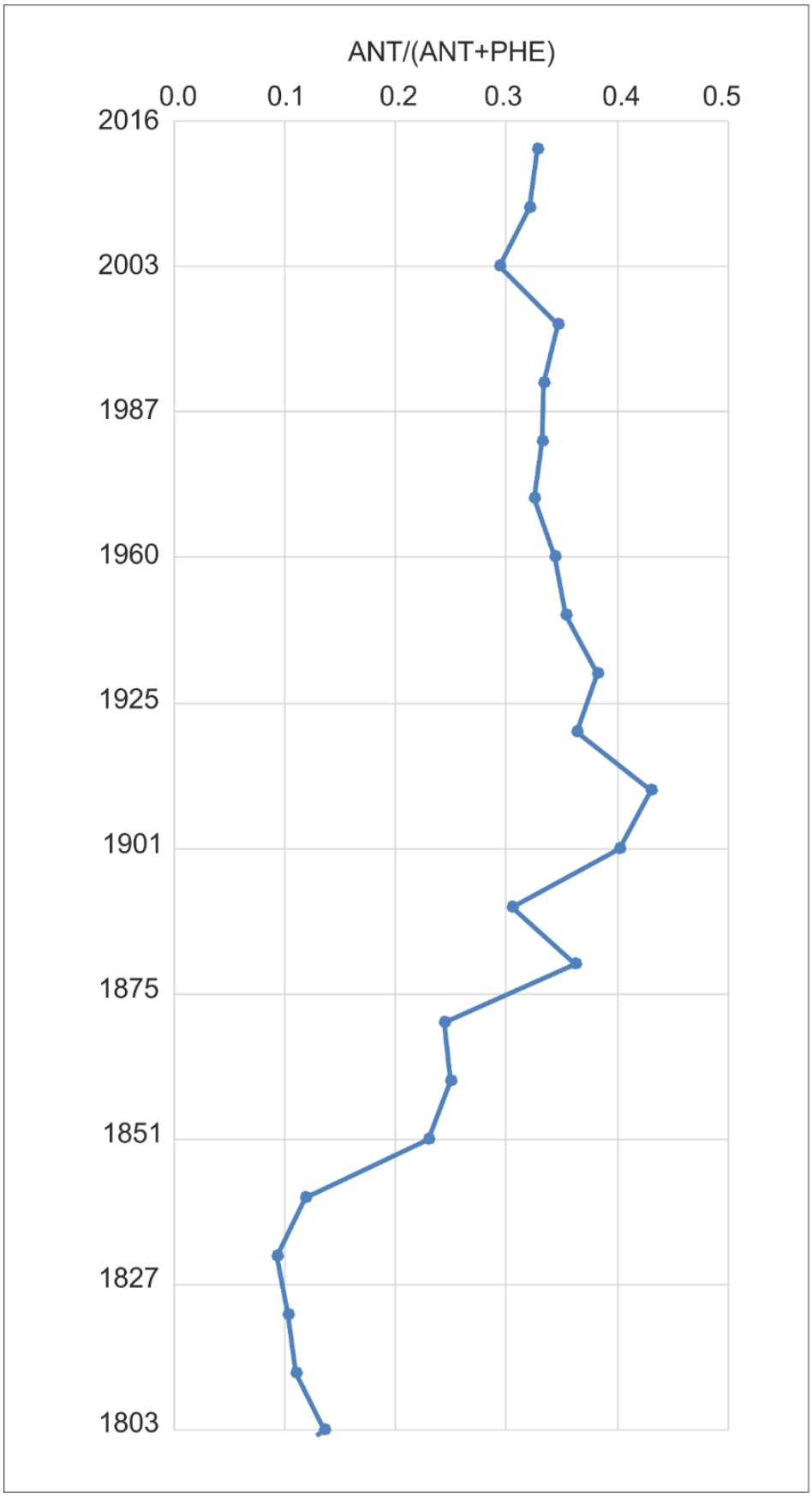

Figure S8.2: Plot of Anthracene/(Anthracene + Phenanthrene) along the AB01 sediment core.

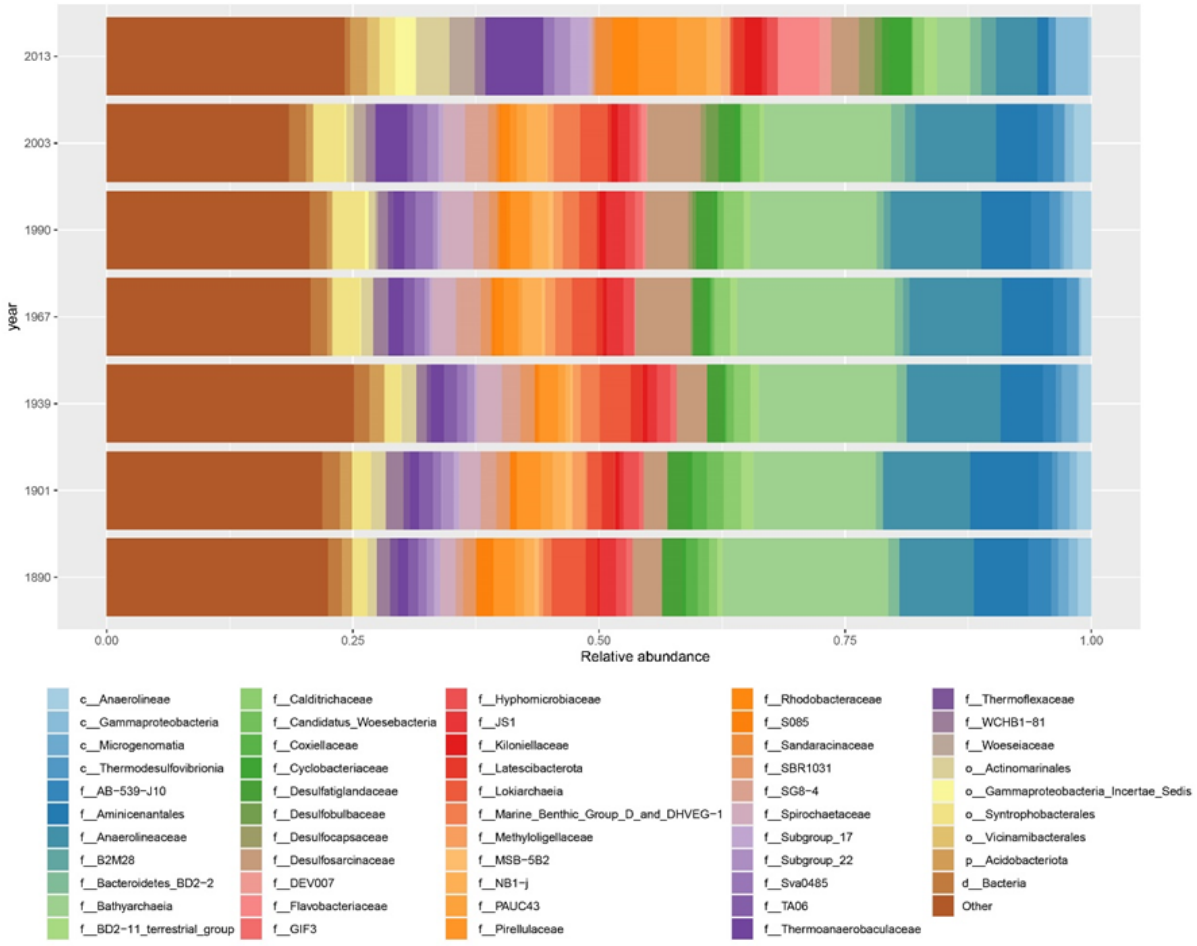

Figure S8.3: Bar-plot showing the taxonomic composition of the bacterial assemblages along the AB01 sediment core at the lowest taxonomic level in terms of sequence contribution to each bacterial taxon. Taxa contributing less than 1% were summed and indicated as "Other". Taxa names are preceded by a letter according to the maximum depth of taxonomic assignment: "d" for domain, "c" for class, "o" for order and "f" for family.

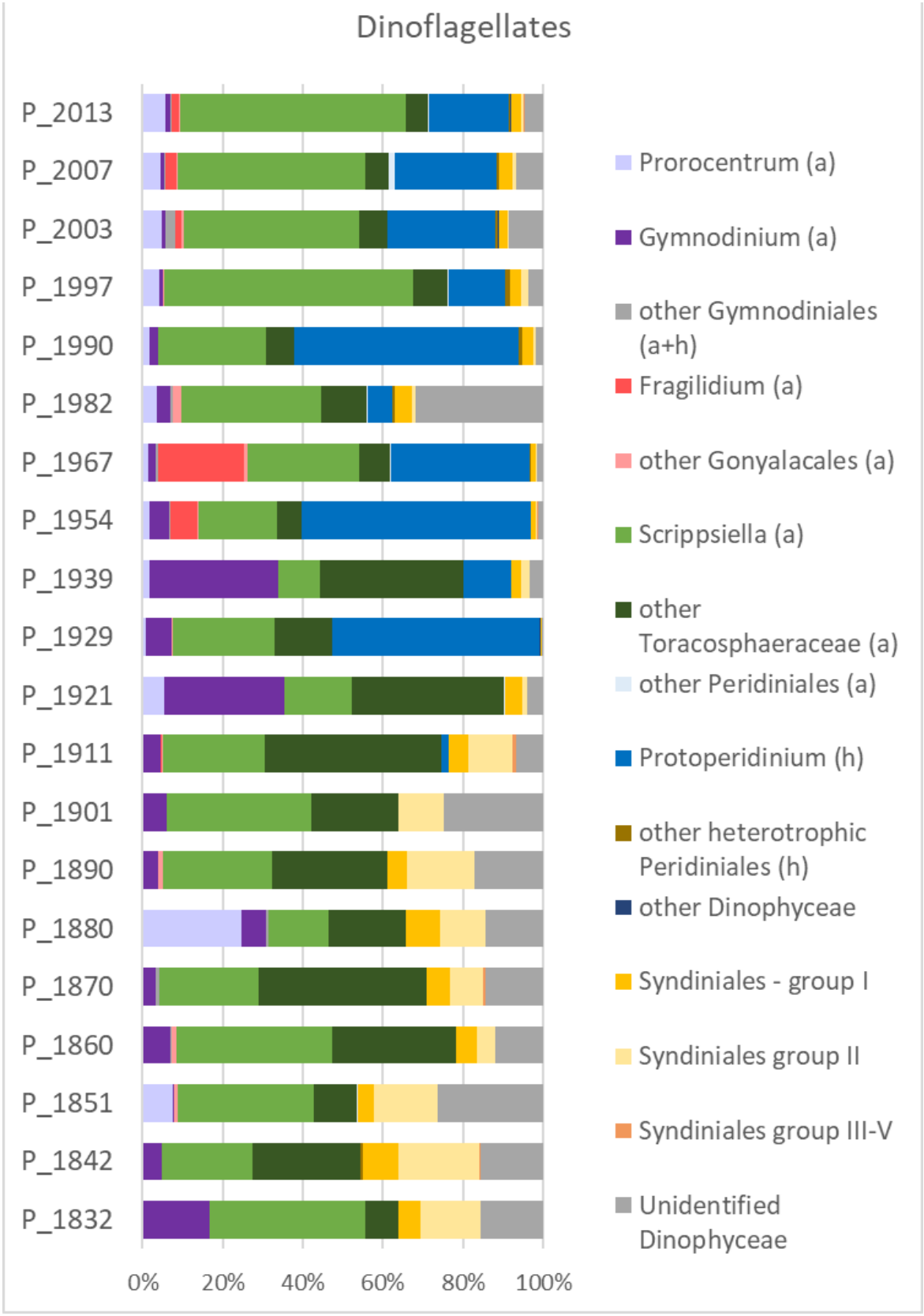

Figure S8. 4: Dinoflagellates in the sediment layers of the AB01 core inferred by metabarcoding of V9 18S rDNA from sedimentary DNA. Taxa are expressed as relative abundance of reads from the normalized dataset. (a) = taxa with chloroplasts; (h) = heterotrophic taxa.

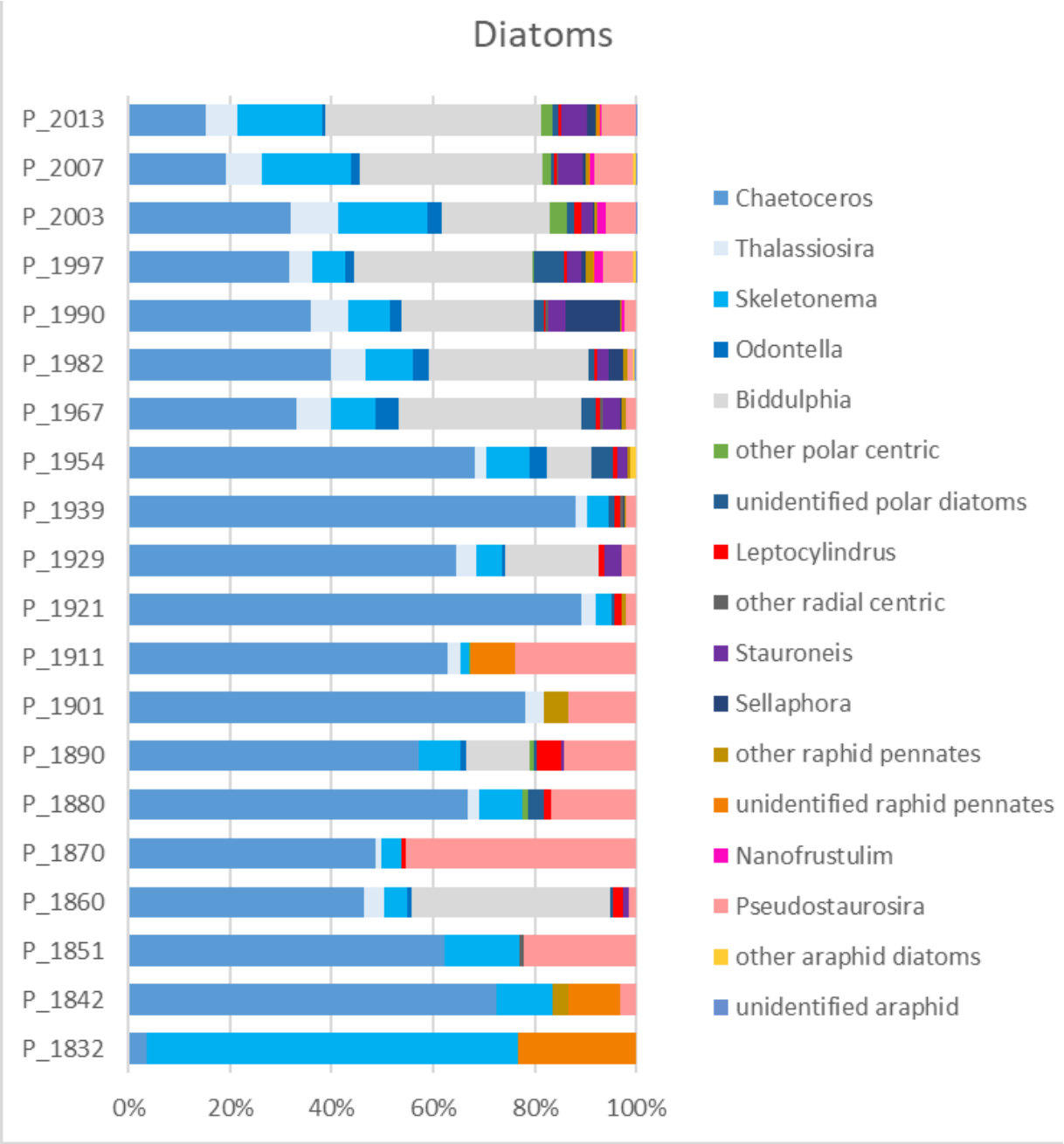

Figure S8. 5: Diatoms in the sediment layers of the AB01 core inferred by metabarcoding of V9 18S rDNA from sedimentary DNA. Taxa are expressed as relative abundance of reads from the normalized dataset.

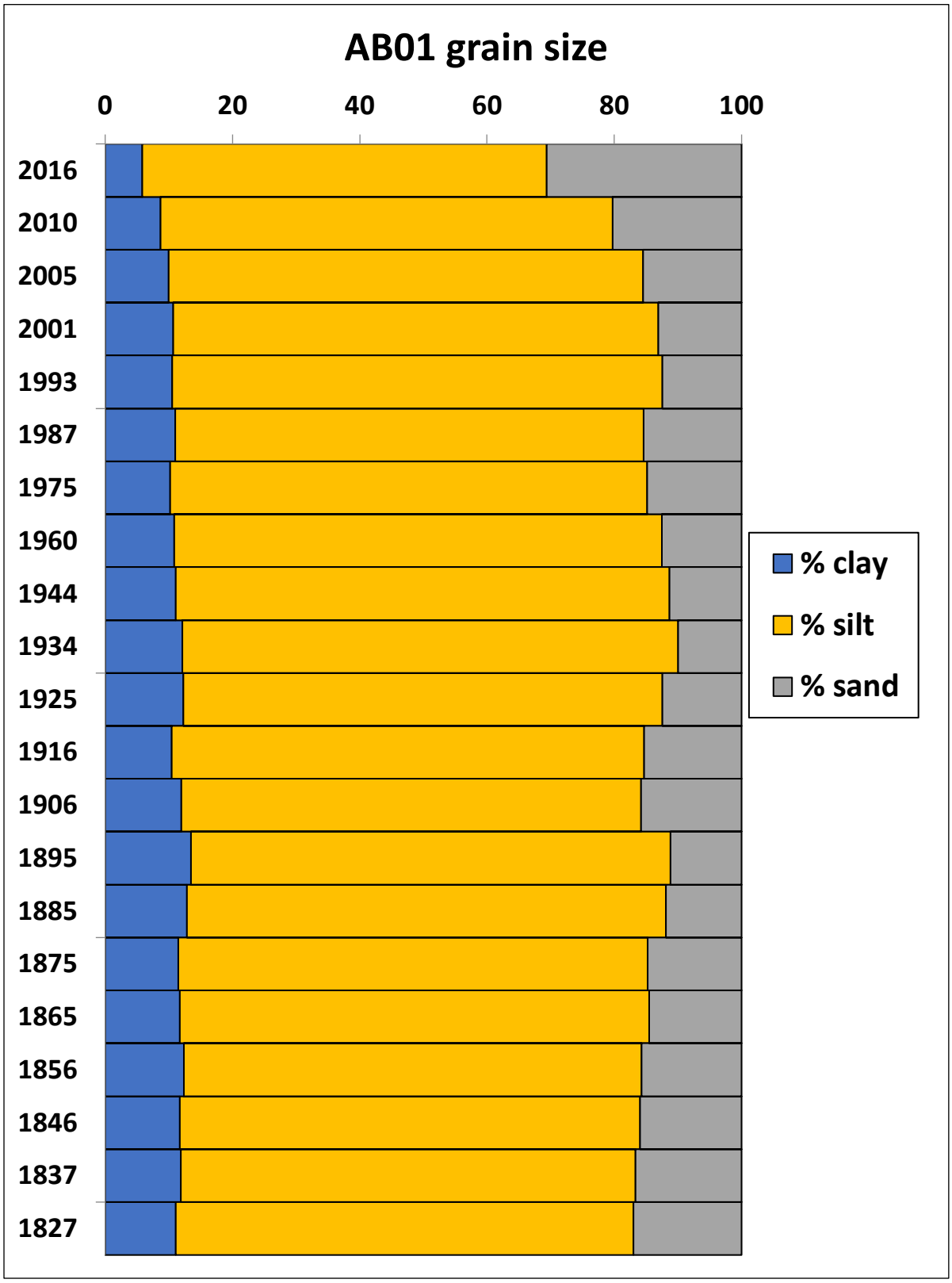

Figure S8. 6: AB01 sediment core grain-size plot: volume percentage of sand, silt and clay (on x axe) vs ages (on y axe) of each odd sediment layer (1 cm resolution) as derived from the geochronology (see Supplementary Table 1).

Chapter 9: Discussion, challenges and applications

In the previous chapters, the analyses of sedimentary DNA, regardless of its age (recent or ancient) or origin (pelagic, benthic), were presented with the main objective: identifying biodiversity. All these studies have raised many questions. Here, I would like to discuss some of them, highlighting strengths and weaknesses of sedimentary DNA approach. First, the issues related to the impact of taphonomic processes on the composition of sedimentary DNA, then those raised by the taxonomy identification of sedDNA sequences, and technical biases. Finally, the issues associated with the practical applications of past and present sedimentary DNA analyses are discussed.

9.1. Taphonomy

Taphonomic processes refer to the source of sedimentary DNA, to its transfer from organisms living in sediments or deposited and buried there, and its preservation in sediments over time (Giguet-Covex et al. 2019; Jia et al. 2022). These processes are key for the interpretation of modern and ancient sedDNA data, but they are poorly understood. In this section, I will discuss some issues related to the DNA taphonomy that have been raised during my study.

9.1.1. Pelagic vs benthic source of sedimentary DNA

Sediments are a DNA repository of benthic (autochthonous) and pelagic organisms (allochthonous). So, any interpretation of sedimentary DNA requires a discrimination between planktonic and benthic sequences. This has been done in chapter 5, where amplicons from Tara Oceans surface samples were used to separate the true deep-sea benthic species from planktonic ones. This study shows that nearly ¼ of the ASVs detected in deep-sea sediments belong to pelagic species, revealing a trend towards the increase of the proportion of pelagic ASVs in sediments from low to high latitudes.

At the same time, chapter 3 shows a variation between planktonic and benthic foraminiferal DNA in the surface sediments (< 10 cm). The DNA of both groups competes during the PCR amplification and sequencing. On the surface, benthic DNA being native and recent, amplifies

preferentially. Below 10 cm, both groups become "ancient", and the proportion of pelagic vs benthic DNA reflects the flux of planktonic foraminifera in sediments.

The question is to which extend sedimentary DNA record reflects the diversity at the sea surface. According to chapters 3 and 5, pelagic DNA transfer occurs vertically from the surface to the bottom, probably by the aggregation of organic particles (Cordier et al. 2022). This has been confirmed by other studies showing that sedimentary DNA preserves patterns of planktonic foraminifera macroecology (Morard et al. 2017). However, pelagic DNA as well as the organisms' skeletons during their gravitational sinking may encounter underwater currents and turbulences. In Chapter 3, we show that foraminiferal community in sedDNA samples can be differentiated between a cold current (the Labrador Current), a warm current (the North Atlantic Current), or a mixture of currents. These results confirm that sedimentary DNA can be used to reconstruct spatial and temporal changes in marine habitats as it provides indirect access to pelagic diversity.

Another question concerns the DNA leaching through the sediment. Studies of terrestrial sites known to have hosted animals at one time (zoos and former farms in Greenland) show no trace of DNA in the sediment layers before animal occupation (Hebsgaard et al. 2009; Andersen et al. 2012). Similarly, in the sub-Antarctic islands, rabbit DNA only appears in sediments in the layers corresponding to their introduction (Ficetola et al. 2018) .To our knowledge, there are no studies on leaching DNA in marine environment. It is reasonable to assume that there is some vertical transfer in surface sediments linked to the bioturbation. However, the impact of this transfer on DNA stratification might not be so important. As illustrated in chapters 4 and 8, there is a clear difference between foraminiferal communities in surface and lower layers (below 20 cm). None of our studies show evidence of DNA percolation in the marine sediments. This is in concordance with studies in lake sediments showing that large molecules such as DNA are immobilized in sediments particles (Smol 2009). In marine sites it is also possible that once attached to a sediment particle, the free DNA or cellular debris does not move and settles with sedimentation. Evidently, more studies are needed on this subject.

9.1.2. DNA preservation of soft vs hard shelled organisms

In contrast to the fossil record, the sedimentary DNA preserves genetic signal from both soft-walled and hard-shelled organisms. In consequence, the morphological and molecular assemblages are not precisely the same. This is especially visible in the case of soft-walled monothalamous foraminifera, which are abundantly present in sedDNA studies, while they are

almost non-existent in the fossil record. This is also the case of acantharia, a group of radiolarians whose strontium skeletons dissolve after their death, and thus acantharia are never found in fossilized form (Decelle et al. 2012b). Indeed, the metabarcoding datasets of radiolarians are often composed mainly of acantharia (Lejzerowicz et al. 2013), yet this may also depend on primer specificity (chapter 4). As illustrated by several studies, many clades of foraminifera and radiolaria are only known from environmental samples, often from extreme environments such as abyssal plains, hydrothermal vents or high-latitude ecosystems (Pawlowski et al. 2011; Biard 2022). There is growing evidence, at least for benthic foraminifera, that many of these clades represent non-skeletonized soft-walled organisms (Holzmann et al. 2022). Sedimentary DNA studies are therefore complementary to microfossil data providing a global vision of biodiversity without selecting species having skeletons or not.

The question is whether the taxonomic composition of sedDNA assemblages could be biased towards non-skeletonized taxa, which DNA would be easier to isolate. It has been suggested that the macroinvertebrates taxa with heavily sclerotized exoskeleton are less represented in metabarcoding datasets (Martins et al. 2021). There is no comparable data regarding foraminifera or radiolaria. Yet, it has been shown that pre-processing (sieving) of sediment samples could enhance the proportion of hard-shelled calcareous foraminifera (Nguyen et al. 2022). If this is confirmed by further studies, it might suggest that the high proportion of soft-walled or naked monothalamous foraminifera in metabarcoding data is caused by their overabundance in small sediment fractions rather than the lack of skeleton.

9.1.3. Sedimentary DNA degradation

Sedimentary DNA degradation could also play a role in taphonomic processes. The DNA fragmentation can be caused by different abiotic and biotic processes. Theoretically, sites with low temperatures (polar regions), anoxia, less ultraviolet (UV) and low microbial activity are more suitable for DNA preservation. The preservation of DNA can also depend on sediment mineralogy. It is known that DNA that bound to fine-grained sediment such as clay minerals will be better preserved (Kanbar et al. 2020). However, some studies show a DNA preservation in coarser sediments as sand, but with high salt concentrations (Lorenz and Wackernagel 1987, Naviaux et al. 2005). The DNA can also bound to humic acids (Crecchio and Stotzky 1998). Yet, the humic acids act as inhibitors during DNA isolation and during PCR amplification, and therefore their presence is not suitable for metabarcoding studies (Boere et al. 2011). Besides mineralogy, the pH values can also influence the DNA bindings properties, for example by

decreasing its adsorption capacities in acidic conditions (Jia et al. 2022). In anoxic environments the microbial activity is reduced and favours DNA preservation. However, several studies (including chapter 4) demonstrate that ancient DNA can also be isolated from sediments deposited under oxygenated conditions (Lejzerowicz et al. 2013; More et al. 2018).

Even in most favourable conditions, DNA in sediments tends to degrade and to fragmentize over time. It is assumed that the DNA degradation affects all organisms in the same way. However, to my knowledge there are no studies that thoroughly examine the relation between specific cellular features and DNA degradation.

On the other hand, a well-documented aspect is the competition between DNA of living organisms and more or less degraded DNA of dead organisms buried in the sediments. In general, targeting small fragments such as V9 in 18S for eukaryotes (<120 bp) promote amplification of degraded sedaDNA, which are known to be in the size range of 500 bp (Boere et al. 2011) to 40 bp (Armbrecht et al. 2020, 2021b). However, there is no guarantee of avoiding preferential amplification of better-preserved recent DNA, especially derived from organisms grown after the coring (Selway et al., 2022). In this case, specific primers can be used to amplify targeted DNA and avoid unwanted contaminants or taxa (Chapter 4). Another option could be the use of metagenomics tools that enable the separation of targeted DNA in silico through a bioinformatic process.

9.2. Taxonomic Identification

An important problem in sedimentary DNA metabarcoding studies is the taxonomic assignment of sequences (ASV or OTU). Naming sequences is necessary when inventorying biodiversity or when using sequence data to infer biotic indices based on ecological values or categories assigned to morphospecies. In most of studies, the unassigned sequences are simply removed from analyses.

9.2.1. Database

The incompleteness of reference databases is the key factor limiting taxonomic assignments. Although curated taxonomic databases, such as the PR2 - Protist Ribosomal Reference database or PFR2 – Planktonic Foraminifera Ribosomal Reference database, have been developed and are continuously updated, these databases include only a small part of diversity corresponding to the species that can be isolated and sometimes cultivated. Many environmental sequences

remain undetermined, especially in poorly explored extreme environments such as hadal regions (Cordier et al. 2019a; Jażdżewska and Mamos 2019) or polar regions (Habura et al. 2004; Nguyen et al. 2022). There are important gaps in regional barcoding databases, as illustrated by our analysis of vascular plants from Papua New Guinea, which DNA was identified in marine sediments (Chapter 4). Only 41 % of morphologically identified species appear in NCBI GenBank, leaving unassigned many plant sequences present in metabarcoding data. Furthermore, the databases can be complete for one of the markers only and include preferentially some taxonomic groups. Thus, mitochondrial COI sequences in Barcode of Life Data (BOLD) have a very good coverage for terrestrial biodiversity (especially insects) but contain limited number of sequences of marine species (Ratnasingham and Hebert 2007; Kvist 2013; Gibson et al. 2015). Indeed, some groups such marine nematodes are underrepresent in databases as shown in (Holovachov et al. 2017). Some phyla could also be better represented in reference databases that others because they live in more accessible environments (coastal waters) and because their size is larger (benthic macrofauna vs microeukaryotes). Furthermore, gap analysis studies reveal some misidentifications or incomplete sequences in large repositories as BOLD or NCBI Genbank (Weigand et al. 2019; Paz and Rinkevich 2021).

9.2.2. Taxonomic resolution

Taxonomic identification also largely depends on the taxonomic resolution of used marker. Some markers can be more resolutive for some taxa. For instance, the mitochondrial COI markers discriminate better metazoan species than the ribosomal 18S, because of its higher evolution rate (Leray et al. 2013). It seems that 18S underestimates metazoan diversity (Tang et al. 2012; Leray and Knowlton 2016). However, the opposite is observed in the case of foraminifera, where the 18S marker is well known to have a high taxonomic resolution (Pawlowski and Lecroq 2010), while the resolution of COI marker is limited (Macher et al. 2022). In foraminifera, the 18S marker identifies even cryptic planktonic species and polymorphism in some species (Pillet et al. 2012; Weber and Pawlowski 2014). However, its use in metabarcoding can be limited by the specificity of primers that do not recognize such taxa as spinose planktonics (chapter 3), lagenids and some species of miliolids. In the latter case, the short region in COI marker seems more efficient (Girard et al. 2022). Thus, the use of multiple markers could be recommended to better asses the biodiversity, instead of having false negatives, especially if some taxa are reluctant to be amplified by specific primers.

The choice of markers, especially for sedaDNA studies, is a compromise between taxonomic resolution and size of amplified fragment. A longer marker would be more resolutive but could be more difficult to amplify. This is the case of the shorter region of P6 loop in chloroplast trnL (Taberlet et al. 2007) and long and high resolutive rbcL (Newmaster et al. 2006). The TrnL marker works better with degraded samples and for this reason, it is preferred for the detection of vascular plants in ancient samples (Clarke et al. 2020; Rijal et al. 2021). In foraminifera also the longer 18S marker including two variable regions (37f and 41F) provides a better resolution. Yet, the shorter 37F region that discriminates most of species is more suitable for degraded samples (Lejzerowicz et al. 2013; Pawlowska et al. 2016).

Specific markers will have higher taxonomic resolution than universal ones as it is designed for targeting specific phyla. For instance, 18S-V9 and 18S-V4, a highly variable region is target for eukaryotic markers. Nonetheless, inside these regions, primers are designed to better target a specific group such as radiolarians in V4 (Decelle et al. 2012a) or ciliates in V9 (Xu et al. 2014).

The capacity of a marker to provide a higher resolution level is necessary in some applications such as in biomonitoring where a species is associated to an ecological value. In this case higher level taxonomic assignment (family or genus) might be misleading as species having different disturbance sensitivities to their environment could be regroup under the same value and distort the ecological status of a given sample/site.

9.2.3. Dealing with unassigned sequences

Metabarcoding data comprise often a significant proportion of unassigned sequences depending on the used marker, targeted taxa, or studied ecosystems (Cordier et al., 2019, 2022). The proportion of unassigned sequences can be particularly high if the taxonomic resolution of the marker is high (e.g., COI) and the database is poor. This is also the case when the sediment samples come from poorly explored extreme environments know to bear high biodiversity (e.g., 2/3 of benthic sequences from abyssal plains were not assigned, Cordier et al. 2022). Unassigned sequences are often interpreted as rare species, chimaeras created during amplification (PCR) or tag jumping (Caron and Countway, 2009; Pedrós-Alió, 2007), or other type of artefacts (Brown et al., 2015c). They are usually removed because they do not provide any information about species ecology or functional groups.

Many approaches have been proposed to conduct taxonomic assignment. These include the alignment-based methods comparing the percentage of similarity of the sequence with the reference, such as Blast (Camacho et al. 2009), probabilistic/algorithm-based methods such as usearch (Edgar 2010) or vsearch (Rognes et al. 2016), machine-learning based methods which train the reference database, such as IDTAXA (Murali et al. 2018), and phylogeny-based methods, which place the sequence in a phylogenetic tree of the whole gene (e.g. 18S) such as HmmUFOtu (Zheng et al. 2018). Some authors combine several methods to assign their sequences and cross-validate the taxonomic assignments (Lejzerowicz et al. 2021).

However, when the diversity is very high and database is poor, none of these methods can efficiently reduce the number of unassigned sequences. Chapter 6 introduces an alternative solution classifying the unassigned deep-sea foraminiferal sequences into lineages and clades based on their genetic signature. As shown by our case study, this method provides some information about the ecology and biogeography of unassigned metabarcodes enabling their use in broader biodiversity studies.

9.3. Technical biases

The sedimentary DNA metabarcoding analyses are subject to many technical biases regardless of the sediment age. Here I will briefly discuss those related to quantitative aspects such as the number of replicates, the volume or weight of extracted sediments, the number of PCR replicates, the sequencing depth, and finally, filtering thresholds.

When sampling for sedimentary DNA, the key question is how many samples (replicates) are needed to achieve the best resolution. Many samples and replicates per core are often required for better temporal and spatial resolution. When species occurrence depends on seasonal activity, samples should be taken in different seasons. Otherwise, the composition would be erroneous or valid for a given season only (Buxton et al. 2018).

Different protocols, reagents, and commercial kits have been compared for sedimentary DNA extraction, showing variation in the species identified and their abundance (reviewed in Armbrecht et al. 2019, 2020, Pawlowski et al. 2021). The volume of sediment used for metabarcoding studies is of particular concern. In some studies, 1 g is suggested to be sufficient for meiofauna, but not for macrofauna, for which 10 g would be needed (Penton et al. 2016; Nascimento et al. 2018; Kang et al. 2021). Still, other authors prefer several technical replicates to compensate for the heterogeneity of the DNA in the sediment samples (Hestetun et al. 2021).

The number of PCRs can influence how rare sequences are amplified. To reliable estimate the presence or absence of a species, simulations recommend at least 8 PCRs (Ficetola et al. 2015). If the species appears in several PCR replicates, it is considered an existing species. However, other experimental studies have shown that three replicates are sufficient (van den Bulcke et al. 2021).

Regarding the sequencing depth, i.e., number of reads obtained for a given sample, it can strongly impact the detection of rare taxa. Recent advances in HTS technologies allows to obtain one million of reads per sample using Novaseq (Singer et al. 2019), instead of commonly used 100'000 reads per sample with Miseq (Cordier et al. 2021)

After sequencing, the way the data is processed will also influence the results depending on the filtering thresholds applied. When clustering, 97% or 99% thresholds will give a different total number of sequences (Alberdi et al. 2018). Eliminating singletons or retaining sequences that occur >10, >100 or >1000 times will influence rare taxa. The number of reads can increase if some species have multiple copies of the amplified gene and are overrepresented. Bailet et al. (2020) compared pipelines used by six laboratories to illustrate bioinformatics' bias. They showed the differences between the numbers of assigned/unassigned reads and hence the discrepancies in the biotic indices inferred from these assignments.

9.4. Applications

In this section, I point out three most common applications of sedimentary DNA, including biodiversity surveys, biomonitoring and bioassessment of present and past environmental impacts.

9.4.1. Biodiversity survey

One of the main applications of sedimentary DNA metabarcoding is to inventory benthic diversity. We show the power of this tool in the case of deep ocean floor diversity. Thanks to metabarcoding surveys, higher diversity has been detected in deep-sea sediments (Cordier et al. 2019a; Lejzerowicz et al. 2021). In those studies, many sequences remain unassigned, meaning a large part of deep-sea diversity still needs to be identified. The sediment DNA reveals the presence of taxa that are either too small or too fragile (i.e, soft-bodied) to be preserved and identify using conventional methods. The existence of a hidden diversity in deep-sea sediments was shown in chapters 5 and 6, confirming numerous other studies detecting species new to

science in almost all deep-sea biodiversity surveys (Danovaro et al. 2010; Goineau and Gooday 2017).

Besides revealing new species, sedimentary DNA metabarcoding can also be used to detect endangered or invasive species. The sensitivity of eDNA-based approaches allows to early detect the presence of non-indigenous species in marine environment (Holman et al. 2019; van den Heuvel-Greve et al. 2021). With such studies a baseline record of those species could be done, providing a basis for further spatial and temporal monitoring of their distribution. However, as mentioned previously, the detection efficacy depends on the marker and primer specificity, which can be responsible for false negatives and erroneous appreciation of species occurrence.

We must keep in mind that metabarcoding data gives semi-quantitative rather than quantitative values. The number of reads cannot be assimilated to number of specimens or species. It might depend on the biomass of different taxa, especially in the case of macrofauna (Patrício et al. 2012). In addition, multiple copies of genes of some species might also increase the number of reads, as illustrated in some foraminifera (Weber and Pawlowski 2014). Metabarcoding data give more reliable information on the presence/absence of taxa. Nevertheless, the relative proportion of species in metabarcoding data allows us to estimate whether a species is abundant or rare.

9.4.2. Biomonitoring of ecological status

Metabarcoding of sedDNA allow to rapidly characterize benthic marine community and can be used to determine the ecological quality status (EcoQS) of disturbed and undisturbed sites (Taberlet et al. 2018; Pawlowski et al. 2021a). Conventional biotic indices are based on species presence/absence or community composition with species assigned to ecological values or categories depending on their response to different types of stressors (Pawlowski et al. 2018). There are selected taxonomic groups that serve as bioindicators, and biotic indices are calculated based on these groups. Among them, the most commonly used is the Azti Marine Biotic Index (AMBI; (Borja et al. 2000) that currently listed over 11,000 taxa with assigned ecological values. Foraminifera are used in the Foram-AMBI comprising the list of indicator species classified into five ecological quality groups and for each type of environment (i.e., coastal, arctic, tropical, etc.) (Jorissen et al. 2018).

Similarly, to morphology-based biomonitoring, metabarcoding data generated from unpolluted/polluted sites can be correlated to physicochemical factors, such as TOC, temperature, pH and can be used to infer *de novo* biotic indices. Thus, bacterial assemblages were used to propose a new biotic index (microgAMBI) for coastal environments (Ban and Alder 2008; Aylagas et al. 2017; Borja 2018), while eukaryotic assemblages were used to infer a new index for assessing the impact of offshore oil and gas platforms (Mauffrey et al. 2021). Benthic foraminifera community composition was also integrated in environmental impact studies of fish farms (Pawlowski et al. 2014), or oil and gas platforms (Laroche et al. 2016; Frontalini et al. 2020b).

Following the same idea of using metabarcoding data to infer biotic indices, in chapter 7 we used g-Foram-AMBI index including Foram-Ambi species only and g-Foram-Motus including all the molecular dataset. By associating TOC concentration with metabarcoding assemblages it was possible to deduce and define EcoQS values and to classify the proximal sites as poor to bad. The results were consistent with geochemical values of heavy metals pollution confirming the usefulness of this approach. However, only a small portion of sequences assigned to level species could be used in our studies because of the lack of indicator values for most of metabarcodes. To solve this problem, one solution would be to identify the potential bioindicators in molecular assemblages. This could be done for instance by conducting mesocosm experiments, such as the studies analysing the impact of chromium and mercury pollution on foraminifera (Frontalini et al. 2018b; Greco et al. 2022). Alternative solution would be to assign the ecological status to a training metabarcoding dataset, including all sequences and to further use it for predicting ecological status of other samples. Several studies show the efficiency of such approach based on supervised machine learning (Cordier et al. 2018; Apothéloz-Perret-Gentil et al. 2021).

9.4.3. Assessing past anthropogenic and climate impacts.

As the biodiversity changes through time are registered in sediments, the sedimentary DNA provides an overview of past ecosystems. Short-term and long-term changes can be assessed using sedimentary DNA. Most of shorter events are related to human activities, while long-term events are mainly related to climate changes.

Short-term events include the period of industrialization, during which the development of ports or industrial complexes in coastal areas has disturbed the benthic habitats. The sediments are loaded with heavy metals and other pollutants, which affect benthic habitats, impacting the

diversity and richness of benthic communities. The chapter 8 presents the biological history of one of the most polluted sites in Europe (Bagnoli Bay) from the pre-industrial period up to the present day. As shown by our study the development of the industrial complex was followed by the decrease of some groups of protists and disappearance of marine plants (*Posidonia oceanica*). Protists (dinoflagellates and stramenopiles) shifts during pollution peaks were also reported in the Brest Bay, in relation to heavy metals pollution accumulated during the World War II (Siano et al. 2021). SedaDNA analysis offers the possibility of assessing biodiversity before the alteration of habitats by human activities (i.e., the industrial period) and provides reference conditions. Hence, this type of study could help in the renaturation of a polluted area.

Long-term records exceed thousands of years and in this case, most often the change in the habitat is related to the climate. In continental environments (i.e lakes), floral or faunal changes linked to deglaciation or glaciation periods were recorded (Parducci et al. 2012; Clarke et al. 2020). In marine environments, the fact that pelagic organisms are also deposited in the sediments, their signal allows to reconstruct past environments related to the sea surface. The sea-ice changes could be followed in Arctic using eukaryotic ancient DNA (de Schepper et al. 2019). Similarly, diatom shifts from sea-ice species to open-ocean species in Antarctic were found (Armbrecht et al. 2022). So, sedaDNA could evaluate the sea-ice transitions and also the paleo-productivity at a given time. In those studies, the molecular assemblages are coherent with reconstructed physical-chemical conditions of sea-surface (i.e Temperature). Thus, in Zimmermann et al. 2021 the diatoms assemblages are well correlated with the SST (sea surface temperatures) and salinity. So, there is a possibility to infer paleo-SST and salinity through sedDNA and palaeoceanographic changes in general as in chapter 3, the downcore planktonic signal correspond to the different ocean currents on the surface.

Using sedimentary ancient DNA, it is possible not only to reconstruct past climates, but also to study how communities and particular species have responded to climate changes. In Pawłowska et al. 2020, a planktonic foraminifer, *N. pachyderma*, shows a different occurrence and abundance of genetic types between glacial and interglacial periods. In chapter 4 of this thesis, Shannon indices of marine protists and plants show high diversity in the glacial periods in the tropical site. Today, the tropical basins are still warming, and a further loss of biodiversity can be expected. Access to such periods allows us to extrapolate or correlate with current and future warming.

Chapter 10: : Conclusions and future perspectives

In this thesis, I attempted to show that the DNA metabarcoding is a powerful tool for assessing biodiversity in recent and ancient sediments. This was done through the paleo-metabarcoding studies focusing on DNA preservation of planktonic and benthic foraminifera in subsurface sediments, as well as on reconstructing the diversity of marine protists and terrestrial flora over a broad time scale covering several glacial and interglacial intervals. Subsequently, we discriminate between pelagic and benthic sequences in surface sediments and proposed a way to deal with the unknown sequences that dominate benthic diversity. And finally, we used sediment DNA metabarcoding to assess ecological status of one of the most polluted marine sites in Europe and to infer its reference conditions prior to industrialization.

After discussing in the previous sections some of the challenges related to sedimentary DNA metabarcoding, I would like to point out here some new technologies and approaches that may help to solve the current challenges and open new avenues for wider applications of sedimentary DNA studies.

First, it is very likely that metabarcoding will be soon replaced or at least complemented by metagenomics based on shotgun sequencing and capture hybridisation. Contrary to metabarcoding, which targets selected taxa through PCR amplification, the metagenomics allows sequencing all DNA fragments present in the sample. In the case of paleogenomics, the shotgun sequencing allows authentication of ancient DNA by size selection, as only the shortest fragments would correspond to ancient DNA. As taxa of interest may sometimes be under-represented in the sample, hybridisation capture has recently been introduced (Gasc and Peyret 2018; Günther et al. 2021). This technique enables enrichment prior to sequencing with targeted taxa 'captured' by magnetic baits (probes). So far, many studies have shown that several taxa can be enriched at once. It is likely that hybridisation capture will become a commonly used method as it avoids the bias of PCR and solves the problem of DNA concentration (Schulte et al. 2021; Armbrecht et al. 2021a).

Second, the continuous advances in high-throughput sequencing technologies are revolutionizing the domain, by enabling generation of sequencing data in real time and incredibly high number. The Oxford Nanopore MinION now offers the option of sequencing sedimentary DNA or RNA on site (Millán-Aguiñaga et al. 2019; Tennant et al. 2022), bypassing

the problems of contamination, transport, storage, and making the data available instantaneously (Tennant et al. 2022). Although the MinION reads are long, so not applicable to very old sediments, and the error rate remains relatively high (7-14%) compared to Illumina platforms (<0.1%), it is foreseen that the sequence quality will be improved very soon (Maggiori et al. 2021). Now it is also possible to increase the sequencing depth with such platforms as Illumina's Novaseq. For example, with Novaseq 6000 it is possible to obtain 20 billion unique reads per run instead of 15 million with Miseq. The greater the sequencing depth, the greater the number of taxa will be detected. Consequently, more diversity can be detected (40% more metazoans according to (Singer et al. 2019)) and the detectability in low-yielding DNA samples can also be improved.

Finally, as in other disciplines, machine learning can help applying sedimentary DNA data in environmental impact assessment and biomonitoring (Vacher et al. 2016; Cordier et al. 2018; Frühe et al. 2021). Machine learning can be used to identify new bioindicators, but more importantly, to overcome the incompleteness of databases by integrating all sequences, including those not assigned to ecological values. This would provide more complete predictive models and better prediction of ecological status on new samples. Future developments of environmental genomics using machine learning could thus help in the discovery and monitoring of interactions between species and their environment.

To conclude, it should be also mentioned that working with ancient and recent sedimentary DNA implies an integration of multidisciplinary fields and experts. Taxonomists are needed to identify species for DNA barcoding, ecologists to understand how the species responds to changes in its environment, sedimentologists and paleoecologists to understand sedimentation, sediment origin and paleoecological changes and finally molecular biologists and bioinformaticians to process the huge data generated by HTS and to implement various analytic tools. Although many aspects of sediment DNA analyses could be automated, the collaborative efforts from all these experts are necessary to generate high-quality data and ensure their accurate interpretation.

References

Acevedo-Trejos E, Marañón E, Merico A (2018) Phytoplankton size diversity and ecosystem function relationships across oceanic regions. Proceedings of the Royal Society B: Biological Sciences. https://doi.org/10.1098/rspb.2018.0621

Agusti S, González-Gordillo JI, Vaqué D, Estrada M, Cerezo MI, Salazar G, Gasol JM, Duarte CM (2015) Ubiquitous healthy diatoms in the deep sea confirm deep carbon injection by the biological pump. Nature Communications 2015 6:1 6:1–8. https://doi.org/10.1038/ncomms8608

Alberdi A, Aizpurua O, Gilbert MTP, Bohmann K (2018) Scrutinizing key steps for reliable metabarcoding of environmental samples. Methods in Ecology and Evolution 9:134–147. https://doi.org/10.1111/2041-210x.12849

Albuquerque P, Mendes M v., Santos CL, Moradas-Ferreira P, Tavares F (2009) DNA signature-based approaches for bacterial detection and identification. Science of The Total Environment 407:3641–3651. https://doi.org/10.1016/j.scitotenv.2008.10.054

Al-Enezi E, Francescangeli F, Balassi E, Borderie S, Al-Hazeem S, Al-Salameen F, Boota Anwar A, Pawlowski J, Frontalini F (2022) Benthic foraminifera as proxies for the environmental quality assessment of the Kuwait Bay (Kuwait, Arabian Gulf): Morphological and metabarcoding approaches. Science of The Total Environment 833:155093. https://doi.org/10.1016/j.scitotenv.2022.155093

Alve E (1995) Benthic foraminiferal responses to estuarine pollution; a review. The Journal of Foraminiferal Research 25:190–203. https://doi.org/10.2113/gsjfr.25.3.190

Alve E, Lepland A, Magnusson J, Backer-Owe K (2009) Monitoring strategies for re-establishment of ecological reference conditions: Possibilities and limitations. Marine Pollution Bulletin 59:297–310. https://doi.org/10.1016/J.MARPOLBUL.2009.08.011

Alve E, Korsun S, Schönfeld J, Dijkstra N, Golikova E, Hess S, Husum K, Panieri G (2016) Foram-AMBI: A sensitivity index based on benthic foraminiferal faunas from North-East Atlantic and Arctic fjords, continental shelves and slopes. Marine Micropaleontology122:1–12. https://doi.org/10.1016/J.MARMICRO.2015.11.001

Alve E, Hess S, Bouchet VMP, Dolven JK, Rygg B (2019) Intercalibration of benthic foraminiferal and macrofaunal biotic indices: An example from the Norwegian Skagerrak coast (NE North Sea). Ecological Indicators 96:107–115. https://doi.org/10.1016/J.ECOLIND.2018.08.037

Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM (2009) A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA Genes. PLoS One. https://doi.org/10.1371/JOURNAL.PONE.0006372

Andersen K, Bird KL, Rasmussen M, Haile J, Breuning-Madsen H, Kjær KH, Orlando L, Gilbert MTP, Willerslev E (2012) Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. Molecular Ecology 21:1966–1979. https://doi.org/10.1111/J.1365-294X.2011.05261.X

André A, Quillévéré F, Morard R, Ujiié Y, Escarguel G, de Vargas C, de Garidel-Thoron T, Douady CJ (2014) SSU rDNA Divergence in Planktonic Foraminifera: Molecular Taxonomy and Biogeographic Implications. PLoS One 9:e104641. https://doi.org/10.1371/journal.pone.0104641

Antich A, Palacin C, Wangensteen OS, Turon X (2021) To denoise or to cluster, that is not the question: optimizing pipelines for COI metabarcoding and metaphylogeography. BMC Bioinformatics 22:1–24. https://doi.org/10.1186/S12859-021-04115-6

Appleby PG, Oldfield F (1978) The calculation of lead-210 dates assuming a constant rate of supply of unsupported 210Pb to the sediment. Catena (Amst) 5:1–8. https://doi.org/10.1016/S0341-8162(78)80002-2

Arienzo M, Donadio C, Mangoni O, Bolinesi F, Stanislao C, Trifuoggi M, Toscanesi M, di Natale G, Ferrara L (2017) Characterization and source apportionment of polycyclic aromatic hydrocarbons (pahs) in the sediments of gulf of Pozzuoli (Campania, Italy). Marine Pollution Bulletin 124:480–487. https://doi.org/10.1016/J.MARPOLBUL.2017.07.006

Armbrecht L, Herrando-Pérez S, Eisenhofer R, Hallegraeff GM, Bolch CJS, Cooper A (2020) An optimized method for the extraction of ancient eukaryote DNA from marine sediments. Molecular Ecology Resources 20:906–919. https://doi.org/10.1111/1755-0998.13162

Armbrecht L, Hallegraeff G, Bolch CJS, Woodward C, Cooper A (2021a) Hybridisation capture allows DNA damage analysis of ancient marine eukaryotes. Scientific Reports 2021 11:1 11:1–14. https://doi.org/10.1038/s41598-021-82578-6

Armbrecht L, Eisenhofer R, Utge J, Sibert EC, Rocha F, Ward R, José J, Karlusich P, Tirichine L, Norris R, Summers M, Bowler C (2021b) Paleo-diatom composition from Santa Barbara Basin deep-sea sediments: a comparison of 18S-V9 and diat-rbcL metabarcoding

vs shotgun metagenomics. ISME Communications 2021 1:1 1:1–10. https://doi.org/10.1038/s43705-021-00070-8

Armbrecht L, Weber ME, Raymo ME, Peck VL, Williams T, Warnock J, Kato Y, Hernández-Almeida I, Hoem F, Reilly B, Hemming S, Bailey I, Martos YM, Gutjahr M, Percuoco V, Allen C, Brachfeld S, Cardillo FG, Du Z, Fauth G, Fogwill C, Garcia M, Glüder A, Guitard M, Hwang J-H, Iizuka M, Kenlee B, O'Connell S, Pérez LF, Ronge TA, Seki O, Tauxe L, Tripathi S, Zheng X (2022) Ancient marine sediment DNA reveals diatom transition in Antarctica. Nature Communications 2022 13:1 13:1–14. https://doi.org/10.1038/s41467-022-33494-4

Armbrecht LH, Coolen MJL, Lejzerowicz F, George SC, Negandhi K, Suzuki Y, Young J, Foster NR, Armand LK, Cooper A, Ostrowski M, Focardi A, Stat M, Moreau JW, Weyrich LS (2019) Ancient DNA from marine sediments: Precautions and considerations for seafloor coring, sample handling and data generation. Earth Sci Rev 196:102887. https://doi.org/10.1016/J.EARSCIREV.2019.102887

Armiento G, Caprioli R, Cerbone A, Chiavarini S, Crovato C, de Cassan M, de Rosa L, Montereali MR, Nardi E, Nardi L, Pezza M, Proposito M, Rimauro J, Salerno A, Salluzzo A, Spaziani F, Zaza F (2020) Current status of coastal sediments contamination in the former industrial area of Bagnoli-Coroglio (Naples, Italy). Chemistry and Ecology 36:579–597. https://doi.org/10.1080/02757540.2020.1747448

Armiento G, Barsanti M, Caprioli R, Chiavarini S, Conte F, Crovato C, de Cassan M, Delbono I, Montereali MR, Nardi E, Parrella L, Pezza M, Proposito M, Rimauro J, Schirone A, Spaziani F (2022) Heavy metal background levels and pollution temporal trend assessment within the marine sediments facing a brownfield area (Gulf of Pozzuoli, Southern Italy). Environmental Monitoring and Assessment 2022 194:11 194:1–21. https://doi.org/10.1007/S10661-022-10480-3

Ascione A, Aucelli PP, Cinque A, di Paola G, Mattei G, Ruello M, Russo Ermolli E, Santangelo N, Valente E (2021) Geomorphology of Naples and the Campi Flegrei: human and natural landscapes in a restless land. Journal of Maps 17:18–28. https://doi.org/10.1080/17445647.2020.1768448

Ausili A, Bergamin L, Romano E (2020) Environmental Status of Italian Coastal Marine Areas Affected by Long History of Contamination. Frontiers in Environmental Sciences 8:34. https://doi.org/10.3389/FENVS.2020.00034

Aylagas E, Borja Á, Tangherlini M, Dell'Anno A, Corinaldesi C, Michell CT, Irigoien X, Danovaro R, Rodríguez-Ezpeleta N (2017) A bacterial community-based index to assess the ecological status of estuarine and coastal environments. Marine Pollution Bulletin 114:679–688. https://doi.org/10.1016/J.MARPOLBUL.2016.10.050

Bailet B, Apothéloz-Perret-Gentil L, Baričević A, Chonova T, Franc A, Frigerio JM, Kelly M, Mora D, Pfannkuchen M, Proft S, Ramon M, Vasselon V, Zimmermann J, Kahlert M (2020) Diatom DNA metabarcoding for ecological assessment: Comparison among bioinformatics pipelines used in six European countries reveals the need for standardization. Science of The Total Environment 745:140948. https://doi.org/10.1016/J.SCITOTENV.2020.140948

Ban N, Alder J (2008) How wild is the ocean? Assessing the intensity of anthropogenic marine activities in British Columbia, Canada. Aquatic Conservation 18:55–85. https://doi.org/10.1002/AQC.816

Barbier EB, Moreno-Mateos D, Rogers AD, Aronson J, Pendleton L, Danovaro R, Henry LA, Morato T, Ardron J, van Dover CL (2014) Ecology: Protect the deep sea. Nature 2014 505:7484 505:475–477. https://doi.org/10.1038/505475a

Barnosky AD, Matzke N, Tomiya S, Wogan GOU, Swartz B, Quental TB, Marshall C, McGuire JL, Lindsey EL, Maguire KC, Mersey B, Ferrer EA (2011) Has the Earth's sixth mass extinction already arrived? Nature 2011 471:7336 471:51–57. https://doi.org/10.1038/nature09678

Barras C, Jorissen FJ, Labrune C, Andral B, Boissery P (2014) Live benthic foraminiferal faunas from the French Mediterranean Coast: Towards a new biotic index of environmental quality. Ecological Indicators 36:719–743. https://doi.org/10.1016/J.ECOLIND.2013.09.028

Barrenechea Angeles I, Lejzerowicz F, Cordier T, Scheplitz J, Kucera M, Ariztegui D, Pawlowski J, Morard R (2020) Planktonic foraminifera eDNA signature deposited on the seafloor remains preserved after burial in marine sediments. Scientific reports 10.1 (2020): 1-12. https://doi.org/10.1038/s41598-020-77179-8

Barton AD, Pershing AJ, Litchman E, Record NR, Edwards KF, Finkel Z v., Kiørboe T, Ward BA (2013) The biogeography of marine plankton traits. Ecology letters 16:522–534. https://doi.org/10.1111/ELE.12063

Beaufort L, Chen Min-Te, Droxler André W. (2005) Physical properties of sediment core MD05-2950. https://doi.pangaea.de/10.1594/PANGAEA.634167 . Accessed 14 Oct 2022

Bergamin L, Romano E, Gabellini M, Ausili A, Carboni MG (2003) Chemical-physical and ecological characterisation in the environmental project of a polluted coastal area: the Bagnoli case study. Mediterranean Marine Science 4:5–20. https://doi.org/10.12681/mms.225

Bergamin L, Romano E, Magno MC, Ausili A, Gabellini M (2005) Pollution monitoring of Bagnoli Bay (Tyrrhenian Sea, Naples, Italy), a sedimentological, chemical and ecological approach. Aquatic Ecosystem Health and Management 8:293–302. https://doi.org/10.1080/14634980500220866

Bertocci I, Dell'Anno A, Musco L, Gambi C, Saggiomo V, Cannavacciuolo M, lo Martire M, Passarelli A, Zazo G, Danovaro R (2019) Multiple human pressures in coastal habitats: variation of meiofaunal assemblages associated with sewage discharge in a post-industrial area. Science of The Total Environment 655:1218–1231. https://doi.org/10.1016/J.SCITOTENV.2018.11.121

Biard T (2022) Diversity and ecology of Radiolaria in modern oceans. Environmental Microbiology 24:2179–2200. https://doi.org/10.1111/1462-2920.16004

Biard T, Bigeard E, Audic S, Poulain J, Gutierrez-Rodriguez A, Pesant S, Stemmann L, Not F (2017) Biogeography and diversity of Collodaria (Radiolaria) in the global ocean. The ISME Journal 2017 11:6 11:1331–1344. https://doi.org/10.1038/ismej.2017.12

Bienhold C, Zinger L, Boetius A, Ramette A (2016) Diversity and biogeography of bathyal and abyssal seafloor bacteria. PLoS One 11:e0148016. https://doi.org/10.1371/journal.pone.0148016

Billett DSM, Lampitt RS, Rice AL, Mantoura RFC (1983) Seasonal sedimentation of phytoplankton to the deep-sea benthos. Nature 1983 302:5908 302:520–522. https://doi.org/10.1038/302520a0

Birks HJB, Line JM, Juggins S, Stevenson AC, ter Braak CJF (1990) Diatoms and pH reconstruction. Philosophical Transactions of the Royal Society of London B, Biological Sciences 327:263–278. https://doi.org/10.1098/RSTB.1990.0062

Blanchet H, Lavesque N, Ruellet T, Dauvin JC, Sauriau PG, Desroy N, Desclaux C, Leconte M, Bachelet G, Janson AL, Bessineton C, Duhamel S, Jourde J, Mayot S, Simon S, de Montaudouin X (2008) Use of biotic indices in semi-enclosed coastal ecosystems and transitional waters habitats—Implications for the implementation of the European Water Framework Directive. Ecological Indicators 8:360–372. https://doi.org/10.1016/J.ECOLIND.2007.04.003

Blum SAE, Lorenz MG, Wackernagel W (1997) Mechanism of Retarded DNA Degradation and Prokaryotic Origin of DNases in Nonsterile Soils. Systematic and applied microbiology 20:513–521. https://doi.org/10.1016/S0723-2020(97)80021-5

Boere AC, Rijpstra WIC, de Lange GJ, Sinninghe Damsté JS, Coolen MJL (2011) Preservation potential of ancient plankton DNA in Pleistocene marine sediments. Geobiology 9:377–393. https://doi.org/10.1111/J.1472-4669.2011.00290.X

Boessenkool S, Mcglynn G, Epp LS, Taylor D, Pimentel M, Gizaw A, Nemomissa S, Brochmann C, Popp M (2014) Use of Ancient Sedimentary DNA as a Novel Conservation Tool for High-Altitude Tropical Biodiversity. Conservation Biology 28:446–455. https://doi.org/10.1111/COBI.12195

Boeuf D, Edwards BR, Eppley JM, Hu SK, Poff KE, Romano AE, Caron DA, Karl DM, DeLong EF (2019) Biological composition and microbial dynamics of sinking particulate organic matter at abyssal depths in the oligotrophic open ocean. Proceedings of the National Academy of Sciences 116:11824–11832. https://doi.org/10.1073/PNAS.1903080116

Boltovskoy D, Correa N (2017) Planktonic equatorial diversity troughs: fact or artifact? Latitudinal diversity gradients in Radiolaria. Ecology 98:112–124. https://doi.org/10.1002/ECY.1623

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang K bin, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu YX, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik A v., Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld

JR, Zhang Y, Zhu Q, Knight R, Caporaso JG (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology 2019 37:8 37:852–857. https://doi.org/10.1038/s41587-019-0209-9

Borja A (2018) Testing the efficiency of a bacterial community-based index (microgAMBI) to assess distinct impact sources in six locations around the world. Ecological Indicators 85:594–602. https://doi.org/10.1016/J.ECOLIND.2017.11.018

Borja A, Muxika I (2005) Guidelines for the use of AMBI (AZTI's Marine Biotic Index) in the assessment of the benthic ecological quality. Marine Pollution Bulletin 50:787–789. https://doi.org/10.1016/J.MARPOLBUL.2005.04.040

Borja A, Franco J, Pérez V (2000) A Marine Biotic Index to Establish the Ecological Quality of Soft-Bottom Benthos Within European Estuarine and Coastal Environments. Marine Pollution Bulletin 40:1100–1114. https://doi.org/10.1016/S0025-326X(00)00061-8

Borja Á, Dauer DM, Grémare A (2012) The importance of setting targets and reference conditions in assessing marine ecosystem quality. Ecological Indicators 12:1–7. https://doi.org/10.1016/J.ECOLIND.2011.06.018

Bouchet VMP, Sauriau PG (2008) Influence of oyster culture practices and environmental conditions on the ecological status of intertidal mudflats in the Pertuis Charentais (SW France): A multi-index approach. Marine Pollution Bulletin 56:1898–1912. https://doi.org/10.1016/J.MARPOLBUL.2008.07.010

Bouchet VMP, Alve E, Rygg B, Telford RJ (2012) Benthic foraminifera provide a promising tool for ecological quality assessment of marine waters. Ecological Indicators 23:66–75. https://doi.org/10.1016/J.ECOLIND.2012.03.011

Bouchet VMP, Goberville E, Frontalini F (2018) Benthic foraminifera to assess Ecological Quality Statuses in Italian transitional waters. Ecological Indicators 84:130–139. https://doi.org/10.1016/J.ECOLIND.2017.07.055

Bouchet VMP, Frontalini F, Francescangeli F, Sauriau PG, Geslin E, Martins MVA, Almogi-Labin A, Avnaim-Katav S, di Bella L, Cearreta A, Coccioni R, Costelloe A, Dimiza MD, Ferraro L, Haynert K, Martínez-Colón M, Melis R, Schweizer M, Triantaphyllou M v., Tsujimoto A, Wilson B, Armynot du Châtelet E (2021) Indicative value of benthic foraminifera for biomonitoring: Assignment to ecological groups of sensitivity to total organic carbon of species from European intertidal areas and transitional waters. Marine Pollution Bulletin 164:112071. https://doi.org/10.1016/J.MARPOLBUL.2021.112071

Brandt MI, Trouche B, Quintric L, Günther B, Wincker P, Poulain J, Arnaud-Haond S (2021) Bioinformatic pipelines combining denoising and clustering tools allow for more comprehensive prokaryotic and eukaryotic metabarcoding. Molecular Ecology Resources 21:1904–1921. https://doi.org/10.1111/1755-0998.13398

Breitburg D, Levin LA, Oschlies A, Grégoire M, Chavez FP, Conley DJ, Garçon V, Gilbert D, Gutiérrez D, Isensee K, Jacinto GS, Limburg KE, Montes I, Naqvi SWA, Pitcher GC, Rabalais NN, Roman MR, Rose KA, Seibel BA, Telszewski M, Yasuhara M, Zhang J (2018) Declining oxygen in the global ocean and coastal waters. Science (1979). https://doi.org/10.1126/SCIENCE.AAM7240

Bremond L, Favier C, Ficetola GF, Tossou MG, Akouégninou A, Gielly L, Giguet-Covex C, Oslisly R, Salzmann U (2017) Five thousand years of tropical lake sediment DNA records from Benin. Quaternary Science Reviews 170:203–211. https://doi.org/10.1016/J.QUASCIREV.2017.06.025

Bribiesca-Contreras G, Dahlgren TG, Amon DJ, Cairns S, Drennan R, Durden JM, Eléaume MP, Hosie AM, Kremenetskaia A, McQuaid K, O'hara TD, Rabone M, Simon-Lledó E, Smith CR, Watling L, Wiklund H, Glover AG (2022) Benthic megafauna of the western Clarion-Clipperton Zone, Pacific Ocean. Zookeys 2022:1–110. https://doi.org/10.3897/zookeys.1113.82172

Briggs L (2020) Ancient DNA research in maritime and underwater archaeology: Pitfalls, promise, and future directions. Open Quaternary 6:1–14. https://doi.org/10.5334/OQ.71

Brosnahan ML, Ralston DK, Fischer AD, Solow AR, Anderson DM (2017) Bloom termination of the toxic dinoflagellate Alexandrium catenella: Vertical migration behavior, sediment infiltration, and benthic cyst yield. Limnology and oceanography 62:2829–2849. https://doi.org/10.1002/LNO.10664

Buxton AS, Groombridge JJ, Griffiths RA (2018) Seasonal variation in environmental DNA detection in sediment and water samples. PLoS One 13:e0191737. https://doi.org/10.1371/JOURNAL.PONE.0191737

Buytaert W, Cuesta-Camacho F, Tobón C (2011) Potential impacts of climate change on the environmental services of humid tropical alpine regions. Global Ecology and Biogeography 20:19–33. https://doi.org/10.1111/J.1466-8238.2010.00585.X

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

Callahan BJ, McMurdie PJ, Holmes SP (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME Journal 11:2639–2643. https://doi.org/10.1038/ismej.2017.119

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: Architecture and applications. BMC Bioinformatics 10:1–9. https://doi.org/10.1186/1471-2105-10-421

Cámara-Leret R, Frodin DG, Adema F, Anderson C, Appelhans MS, Argent G, Arias Guerrero S, Ashton P, Baker WJ, Barfod AS, Barrington D, Borosova R, Bramley GLC, Briggs M, Buerki S, Cahen D, Callmander MW, Cheek M, Chen CW, Conn BJ, Coode MJE, Darbyshire I, Dawson S, Dransfield J, Drinkell C, Duyfjes B, Ebihara A, Ezedin Z, Fu LF, Gideon O, Girmansyah D, Govaerts R, Fortune-Hopkins H, Hassemer G, Hay A, Heatubun CD, Hind DJN, Hoch P, Homot P, Hovenkamp P, Hughes M, Jebb M, Jennings L, Jimbo T, Kessler M, Kiew R, Knapp S, Lamei P, Lehnert M, Lewis GP, Linder HP, Lindsay S, Low YW, Lucas E, Mancera JP, Monro AK, Moore A, Middleton DJ, Nagamasu H, Newman MF, Nic Lughadha E, Melo PHA, Ohlsen DJ, Pannell CM, Parris B, Pearce L, Penneys DS, Perrie LR, Petoe P, Poulsen AD, Prance GT, Quakenbush JP, Raes N, Rodda M, Rogers ZS, Schuiteman A, Schwartsburd P, Scotland RW, Simmons MP, Simpson DA, Stevens P, Sundue M, Testo W, Trias-Blasi A, Turner I, Utteridge T, Walsingham L, Webber BL, Wei R, Weiblen GD, Weigend M, Weston P, de Wilde W, Wilkie P, Wilmot-Dear CM, Wilson HP, Wood JRI, Zhang LB, van Welzen PC (2020) New Guinea has the world's richest island flora. Nature 2020 584:7822 584:579–583. https://doi.org/10.1038/s41586-020-2549-5

Capo E, Giguet-Covex C, Rouillard A, Nota K, Heintzman PD, Vuillemin A, Ariztegui D, Arnaud F, Belle S, Bertilsson S, Bigler C, Bindler R, Brown AG, Clarke CL, Crump SE, Debroas D, Englund G, Ficetola GF, Garner RE, Gauthier J, Gregory-Eaves I, Heinecke L, Herzschuh U, Ibrahim A, Kisand V, Kjær KH, Lammers Y, Littlefair J, Messager E, Monchamp ME, Olajos F, Orsi W, Pedersen MW, Rijal DP, Rydberg J, Spanbauer T, Stoof-Leichsenring KR, Taberlet P, Talas L, Thomas C, Walsh DA, Wang Y, Willerslev E, van Woerkom A, Zimmermann HH, Coolen MJL, Epp LS, Domaizon I, Alsos IG, Parducci L (2021) Lake Sedimentary DNA Research on Past Terrestrial and Aquatic Biodiversity: Overview and Recommendations. Quaternary 2021, Vol 4, Page 6 4:6. https://doi.org/10.3390/QUAT4010006

Carraro L, Mächler E, Wüthrich R, Altermatt F (2020) Environmental DNA allows upscaling spatial patterns of biodiversity in freshwater ecosystems. Nature Communications 2020 11:1 11:1–12. https://doi.org/10.1038/s41467-020-17337-8

Cavaliere M, Barrenechea Angeles I, Montresor M, Bucci C, Brocani L, Balassi E, Margiotta F, Francescangeli F, Bouchet VMP, Pawlowski J, Frontalini F (2021) Assessing the ecological quality status of the highly polluted Bagnoli area (Tyrrhenian Sea, Italy) using foraminiferal eDNA metabarcoding. Science of The Total Environment 790:147871. https://doi.org/10.1016/J.SCITOTENV.2021.147871

Cicchella D, de Vivo B, Lima A (2005) Background and baseline concentration values of elements harmful to human health in the volcanic soils of the metropolitan and provincial areas of Napoli (Italy). Geochemistry: Exploration, Environment, Analysis 5:29–40. https://doi.org/10.1144/1467-7873/03-042

Cirillo C, Bertoli B, Acampora G, Marcolongo L (2022) Bagnoli Urban Regeneration through Phytoremediation. Encyclopedia 2022, Vol 2, Pages 882-892 2:882–892. https://doi.org/10.3390/ENCYCLOPEDIA2020058

Clarke CL, Edwards ME, Brown AG, Gielly L, Lammers Y, Heintzman PD, Ancin-Murguzur FJ, Bråthen KA, Goslar T, Alsos IG (2019) Holocene floristic diversity and richness in northeast Norway revealed by sedimentary ancient DNA (sedaDNA) and pollen. Boreas 48:299–316. https://doi.org/10.1111/bor.12357

Clarke CL, Alsos IG, Edwards ME, Paus A, Gielly L, Haflidason H, Mangerud J, Regnéll C, Hughes PDM, Svendsen JI, Bjune AE (2020) A 24,000-year ancient DNA and pollen record from the Polar Urals reveals temporal dynamics of arctic and boreal plant communities. Quaternary Science Reviews 247:106564. https://doi.org/10.1016/J.QUASCIREV.2020.106564

Consalvey M, Clark MR, Rowden AA, Stocks KI (2010) Life on Seamounts. In: Life in the World's Oceans: Diversity, Distribution, and Abundance. John Wiley & Sons, Ltd, pp 123–138. https://doi.org/10.1002/9781444325508

Coolen MJL, Orsi WD, Balkema C, Quince C, Harris K, Sylva SP, Filipova-Marinova M, Giosan L (2013) Evolution of the plankton paleome in the Black Sea from the Deglacial to Anthropocene. Proceedings of the National Academy of Sciences 110:8609–14. https://doi.org/10.1073/pnas.1219283110

Cordier T, Forster D, Dufresne Y, Martins CIM, Stoeck T, Pawlowski J (2018) Supervised machine learning outperforms taxonomy-based environmental DNA metabarcoding

applied to biomonitoring. Molecular Ecology Resources 18:1381–1391. https://doi.org/10.1111/1755-0998.12926

Cordier T, Barrenechea I, Lejzerowicz F, Reo E, Pawlowski J (2019a) Benthic foraminiferal DNA metabarcodes significantly vary along a gradient from abyssal to hadal depths and between each side of the Kuril-Kamchatka trench. Progress in Oceanography 178:102175. https://doi.org/10.1016/J.POCEAN.2019.102175

Cordier T, Frontalini F, Cermakova K, Apothéloz-Perret-Gentil L, Treglia M, Scantamburlo E, Bonamin V, Pawlowski J (2019b) 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, Angeles IB, Henry N, Lejzerowicz F, Berney C, Morard R, Brandt A, Cambon-Bonavita MA, Guidi L, Lombard F, Arbizu PM, Massana R, Orejas C, Poulain J, Smith CR, Wincker P, Arnaud-Haond S, Gooday AJ, de Vargas C, Pawlowski J (2022) Patterns of eukaryotic diversity from the surface to the deep-ocean sediment. Science Advances 8:9309. https://doi.org/10.1126/SCIADV.ABJ9309

Corinaldesi C, Dell'Anno A, Danovaro R (2007) Early diagenesis and trophic role of extracellular DNA in different benthic ecosystems. Limnology and oceanography 52:1710–1717. https://doi.org/10.4319/lo.2007.52.4.1710

Corinaldesi C, Beolchini F, Dell'Anno a (2008) Damage and degradation rates of extracellular DNA in marine sediments: implications for the preservation of gene sequences. Molecular Ecology 17:3939–51. https://doi.org/10.1111/j.1365-294X.2008.03880.x

Corinaldesi C, Barucca M, Luna GM, Dell'Anno A (2011) Preservation, origin and genetic imprint of extracellular DNA in permanently anoxic deep-sea sediments. Molecular Ecology 20:642–54. https://doi.org/10.1111/j.1365-294X.2010.04958.x

Corliss BH, Emerson S (1990) Distribution of rose bengal stained deep-sea benthic foraminifera from the Nova Scotian continental margin and Gulf of Maine. Deep Sea Research Part A Oceanographic Research Papers 37:381–400. https://doi.org/10.1016/0198-0149(90)90015-N

Costa C, Fanelli E, Marini S, Danovaro R, Aguzzi J (2020) Global Deep-Sea Biodiversity Research Trends Highlighted by Science Mapping Approach. Frontiers in Marine Science 7:384. https://doi.org/10.3389/FMARS.2020.00384

Costanza R, D'Arge R, de Groot R, Farber S, Grasso M, Hannon B, Limburg K, Naeem S, O'Neill R v., Paruelo J, Raskin RG, Sutton P, van den Belt M (1997) The value of the world's ecosystem services and natural capital. Nature 1997 387:6630 387:253–260. https://doi.org/10.1038/387253a0

Crecchio C, Stotzky G (1998) Binding of DNA on humic acids: Effect on transformation of Bacillus subtilis and resistance to DNase. Soil Biology and Biochemistry 30:1061–1067. https://doi.org/10.1016/S0038-0717(97)00248-4

Dabney J, Meyer M, Pääbo S (2013) Ancient DNA Damage. Cold Spring Harbor perspectives in biology 5.(7). https://doi.org/10.1101/CSHPERSPECT.A012567

Damiani V, Baudo R, de Rosa S, de Simone R, Ferretti O, Izzo G, Serena F (1987) A case study: Bay of Pozzuoli (Gulf of Naples, Italy). Hydrobiologia 1987 149:1 149:201–211. https://doi.org/10.1007/BF00048661

Danovaro R, Company JB, Corinaldesi C, D'Onghia G, Galil B, Gambi C, Gooday AJ, Lampadariou N, Luna GM, Morigi C, Olu K, Polymenakou P, Ramirez-Llodra E, Sabbatini A, Sardá F, Sibuet M, Tselepides A (2010) Deep-Sea Biodiversity in the Mediterranean Sea: The Known, the Unknown, and the Unknowable. PLoS One 5:e11832. https://doi.org/10.1371/JOURNAL.PONE.0011832

Danovaro R, Molari M, Corinaldesi C, Dell'Anno A (2016) Macroecological drivers of archaea and bacteria in benthic deep-sea ecosystems. Science Advances 2.4. https://doi.org/10.1126/SCIADV.1500961

Danovaro R, Fanelli E, Aguzzi J, Billett D, Carugati L, Corinaldesi C, Dell'Anno A, Gjerde K, Jamieson AJ, Kark S, McClain C, Levin L, Levin N, Ramirez-Llodra E, Ruhl H, Smith CR, Snelgrove PVR, Thomsen L, van Dover CL, Yasuhara M (2020a) Ecological variables for developing a global deep-ocean monitoring and conservation strategy. Nature Ecology & Evolution 2020 4:2 4:181–192. https://doi.org/10.1038/s41559-019-1091-z

Darling KF, Wade CM (2008) The genetic diversity of planktic foraminifera and the global distribution of ribosomal RNA genotypes. Marine Micropaleontology 67:216–238. https://doi.org/10.1016/J.MARMICRO.2008.01.009

Darling KF, Wade CM, Kroon D, Brown AJL (1997) Planktic foraminiferal molecular evolution and their polyphyletic origins from benthic taxa. Marine Micropaleontology 30:251–266. https://doi.org/10.1016/S0377-8398(96)00057-6

de Jonge DSW, Merten V, Bayer T, Puebla O, Reusch TBH, Hoving HJT (2021) A novel metabarcoding primer pair for environmental DNA analysis of Cephalopoda (Mollusca)

targeting the nuclear 18S rRNA region. Royal Society open science https://doi.org/10.1098/RSOS.201388

De Schepper S, Ray JL, Skaar KS, Sadatzki H, Ijaz UZ, Stein R, Larsen A (2019) The potential of sedimentary ancient DNA for reconstructing past sea ice evolution. The ISME Journal 13:2566–2577. https://doi.org/10.1038/s41396-019-0457-1

de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E, Berney Ć, le Bescot N, Probert I, Carmichael M, Poulain J, Romac S, Colin S, Aury JM, Bittner L, Chaffron S, Dunthorn M, Engelen S, Flegontova O, Guidi L, Horák A, Jaillon O, Lima-Mendez G, Lukeš J, Malviya S, Morard R, Mulot M, Scalco E, Siano R, Vincent F, Zingone A, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Acinas SG, Bork P, Bowler C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Raes J, Sieracki ME, Speich S, Stemmann L, Sunagawa S, Weissenbach J, Wincker P, Karsenti E, Boss E, Follows M, Karp-Boss L, Krzic U, Reynaud EG, Sardet C, Sullivan MB, Velayoudon D (2015) Eukaryotic plankton diversity in the sunlit ocean. Science. https://doi.org/10.1126/SCIENCE.1261605

de Vivo B, Lima A (2008) Characterization and remediation of a brownfield site: the Bagnoli case in Italy. Environmental Geochemistry: Site Characterization, Data Analysis and Case Histories 355–385. https://doi.org/10.1016/B978-0-444-53159-9.00015-2

Decelle J, Suzuki N, Mahé F, de Vargas C, Not F (2012a) Molecular Phylogeny and Morphological Evolution of the Acantharia (Radiolaria). Protist 163:435–450. https://doi.org/10.1016/J.PROTIS.2011.10.002

Decelle J, Siano R, Probert I, Poirier C, Not F (2012b) Multiple microalgal partners in symbiosis with the acantharian Acanthochiasma sp. (Radiolaria). Symbiosis 58:233–244. https://doi.org/10.1007/S13199-012-0195-X

del Carmen Gomez Cabrera M, Young JM, Roff G, Staples T, Ortiz JC, Pandolfi JM, Cooper A (2019) Broadening the taxonomic scope of coral reef palaeoecological studies using ancient DNA. Molecular Ecology 28:2636–2652. https://doi.org/10.1111/MEC.15038

Delbono I, Barsanti M, Schirone A, Conte F, Delfanti R (2016) 210Pb mass accumulation rates in the depositional area of the Magra River (Mediterranean Sea, Italy). Continental Shelf Research 124:35–48. https://doi.org/10.1016/j.csr.2016.05.010

Dell'Anno A, Stefano B, Danovaro R (2002) Quantification, base composition, and fate of extracellular DNA in marine sediments. Limnology and oceanography 47:899–905. https://doi.org/10.4319/LO.2002.47.3.0899

Dijkstra N, Junttila J, Aagaard-Sørensen S (2017) Environmental baselines and reconstruction of Atlantic Water inflow in Bjørnøyrenna, SW Barents Sea, since 1800 CE. Marine environmental research 132:117–131. https://doi.org/10.1016/J.MARENVRES.2017.10.012

Dimiza MD, Triantaphyllou M v., Koukousioura O, Hallock P, Simboura N, Karageorgis AP, Papathanasiou E (2016) The Foram Stress Index: A new tool for environmental assessment of soft-bottom environments using benthic foraminifera. A case study from the Saronikos Gulf, Greece, Eastern Mediterranean. Ecological Indicators 60:611–621. https://doi.org/10.1016/J.ECOLIND.2015.07.030

Dirzo R, Young HS, Galetti M, Ceballos G, Isaac NJB, Collen B (2014) Defaunation in the Anthropocene. Science (1979) 345:401–406. https://doi.org/10.1126/SCIENCE.1251817

Djurhuus A, Closek CJ, Kelly RP, Pitz KJ, Michisaki RP, Starks HA, Walz KR, Andruszkiewicz EA, Olesin E, Hubbard K, Montes E, Otis D, Muller-Karger FE, Chavez FP, Boehm AB, Breitbart M (2020) Environmental DNA reveals seasonal shifts and potential interactions in a marine community. Nature Communications 2020 11:1 11:1–9. https://doi.org/10.1038/s41467-019-14105-1

Dommain R, Andama M, McDonough MM, Prado NA, Goldhammer T, Potts R, Maldonado JE, Nkurunungi JB, Campana MG (2020) The Challenges of Reconstructing Tropical Biodiversity With Sedimentary Ancient DNA: A 2200-Year-Long Metagenomic Record From Bwindi Impenetrable Forest, Uganda. Frontiers in Ecology and Evolution 8:218. https://doi.org/10.3389/FEVO.2020.00218

Dornelas M, Gotelli NJ, McGill B, Shimadzu H, Moyes F, Sievers C, Magurran AE (2014) Assemblage time series reveal biodiversity change but not systematic loss. Science (1979) 344:296–299. https://doi.org/10.1126/SCIENCE.1248484

Drinkwater KF (1996) Atmospheric and oceanic variability in the Northwest Atlantic during the 1980 s and early 1990 s. Journal of Northwest Atlantic fishery science Dartmouth NS 18:77–97. https://doi.org/10.2960/J.v18.a6

Dufresne Y, Lejzerowicz F, Perret-Gentil LA, Pawlowski J, Cordier T (2019) SLIM: A flexible web application for the reproducible processing of environmental DNA metabarcoding data. BMC Bioinformatics. https://doi.org/10.1186/s12859-019-2663-2

Dulias K, Stoof-Leichsenring KR, Pestryakova LA, Herzschuh U (2017) Sedimentary DNA versus morphology in the analysis of diatom-environment relationships. Journal of Paleolimnology 57:51–66. https://doi.org/10.1007/S10933-016-9926-Y

Ebbe B, Billett DSM, Brandt A, Ellingsen K, Glover A, Keller S, Malyutina M, Martínez Arbizu P, Molodtsova T, Rex M, Smith C, Tselepides A (2010) Diversity of Abyssal Marine Life. In: Life in the World's Oceans: Diversity, Distribution, and Abundance. John Wiley & Sons, Ltd, pp 139–160. **https://doi.org/10.1002/9781444325508.ch8**

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461. https://doi.org/10.1093/BIOINFORMATICS/BTQ461

Edgar RC (2018) Updating the 97% identity threshold for 16S ribosomal RNA OTUs. Bioinformatics 34:2371–2375. https://doi.org/10.1093/BIOINFORMATICS/BTY113

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194–200. https://doi.org/10.1093/bioinformatics/btr381

el Kateb A, Stalder C, Martínez-Colón M, Mateu-Vicens G, Francescangeli F, Coletti G, Stainbank S, Spezzaferri S (2020) Foraminiferal-based biotic indices to assess the ecological quality status of the Gulf of Gabes (Tunisia): Present limitations and future perspectives. Ecological Indicators 111:105962. https://doi.org/10.1016/J.ECOLIND.2019.105962

Epp LS, Stoof KR, Trauth MH, Tiedemann R (2010) Historical genetics on a sediment core from a Kenyan lake: Intraspecific genotype turnover in a tropical rotifer is related to past environmental changes. Journal of Paleolimnology 43:939–954. https://doi.org/10.1007/S10933-009-9379-7

Epp LS, Zimmermann HH, Stoof-Leichsenring KR (2019) Sampling and extraction of ancient DNA from sediments. Methods in Molecular Biology 1963:31–44. https://doi.org/10.1007/978-1-4939-9176-1_5

Esling P, Lejzerowicz F, Pawlowski J (2015) Accurate multiplexing and filtering for high-throughput amplicon-sequencing. Nucleic acids research 43:2513–2524. https://doi.org/10.1093/nar/gkv107

Fasciglione P, Barra M, Santucci A, Ciancimino S, Mazzola S, Passaro S (2016) Macrobenthic community status in highly polluted area: a case study from Bagnoli, Naples Bay, Italy. Rendiconti Lincei 27:229–239. https://doi.org/10.1007/S12210-015-0467-5

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

Ficetola GF, Poulenard J, Sabatier P, Messager E, Gielly L, Leloup A, Etienne D, Bakke J, Malet E, Fanget B, Støren E, Reyss JL, Taberlet P, Arnaud F (2018) DNA from lake sediments reveals long-term ecosystem changes after a biological invasion. Science Advances. https://doi.org/10.1126/SCIADV.AAR4292

Flegontova O, Flegontov P, Malviya S, Audic S, Wincker P, de Vargas C, Bowler C, Lukeš J, Horák A (2016) Extreme Diversity of Diplonemid Eukaryotes in the Ocean. Current Biology 26:3060–3065. https://doi.org/10.1016/j.cub.2016.09.031

Floyd RM, Rogers AD, Lambshead PJD, Smith CR (2005) Nematode-specific PCR primers for the 18S small subunit rRNA gene. Molecular Ecology Notes 5:611–612. https://doi.org/10.1111/J.1471-8286.2005.01009.X

Fonseca VG, Sinniger F, Gaspar JM, Quince C, Creer S, Power DM, Peck LS, Clark MS (2017) Revealing higher than expected meiofaunal diversity in Antarctic sediments: a metabarcoding approach. Scientific Reports 2017 7:1 7:1–11. https://doi.org/10.1038/s41598-017-06687-x

Fordham DA, Jackson ST, Brown SC, Huntley B, Brook BW, Dahl-Jensen D, Thomas Gilbert MP, Otto-Bliesner BL, Svensson A, Theodoridis S, Wilmshurst JM, Buettel JC, Canteri E, McDowell M, Orlando L, Pilowsky J, Rahbek C, Nogues-Bravo D (2020) Using paleo-archives to safeguard biodiversity under climate change. Science (1979). https://doi.org/10.1126/SCIENCE.ABC5654

Forster D, Dunthorn M, Stoeck T, Mahé F (2016) Comparison of three clustering approaches for detecting novel environmental microbial diversity. PeerJ 4:e1692. https://doi.org/10.7717/peerj.1692

Forster D, Lentendu G, Filker S, Dubois E, Wilding TA, Stoeck T (2019) Improving eDNA-based protist diversity assessments using networks of amplicon sequence variants. Environmental Microbiology 21:4109–4124. https://doi.org/10.1111/1462-2920.14764

Fossile E, Sabbatini A, Spagnoli F, Caridi F, Dell'Anno A, de Marco R, Dinelli E, Droghini E, Tramontana M, Negri A (2021) Sensitivity of foraminiferal-based indices to evaluate the ecological quality status of marine coastal benthic systems: A case study of the Gulf of Manfredonia (southern Adriatic Sea). Marine Pollution Bulletin 163:111933. https://doi.org/10.1016/J.MARPOLBUL.2020.111933

Francescangeli F, Armynot du Chatelet E, Billon G, Trentesaux A, Bouchet VMP (2016) Palaeo-ecological quality status based on foraminifera of Boulogne-sur-Mer harbour (Pas-

de-Calais, Northeastern France) over the last 200 years. Marine environmental research 117:32–43. https://doi.org/10.1016/J.MARENVRES.2016.04.002

Francescangeli F, Quijada M, Armynot du Châtelet E, Frontalini F, Trentesaux A, Billon G, Bouchet VMP (2020) Multidisciplinary study to monitor consequences of pollution on intertidal benthic ecosystems (Hauts de France, English Channel, France): Comparison with natural areas. Marine environmental research 160:105034. https://doi.org/10.1016/J.MARENVRES.2020.105034

Frenzel P (2019) Fossils of the southern Baltic Sea as palaeoenvironmental indicators in multi-proxy studies. Quaternary International 511:6–21. https://doi.org/10.1016/J.QUAINT.2018.09.014

Frontalini F, Greco M, di Bella L, Lejzerowicz F, Reo E, Caruso A, Cosentino C, Maccotta A, Scopelliti G, Nardelli MP, Losada MT, Armynot du Châtelet E, Coccioni R, Pawlowski J (2018a) Assessing the effect of mercury pollution on cultured benthic foraminifera community using morphological and eDNA metabarcoding approaches. Marine Pollution Bulletin 129:512–524. https://doi.org/10.1016/J.MARPOLBUL.2017.10.022

Frontalini F, Cordier T, Balassi E, Armynot du Chatelet E, Cermakova K, Apothéloz-Perret-Gentil L, Martins MVA, Bucci C, Scantamburlo E, Treglia M, Bonamin V, Pawlowski J (2020a) Benthic foraminiferal metabarcoding and morphology-based assessment around three offshore gas platforms: Congruence and complementarity. Environmental International 144:106049. https://doi.org/10.1016/J.ENVINT.2020.106049

Frontalini F, Cordier T, Balassi E, Armynot du Chatelet E, Cermakova K, Apothéloz-Perret-Gentil L, Martins MVA, Bucci C, Scantamburlo E, Treglia M, Bonamin V, Pawlowski J (2020b) Benthic foraminiferal metabarcoding and morphology-based assessment around three offshore gas platforms: Congruence and complementarity. Environmental International 144:106049. https://doi.org/10.1016/J.ENVINT.2020.106049

Frøslev TG, Kjøller R, Bruun HH, Ejrnæs R, Brunbjerg AK, Pietroni C, Hansen AJ (2017) Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nature Communications 2017 8:1 8:1–11. https://doi.org/10.1038/s41467-017-01312-x

Frühe L, Cordier T, Dully V, Breiner HW, Lentendu G, Pawlowski J, Martins C, Wilding TA, Stoeck T (2021) Supervised machine learning is superior to indicator value inference in monitoring the environmental impacts of salmon aquaculture using eDNA metabarcodes. In: Molecular Ecology. pp 2988–3006. https://doi.org/10.1111/mec.15434

Gambi C, Dell'Anno A, Corinaldesi C, lo Martire M, Musco L, da Ros Z, Armiento G, Danovaro R (2020) Impact of historical contamination on meiofaunal assemblages: The case study of the Bagnoli-Coroglio Bay (southern Tyrrhenian Sea). Marine environmental research 156:104907. https://doi.org/10.1016/J.MARENVRES.2020.104907

Ganser MH, Forster D, Liu W, Lin X, Stoeck T, Agatha S (2021) Genetic Diversity in Marine Planktonic Ciliates (Alveolata, Ciliophora) Suggests Distinct Geographical Patterns – Data From Chinese and European Coastal Waters. Frontiers in Marine Science 8:395. https://doi.org/10.3389/FMARS.2021.643822

Ganser MH, Santoferrara LF, Agatha S (2022) Molecular signature characters complement taxonomic diagnoses: A bioinformatic approach exemplified by ciliated protists (Ciliophora, Oligotrichea). Molecular phylogenetics and evolution. 170:107433. https://doi.org/10.1016/J.YMPEV.2022.107433

García-López R, Cornejo-Granados F, Lopez-Zavala AA, Cota-Huízar A, Sotelo-Mundo RR, Gómez-Gil B, Ochoa-Leyva A (2021) OTUs and ASVs Produce Comparable Taxonomic and Diversity from Shrimp Microbiota 16S Profiles Using Tailored Abundance Filters. Genes 2021, Vol 12, Page 564 12:564. https://doi.org/10.3390/GENES12040564

Gasc C, Peyret P (2018) Hybridization capture reveals microbial diversity missed using current profiling methods. Microbiome 6:61. https://doi.org/10.1186/S40168-018-0442-3

Gibson JF, Shokralla S, Curry C, Baird DJ, Monk WA, King I, Hajibabaei M (2015) Large-Scale Biomonitoring of Remote and Threatened Ecosystems via High-Throughput Sequencing. PLoS One 10:e0138432. https://doi.org/10.1371/JOURNAL.PONE.0138432

Giguet-Covex C, Ficetola GF, Walsh K, Poulenard J, Bajard M, Fouinat L, Sabatier P, Gielly L, Messager E, Develle AL, David F, Taberlet P, Brisset E, Guiter F, Sinet R, Arnaud F (2019) New insights on lake sediment DNA from the catchment: importance of taphonomic and analytical issues on the record quality. Scientific Reports 2019 9:1 9:1–21. https://doi.org/10.1038/s41598-019-50339-1

Girard EB, Langerak A, Jompa J, Wangensteen OS, Macher JN, Renema W (2022) Mitochondrial Cytochrome Oxidase Subunit 1: A Promising Molecular Marker for Species Identification in Foraminifera. Frontiers in Marine Science 9:42. https://doi.org/10.3389/FMARS.2022.809659

Glover AG, Smith CR (2003) The deep-sea floor ecosystem: current status and prospects of anthropogenic change by the year 2025. Environ Conserv 30:219–241. https://doi.org/10.1017/S0376892903000225

Glover AG, Wiklund H, Chen C, Dahlgren TG (2018) Point of view: Managing a sustainable deep-sea 'blue economy' requires knowledge of what actually lives there. Elife. https://doi.org/10.7554/ELIFE.41319

Goineau A, Gooday AJ (2017) Novel benthic foraminifera are abundant and diverse in an area of the abyssal equatorial Pacific licensed for polymetallic nodule exploration. Scientific Reports 2017 7:1 7:1–15. https://doi.org/10.1038/srep45288

Gold Z, Curd EE, Goodwin KD, Choi ES, Frable BW, Thompson AR, Walker HJ, Burton RS, Kacev D, Martz LD, Barber PH (2021) Improving metabarcoding taxonomic assignment: A case study of fishes in a large marine ecosystem. Molecular Ecology Resources 21:2546–2564. https://doi.org/10.1111/1755-0998.13450

Gooday AJ (2001) Benthic Foraminifera. In: Encyclopedia of Ocean Sciences. pp 274–286 https://doi.org/10.1016/B978-0-12-409548-9.09071-0

Gooday AJ (2019) Deep-Sea Benthic Foraminifera. Encyclopedia of Ocean Sciences 684–705. https://doi.org/10.1016/B978-0-12-409548-9.09071-0

Gooday AJ, Rothe N, Pearce RB (2013) New and poorly known benthic foraminifera (Protista, Rhizaria) inhabiting the shells of planktonic foraminifera on the bathyal Mid-Atlantic Ridge. Marine Biology Research 9:447–461. https://doi.org/10.1080/17451000.2012.750365

Gooday AJ, Holzmann M, Caulle C, Goineau A, Kamenskaya O, Weber AAT, Pawlowski J (2017a) Giant protists (xenophyophores, Foraminifera) are exceptionally diverse in parts of the abyssal eastern Pacific licensed for polymetallic nodule exploration. Biological conservation 207:106–116. https://doi.org/10.1016/J.BIOCON.2017.01.006

Gooday AJ, Schoenle A, Dolan JR, Arndt H (2020) Protist diversity and function in the dark ocean – Challenging the paradigms of deep-sea ecology with special emphasis on foraminiferans and naked protists. European journal of protistology 75:125721. https://doi.org/10.1016/J.EJOP.2020.125721

Gooday AJ, Lejzerowicz F, Goineau A, Holzmann M, Kamenskaya O, Kitazato H, Lim SC, Pawlowski J, Radziejewska T, Stachowska Z, Wawrzyniak-Wydrowska B (2021) The Biodiversity and Distribution of Abyssal Benthic Foraminifera and Their Possible Ecological Roles: A Synthesis Across the Clarion-Clipperton Zone. Frontiers in Marine Science 8:532. https://doi.org/10.3389/FMARS.2021.634726

Gordon ES, Goni MA (2003) Sources and distribution of terrigenous organic matter delivered by the Atchafalaya River to sediments in the northern Gulf of Mexico. Geochim Cosmochim Acta 67:2359–2375. https://doi.org/10.1016/S0016-7037(02)01412-6

Grassle JF, Morse-Porteous LS (1987) Macrofaunal colonization of disturbed deep-sea environments and the structure of deep-sea benthic communities. Deep Sea Research Part A Oceanographic Research Papers 34:1911–1950. https://doi.org/10.1016/0198-0149(87)90091-4

Greco M, Lejzerowicz F, Reo E, Caruso A, Maccotta A, Coccioni R, Pawlowski J, Frontalini F (2022) Environmental RNA outperforms eDNA metabarcoding in assessing impact of marine pollution: A chromium-spiked mesocosm test. Chemosphere 298:134239. https://doi.org/10.1016/J.CHEMOSPHERE.2022.134239

Grimm EC (1987) CONISS: a FORTRAN 77 program for stratigraphically constrained cluster analysis by the method of incremental sum of squares. Computers & Geosciences 13:13–35. https://doi.org/10.1016/0098-3004(87)90022-7

Gu Z, Gu L, Eils R, Schlesner M, Brors B (2014) circlize implements and enhances circular visualization in R. Bioinformatics 30:2811–2812. https://doi.org/10.1093/BIOINFORMATICS/BTU393

Guardiola M, Wangensteen OS, Taberlet P, Coissac E, Uriz MJ, Turon X (2016) Spatio-temporal monitoring of deep-sea communities using metabarcoding of sediment DNA and RNA. PeerJ 2016:e2807. https://doi.org/10.7717/PEERJ.2807

Guidi L, Legendre L, Reygondeau G, Uitz J, Stemmann L, Henson SA (2015) A new look at ocean carbon remineralization for estimating deepwater sequestration. Global Biogeochem Cycles 29:1044–1059. https://doi.org/10.1002/2014GB005063

Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, Darzi Y, Audic S, Berline L, Brum JR, Coelho LP, Espinoza JCI, Malviya S, Sunagawa S, Dimier C, Kandels-Lewis S, Picheral M, Poulain J, Searson S, Stemmann L, Not F, Hingamp P, Speich S, Follows M, Karp-Boss L, Boss E, Ogata H, Pesant S, Weissenbach J, Wincker P, Acinas SG, Bork P, de Vargas C, Iudicone D, Sullivan MB, Raes J, Karsenti E, Bowler C, Gorsky G (2016) Plankton networks driving carbon export in the oligotrophic ocean. Nature 532:465–470. https://doi.org/10.1038/nature16942

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, de Vargas C, Decelle J, del Campo J, Dolan JR, Dunthorn M, Edvardsen B, Holzmann M, Kooistra WHCF, Lara E, le Bescot N, Logares R, Mahé F, Massana R, Montresor M, Morard R,

Not F, Pawlowski J, Probert I, Sauvadet AL, Siano R, Stoeck T, Vaulot D, Zimmermann P, Christen R (2013a) The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic acids research 41:D597. https://doi.org/10.1093/NAR/GKS1160

Günther B, MARRE S, Defois C, Merzi T, Blanc P, PEYRET P, Arnaud-Haond S (2021) Capture by hybridization for full-length barcode-based eukaryotic and prokaryotic biodiversity inventories for deep sea ecosystems. Authorea Preprints. https://doi.org/10.22541/AU.161901961.16030002/V1

Habura A, Pawlowski J, Hanes SD, Bowser SS (2004) Unexpected Foraminiferal Diversity Revealed by Small-subunit rDNA Analysis of Antarctic Sediment. Journal of Eukaryotic Microbiology 51:173–179. https://doi.org/10.1111/J.1550-7408.2004.TB00542.X

Hadziavdic K, Lekang K, Lanzen A, Jonassen I, Thompson EM, Troedsson C (2014) Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers. PLoS One 9:87624. https://doi.org/10.1371/JOURNAL.PONE.0087624

Hammer DAT, Ryan PD, Hammer Ø, Harper DAT (2001) Past: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica 4:178.

Hayward BW, le Coze F, Vachard D, Gross O (2022) Foraminifera - The World Foraminifera Database. https://www.marinespecies.org/foraminifera/ . Accessed 23 Sep 2022

He X, Sutherland TF, Pawlowski J, Abbott CL (2019) Responses of foraminifera communities to aquaculture-derived organic enrichment as revealed by environmental DNA metabarcoding. Molecular Ecology 28:1138–1153. https://doi.org/10.1111/MEC.15007

Hebsgaard MB, Gilbert MTP, Arneborg J, Heyn P, Allentoft ME, Bunce M, Munch K, Schweger C, Willerslev E (2009) 'The Farm Beneath the Sand' – an archaeological case study on ancient 'dirt' DNA. Antiquity 83:430–444. https://doi.org/10.1017/S0003598X00098537

Henson SA, Sanders R, Madsen E, Morris PJ, le Moigne F, Quartly GD (2011) A reduced estimate of the strength of the ocean's biological carbon pump. Geophys Res Lett. https://doi.org/10.1029/2011GL046735

Hernández-Almeida I, Cortese G, Yu PS, Chen MT, Kucera M (2017) Environmental determinants of radiolarian assemblages in the western Pacific since the last deglaciation. Paleoceanography 32:830–847. https://doi.org/10.1002/2017PA003159

Hestetun JT, Bye-Ingebrigtsen E, Nilsson RH, Glover AG, Johansen PO, Dahlgren TG (2020) Significant taxon sampling gaps in DNA databases limit the operational use of marine

macrofauna metabarcoding. Marine Biodiversity 50:1–9. https://doi.org/10.1007/S12526-020-01093-5

Hestetun JT, Lanzén A, Dahlgren TG (2021) Grab what you can—an evaluation of spatial replication to decrease heterogeneity in sediment eDNA metabarcoding. PeerJ 9:e11619. https://doi.org/10.7717/PEERJ.11619

Hleap JS, Littlefair JE, Steinke D, Hebert PDN, Cristescu ME (2021) Assessment of current taxonomic assignment strategies for metabarcoding eukaryotes. Molecular Ecology Resources 21:2190–2203. https://doi.org/10.1111/1755-0998.13407

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018) UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular biology and evolution 35:518–522. https://doi.org/10.1093/MOLBEV/MSX281

Holman LE, de Bruyn M, Creer S, Carvalho G, Robidart J, Rius M (2019) Detection of introduced and resident marine species using environmental DNA metabarcoding of sediment and water. Scientific Reports 2019 9:1 9:1–10. https://doi.org/10.1038/s41598-019-47899-7

Holovachov O, Haenel Q, Bourlat SJ, Jondelius U (2017) Taxonomy assignment approach determines the efficiency of identification of OTUs in marine nematodes. Royal Society open science https://doi.org/10.1098/RSOS.170315

Holzmann M, Gooday AJ, Majewski W, Pawlowski J (2022) Molecular and morphological diversity of monothalamous foraminifera from South Georgia and the Falkland Islands: Description of four new species. European journal of protistology85:125909. https://doi.org/10.1016/J.EJOP.2022.125909

Horner RA, Greengrove CL, Davies-Vollum KS, Gawel JE, Postel JR, Cox AM (2011) Spatial distribution of benthic cysts of Alexandrium catenella in surface sediments of Puget Sound, Washington, USA. Harmful Algae 11:96–105. https://doi.org/10.1016/J.HAL.2011.08.004

Hoshino T, Doi H, Uramoto GI, Wörmer L, Adhikari RR, Xiao N, Morono Y, D'Hondt S, Hinrichs KU, Inagaki F (2020) Global diversity of microbial communities in marine sediment. Proceedings of the National Academy of Sciences 117:27587–27597. https://doi.org/10.1073/PNAS.1919139117

Hugenholtz P, Chuvochina M, Oren A, Parks DH, Soo RM (2021) Prokaryotic taxonomy and nomenclature in the age of big sequence data. ISME JOURNAL 15:1879. https://doi.org/10.1038/S41396-021-00941-X

Hussain RHM, Ghani MKA, Khan NA, Siddiqui R, Anuar TS (2022) Acanthamoeba species isolated from marine water in Malaysia exhibit distinct genotypes and variable physiological properties. Journal of Water and Health 20:54–67. https://doi.org/10.2166/WH.2021.128

Ibarbalz FM, Henry N, Brandão MC, Martini S, Busseni G, Byrne H, Coelho LP, Endo H, Gasol JM, Gregory AC, Mahé F, Rigonato J, Royo-Llonch M, Salazar G, Sanz-Sáez I, Scalco E, Soviadan D, Zayed AA, Zingone A, Labadie K, Ferland J, Marec C, Kandels S, Picheral M, Dimier C, Poulain J, Pisarev S, Carmichael M, Pesant S, Acinas SG, Babin M, Bork P, Boss E, Bowler C, Cochrane G, de Vargas C, Follows M, Gorsky G, Grimsley N, Guidi L, Hingamp P, Iudicone D, Jaillon O, Karp-Boss L, Karsenti E, Not F, Ogata H, Poulton N, Raes J, Sardet C, Speich S, Stemmann L, Sullivan MB, Sunagawa S, Wincker P, Pelletier E, Bopp L, Lombard F, Zinger L (2019) Global Trends in Marine Plankton Diversity across Kingdoms of Life. Cell 179:1084-1097.e21. https://doi.org/10.1016/j.cell.2019.10.008

Ingels J, Vanreusel A, Pape E, Pasotti F, Macheriotou L, Arbizu PM, Sørensen MV, Edgcomb VP, Sharma J, Sánchez N, Homoky WB, Woulds C, Leduc D, Gooday AJ, Pawlowski J, Dolan JR, Schratzberger M, Gollner S, Schoenle A, Arndt H, Zeppilli D (2020) Ecological variables for deep-ocean monitoring must include microbiota and meiofauna for effective conservation. Nature Ecology & Evolution 2020 5:1 5:27–29. https://doi.org/10.1038/s41559-020-01335-6

Iversen MH, Ploug H (2010) Ballast minerals and the sinking carbon flux in the ocean: carbon-specific respiration rates and sinking velocity of marine snow aggregates. Biogeosciences 7:2613–2624. https://doi.org/10.5194/bg-7-2613-2010

Jahan S, Strezov V (2019) Assessment of trace elements pollution in the sea ports of New South Wales (NSW), Australia using oysters as bioindicators. Scientific Reports 2019 9:1 9:1–10. https://doi.org/10.1038/s41598-018-38196-w

Jażdżewska AM, Mamos T (2019) High species richness of Northwest Pacific deep-sea amphipods revealed through DNA barcoding. Progress in Oceanography 178:102184. https://doi.org/10.1016/J.POCEAN.2019.102184

Jia W, Anslan S, Chen F, Cao X, Dong H, Dulias K, Gu Z, Heinecke L, Jiang H, Kruse S, Kang W, Li K, Liu S, Liu X, Liu Y, Ni J, Schwalb A, Stoof-Leichsenring KR, Shen W, Tian F, Wang J, Wang Y, Wang Y, Xu H, Yang X, Zhang D, Herzschuh U (2022) Sedimentary ancient DNA reveals past ecosystem and biodiversity changes on the Tibetan Plateau:

Overview and prospects. Quaternary Science Reviews 293:107703. https://doi.org/10.1016/J.QUASCIREV.2022.107703

Jones DOB, Simon-Lledó E, Amon DJ, Bett BJ, Caulle C, Clément L, Connelly DP, Dahlgren TG, Durden JM, Drazen JC, Felden J, Gates AR, Georgieva MN, Glover AG, Gooday AJ, Hollingsworth AL, Horton T, James RH, Jeffreys RM, Laguionie-Marchais C, Leitner AB, Lichtschlag A, Menendez A, Paterson GLJ, Peel K, Robert K, Schoening T, Shulga NA, Smith CR, Taboada S, Thurnherr AM, Wiklund H, Young CR, Huvenne VAI (2021) Environment, ecology, and potential effectiveness of an area protected from deep-sea mining (Clarion Clipperton Zone, abyssal Pacific). Progress in Oceanography 197:102653. https://doi.org/10.1016/J.POCEAN.2021.102653

Jonkers L, Hillebrand H, Kucera M (2019) Global change drives modern plankton communities away from the pre-industrial state. Nature 570:372–375. https://doi.org/10.1038/s41586-019-1230-3

Joos L, Beirinckx S, Haegeman A, Debode J, Vandecasteele B, Baeyen S, Goormachtig S, Clement L, de Tender C (2020) Daring to be differential: metabarcoding analysis of soil and plant-related microbial communities using amplicon sequence variants and operational taxonomical units. BMC Genomics. https://doi.org/10.1186/S12864-020-07126-4

Jorissen F, Nardelli MP, Almogi-Labin A, Barras C, Bergamin L, Bicchi E, el Kateb A, Ferraro L, McGann M, Morigi C, Romano E, Sabbatini A, Schweizer M, Spezzaferri S (2018) Developing Foram-AMBI for biomonitoring in the Mediterranean: Species assignments to ecological categories. Marine Micropaleontology140:33–45. https://doi.org/10.1016/J.MARMICRO.2017.12.006

Juggins S (2022) Rioja : Analysis of Quaternary Science Data v 1.0-5 https://cran.r-project.org/package=rioja

Kanbar HJ, Olajos F, Englund G, Holmboe M (2020) Geochemical identification of potential DNA-hotspots and DNA-infrared fingerprints in lake sediments. Applied Geochemistry 122:104728. https://doi.org/10.1016/J.APGEOCHEM.2020.104728

Kang W, Anslan S, Börner N, Schwarz A, Schmidt R, Künzel S, Rioual P, Echeverría-Galindo P, Vences M, Wang J, Schwalb A (2021) Diatom metabarcoding and microscopic analyses from sediment samples at Lake Nam Co, Tibet: The effect of sample-size and bioinformatics on the identified communities. Ecological Indicators 121:107070. https://doi.org/10.1016/J.ECOLIND.2020.107070

Kassambara A ggpubr: Publication Ready Plots - Articles - STHDA. Accessed 9 Oct 2022

Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in bioinformatics 20:1160–1166. https://doi.org/10.1093/bib/bbx108

Keck F, Millet L, Debroas D, Etienne D, Galop D, Rius D, Domaizon I (2020) Assessing the response of micro-eukaryotic diversity to the Great Acceleration using lake sedimentary DNA. Nature Communications 2020 11:1 11:1–8. https://doi.org/10.1038/s41467-020-17682-8

Kelly RP, Port JA, Yamahara KM, Martone RG, Lowell N, Thomsen PF, Mach ME, Bennett M, Prahler E, Caldwell MR, Crowder LB (2014) Harnessing DNA to improve environmental management. Science (1979) 344:1455–1456. https://doi.org/10.1126/SCIENCE.1251156

Kerrigan Z, D'Hondt S (2022) Patterns of Relative Bacterial Richness and Community Composition in Seawater and Marine Sediment Are Robust for Both Operational Taxonomic Units and Amplicon Sequence Variants. Frontiers in microbiology 13:96. https://doi.org/10.3389/FMICB.2022.796758

Keyes AA, McLaughlin JP, Barner AK, Dee LE (2021) An ecological network approach to predict ecosystem service vulnerability to species losses. Nature Communications 2021 12:1 12:1–11. https://doi.org/10.1038/s41467-021-21824-x

Kirkpatrick JB, Walsh EA, D'Hondt S (2016) Fossil DNA persistence and decay in marine sediment over hundred-thousand-year to million-year time scales. Geology 44:615–618. https://doi.org/10.1130/G37933.1

Kowalska JB, Mazurek R, Gąsiorek M, Zaleski T (2018) Pollution indices as useful tools for the comprehensive evaluation of the degree of soil contamination–A review. Environmental geochemistry and health 40:2395–2420. https://doi.org/10.1007/S10653-018-0106-Z

Krueger J, Foerster V, Trauth MH, Hofreiter M, Tiedemann R (2021) Exploring the Past Biosphere of Chew Bahir/Southern Ethiopia: Cross-Species Hybridization Capture of Ancient Sedimentary DNA from a Deep Drill Core. Front Earth Sci (Lausanne) 9:775. https://doi.org/10.3389/FEART.2021.683010

Kucera M (2007) Chapter Six Planktonic Foraminifera as Tracers of Past Oceanic Environments. Developments in Marine Geology 1:213–262. https://doi.org/10.1016/S1572-5480(07)01011-1

Kucera M, Darling KF (2002) Cryptic species of planktonic foraminifera: their effect on palaeoceanographic reconstructions. Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 360:695–718. https://doi.org/10.1098/rsta.2001.0962

Kvist S (2013) Barcoding in the dark?: A critical view of the sufficiency of zoological DNA barcoding databases and a plea for broader integration of taxonomic knowledge. Molecular phylogenetics and evolution 69:39–45. https://doi.org/10.1016/J.YMPEV.2013.05.012

Langer MR, Weinmann AE, Makled WA, Könen J, Gooday AJ (2022) New observations on test architecture and construction of Jullienella foetida Schlumberger, 1890, the largest shallow-water agglutinated foraminifer in modern oceans. PeerJ 10:e12884. https://doi.org/10.7717/PEERJ.12884

Lanzén A, Dahlgren TG, Bagi A, Hestetun JT (2021) Benthic eDNA metabarcoding provides accurate assessments of impact from oil extraction, and ecological insights. Ecological Indicators 130:1470–160. https://doi.org/10.1016/j.ecolind.2021.108064

Lara E, Moreira D, Vereshchaka A, López-García P (2009) Pan-oceanic distribution of new highly diverse clades of deep-sea diplonemids. Environmental Microbiology 11:47–55. https://doi.org/10.1111/J.1462-2920.2008.01737.X

Laroche O, Wood SA, Tremblay LA, Ellis JI, Lejzerowicz F, Pawlowski J, Lear G, Atalah J, Pochon X (2016) First evaluation of foraminiferal metabarcoding for monitoring environmental impact from an offshore oil drilling site. Marine environmental research 120:225–235. https://doi.org/10.1016/j.marenvres.2016.08.009

Laroche O, Wood SA, Tremblay LA, Ellis JI, Lear G, Pochon X (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, Kersten O, Smith CR, Goetze E (2020) From Sea Surface to Seafloor: A Benthic Allochthonous eDNA Survey for the Abyssal Ocean. Frontiers in Marine Science 7:682. https://doi.org/10.3389/FMARS.2020.00682

Lecroq B, Lejzerowicz F, Bachar D, Christen R, Esling P, Baerlocher L, Osterås M, Farinelli L, Pawlowski J (2011) Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. Proceedings of

the National Academy of Sciences 108:13177–13182. https://doi.org/10.1073/pnas.1018426108

Lejzerowicz F, Pawlowski J, Fraissinet-Tachet L, Marmeisse R (2010) Molecular evidence for widespread occurrence of Foraminifera in soils. Environmental Microbiology 12:2518–2526. https://doi.org/10.1111/J.1462-2920.2010.02225.X

Lejzerowicz F, Esling P, Majewski W, Szczucinski W, Decelle J, Obadia C, Arbizu PM, Pawlowski J (2013) Ancient DNA complements microfossil record in deep-sea subsurface sediments. Biology letters 9:20130283–20130283. https://doi.org/10.1098/rsbl.2013.0283

Lejzerowicz F, Esling P, Pawlowski J (2014a) 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, Voltski I, Pawlowski J (2015a) Foraminifera of the Kuril-Kamchatka Trench area: The prospects of molecular study. Deep Sea Res 2 Top Stud Oceanogr 111:19–25. https://doi.org/10.1016/j.dsr2.2014.10.003

Lejzerowicz F, Esling P, Pillet L, Wilding TA, Black KD, Pawlowski J (2015b) High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. Scientific Reports 2015 5:1 5:1–10. https://doi.org/10.1038/srep13932

Lejzerowicz F, Gooday AJ, Barrenechea Angeles I, Cordier T, Morard R, Apothéloz-Perret-Gentil L, Lins L, Menot L, Brandt A, Levin LA, Martinez Arbizu P, Smith CR, Pawlowski J (2021) Eukaryotic Biodiversity and Spatial Patterns in the Clarion-Clipperton Zone and Other Abyssal Regions: Insights From Sediment DNA and RNA Metabarcoding. Frontiers in Marine Science 8:536. https://doi.org/10.3389/FMARS.2021.671033

Leng W, von Dobeneck T, Bergmann F, Just J, Mulitza S, Chiessi CM, St-Onge G, Piper DJW (2018) Sedimentary and rock magnetic signatures and event scenarios of deglacial outburst floods from the Laurentian Channel Ice Stream. Quaternary Science Reviews 186:27–46. https://doi.org/10.1016/j.quascirev.2018.01.016

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

Leray M, Knowlton N (2016) Censusing marine eukaryotic diversity in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences. https://doi.org/10.1098/RSTB.2015.0331

Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, Boehm JT, Machida RJ (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. Frontiers in Zoologie 10:1–14. https://doi.org/10.1186/1742-9994-10-34

Leray M, Knowlton N, Machida RJ (2022) MIDORI2: A collection of quality controlled, preformatted, and regularly updated reference databases for taxonomic assignment of eukaryotic mitochondrial sequences. Environmental DNA 4:894–907. https://doi.org/10.1002/EDN3.303

Levin LA, Bris NL (2015) The deep ocean under climate change. Science (1979) 350:766–768. https://doi.org/10.1126/SCIENCE.AAD0126

Levin LA, Amon DJ, Lily H (2020) Challenges to the sustainability of deep-seabed mining. Nat Sustain 3:784–794. https://doi.org/10.1038/s41893-020-0558-x

Lewis SL, Maslin MA (2015) Defining the Anthropocene. Nature 2015 519:7542 519:171–180. https://doi.org/10.1038/nature14258

Lie AAY, Liu Z, Hu SK, Jones AC, Kim DY, Countway PD, Amaral-Zettler LA, Cary SC, Sherr EB, Sherr BF, Gast RJ, Caron DA (2014) Investigating microbial eukaryotic diversity from a global census: Insights from a comparison of pyrotag and full-length sequences of 18S rRNA genes. Appl Environmental Microbiology 80:4363–4373. https://doi.org/10.1128/AEM.00057-14

Lindh M v., Maillot BM, Shulse CN, Gooday AJ, Amon DJ, Smith CR, Church MJ (2017) From the surface to the deep-sea: Bacterial distributions across polymetallic nodule fields in the clarion-clipperton zone of the Pacific Ocean. Frontiers in microbiology 8:1696. https://doi.org/10.3389/FMICB.2017.01696

Lins L, Zeppilli D, Menot L, Michel LN, Bonifácio P, Brandt M, Pape E, Rossel S, Uhlenkott K, Macheriotou L, Bezerra TN, Sánchez N, Alfaro-Lucas JM, Martínez Arbizu P, Kaiser S, Murakami C, Vanreusel A (2021) Toward a reliable assessment of potential ecological impacts of deep-sea polymetallic nodule mining on abyssal infauna. Limnology and oceanography Methods 19:626–650. https://doi.org/10.1002/LOM3.10448

Locarnini RA, Mishonov A V., Antonov JI, Boyer TP, Garcia HE, Baranova OK, Zweng MM, Paver CR, Reagan JR, Johnson DR, Hamilton M, Seidov D (2013) World Ocean Atlas 2013, Volume 1: Temperature. NOAA Atlas NESDIS 73 40 pp. http://doi.org/10.7289/V55X26VD

Logares R, Audic S, Bass D, Bittner L, Boutte C, Christen R, Claverie JM, Decelle J, Dolan JR, Dunthorn M, Edvardsen B, Gobet A, Kooistra WHCF, Mahé F, Not F, Ogata H, Pawlowski J, Pernice MC, Romac S, Shalchian-Tabrizi K, Simon N, Stoeck T, Santini S, Siano R, Wincker P, Zingone A, Richards TA, de Vargas C, Massana R (2014) Patterns of rare and abundant marine microbial eukaryotes. Current Biololgy 24:813–821. https://doi.org/10.1016/J.CUB.2014.02.050

Long ER, MacDonald DD, Severn CG, Hong CB (2000) Classifying probabilities of acute toxicity in marine sediments with empirically derived sediment quality guidelines. Environmental Toxicology and Chemistry 19:2598–2601. https://doi.org/10.1002/ETC.5620191028

Lorenz MG, Wackernagel W (1987) Adsorption of DNA to sand and variable degradation rates of adsorbed DNA. Appl Environmental Microbiology 53:2948–2952. https://doi.org/10.1128/AEM.53.12.2948-2952.1987

Loubere P (1989) Bioturbation and sedimentation rate control of benthic microfossil taxon abundances in surface sediments: A theoretical approach to the analysis of species microhabitats. Marine Micropaleontology 14:317–325. https://doi.org/10.1016/0377-8398(89)90016-9

Lu GY, Wong DW (2008) An adaptive inverse-distance weighting spatial interpolation technique. Comput Geosci 34:1044–1055. https://doi.org/10.1016/J.CAGEO.2007.07.010

Lutz MJ, Caldeira K, Dunbar RB, Behrenfeld MJ (2007) Seasonal rhythms of net primary production and particulate organic carbon flux to depth describe the efficiency of biological pump in the global ocean. Journal of Geophysical Research 112:10011. https://doi.org/10.1029/2006JC003706

Macher JN, Wideman JG, Girard EB, Langerak A, Duijm E, Jompa J, Sadekov A, Vos R, Wissels R, Renema W (2021) First report of mitochondrial COI in foraminifera and implications for DNA barcoding. Scientific Reports 2021 11:1 11:1–9. https://doi.org/10.1038/s41598-021-01589-5

Macher J-N, Bloska DM, Holzmann M, Girard EB, Pawlowski J, Renema W (2022) Mitochondrial cytochrome c oxidase subunit I (COI) metabarcoding of Foraminifera communities using taxon-specific primers. PeerJ 10:e13952. https://doi.org/10.7717/PEERJ.13952

Maggiori C, Raymond-Bouchard I, Brennan L, Touchette D, Whyte L (2021) MinION sequencing from sea ice cryoconites leads to de novo genome reconstruction from

metagenomes. Scientific Reports 2021 11:1 11:1–16. https://doi.org/10.1038/s41598-021-00026-x

Mahé F, Rognes T, Quince C, de Vargas C, Dunthorn M (2015) Swarmv2: Highly-scalable and high-resolution amplicon clustering. PeerJ 2015:e1420. https://doi.org/10.7717/peerj.1420

Marbà N, Díaz-Almela E, Duarte CM (2014) Mediterranean seagrass (Posidonia oceanica) loss between 1842 and 2009. Biological Conservation 176:183–190. https://doi.org/10.1016/J.BIOCON.2014.05.024

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. https://doi.org/10.14806/EJ.17.1.200

Martins FMS, Porto M, Feio MJ, Egeter B, Bonin A, Serra SRQ, Taberlet P, Beja P (2021) Modelling technical and biological biases in macroinvertebrate community assessment from bulk preservative using multiple metabarcoding markers. Molecular Ecology 30:3221–3238. https://doi.org/10.1111/MEC.15620

Mauffrey F, Cordier T, Apothéloz-Perret-Gentil L, Cermakova K, Merzi T, Delefosse M, Blanc P, Pawlowski J (2021) Benthic monitoring of oil and gas offshore platforms in the North Sea using environmental DNA metabarcoding. Molecular Ecology 30:3007–3022. https://doi.org/10.1111/MEC.15698

McClain CR, Rex MA (2015) Toward a Conceptual Understanding of β-Diversity in the Deep-Sea Benthos. Annual Review of Ecology, Evolution, and Systematics 46:623–642. https://doi.org/10.1146/annurev-ecolsys-120213-091640

McClain CR, Nekola JC, Kuhnz L, Barry JP (2011) Local-scale faunal turnover on the deep Pacific seafloor. Marine Ecology Progress Series 422:193–200. https://doi.org/10.3354/meps08924

McKnight DT, Huerlimann R, Bower DS, Schwarzkopf L, Alford RA, Zenger KR (2019) microDecon: A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies. Environmental DNA 1:14–25. https://doi.org/10.1002/EDN3.11

McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One 8:e61217. https://doi.org/10.1371/JOURNAL.PONE.0061217

Meilland J, Howa H, Hulot V, Demangel I, Salaün J, Garlan T (2019) Population dynamics of modern planktonic foraminifera in the western Barents Sea. Biogeosciences Discussions. https://doi.org/https://doi.org/10.5194/bg-2019-429

Mengerink KJ, van Dover CL, Ardron J, Baker M, Escobar-Briones E, Gjerde K, Koslow JA, Ramirez-Llodra E, Lara-Lopez A, Squires D, Sutton T, Sweetman AK, Levin LA (2014) A Call for Deep-Ocean Stewardship. Science (1979) 344:696–698. https://doi.org/10.1126/SCIENCE.1251458

Meyers PA (1994) Preservation of elemental and isotopic source identification of sedimentary organic matter. Chem Geol 114:289–302. https://doi.org/10.1016/0009-2541(94)90059-0

Milker Y, Schmiedl G (2012) A taxonomic guide to modern benthic shelf foraminifera of the western Mediterranean Sea. Palaeontologia Electronica 15:1–134. https://doi.org/10.26879/271

Millán-Aguiñaga N, Soldatou S, Brozio S, Munnoch JT, Howe J, Hoskisson PA, Duncan KR (2019) Awakening ancient polar actinobacteria: Diversity, evolution and specialized metabolite potential. Microbiology (United Kingdom) 165:1169–1180. https://doi.org/10.1099/MIC.0.000845

Moguel B, Pérez L, Alcaraz LD, Blaz J, Caballero M, Muñoz-Velasco I, Becerra A, Laclette JP, Ortega-Guerrero B, Romero-Oliva CS, Herrera-Estrella L, Lozano-García S (2021) Holocene life and microbiome profiling in ancient tropical Lake Chalco, Mexico. Scientific Reports 2021 11:1 11:1–15. https://doi.org/10.1038/s41598-021-92981-8

Monchamp ME, Spaak P, Domaizon I, Dubois N, Bouffard D, Pomati F (2017) Homogenization of lake cyanobacterial communities over a century of climate change and eutrophication. Nature Ecology & Evolution 2017 2:2 2:317–324. https://doi.org/10.1038/s41559-017-0407-0

Moodley L (1990) "Squatter" behaviour in soft-shelled foraminifera. Marine Micropaleontology16:149–153. https://doi.org/10.1016/0377-8398(90)90033-I

Morard R, Darling KF, Mahé F, Audic S, Ujiié Y, Weiner AKM, André A, Seears H a., Wade CM, Quillévéré F, Douady CJ, Escarguel G, de Garidel-Thoron T, Siccha M, Kucera M, de Vargas C (2015) PFR^2 : a curated database of planktonic foraminifera 18S ribosomal DNA as a resource for studies of plankton ecology, biogeography and evolution. Molecular Ecology Resources 15:1472–1485. https://doi.org/10.1111/1755-0998.12410

Morard R, Reinelt M, Chiessi CM, Groeneveld J, Kucera M (2016) Tracing shifts of oceanic fronts using the cryptic diversity of the planktonic foraminifera Globorotalia inflata. Paleoceanography 31:1193–1205. https://doi.org/10.1002/2016PA002977

Morard R, Lejzerowicz F, Darling KF, Lecroq-Bennet B, Pedersen MW, Orlando L, Pawlowski J, Mulitza S, de Vargas C, Kucera M (2017) Planktonic foraminifera-derived

environmental DNA extracted from abyssal sediments preserves patterns of plankton macroecology. Biogeosciences 14:2741–2754. https://doi.org/10.5194/bg-14-2741-2017

Morard R, Mahé F, Romac S, Poulain J, Kucera M, Vargas C De (2018) Surface ocean metabarcoding confirms limited diversity in planktonic foraminifera but reveals unknown hyper-abundant lineages. Scientific Reports 8:2539:1–10. https://doi.org/10.1038/s41598-018-20833-z

Morard R, Vollmar ID NM, Greco MI, Kucera M (2019) Unassigned diversity of planktonic foraminifera from environmental sequencing revealed as known but neglected species. PLOS One 14.3. https://doi.org/10.1371/journal.pone.0213936

More KD, Orsi WD, Galy V, Giosan L, He L, Grice K, Coolen MJL (2018) A 43 kyr record of protist communities and their response to oxygen minimum zone variability in the Northeastern Arabian Sea. Earth and Planetary Science Letters496:248–256. https://doi.org/10.1016/J.EPSL.2018.05.045

More KD, Giosan L, Grice K, Coolen MJL (2019) Holocene paleodepositional changes reflected in the sedimentary microbiome of the Black Sea. Geobiology 17:436–448. https://doi.org/10.1111/GBI.12338

Morgan-Smith D, Clouse MA, Herndl GJ, Bochdansky AB (2013) Diversity and distribution of microbial eukaryotes in the deep tropical and subtropical North Atlantic Ocean. Deep Sea Research Part I: Oceanographic Research Papers 78:58–69. https://doi.org/10.1016/J.DSR.2013.04.010

Morlock MA, Vogel H, Russell JM, Anselmetti FS, Bijaksana S (2021) Quaternary environmental changes in tropical Lake Towuti, Indonesia, inferred from end-member modelling of X-ray fluorescence core-scanning data. Journal of Quaternary Science 36:1040–1051. https://doi.org/10.1002/JQS.3338

Morroni L, d'Errico G, Sacchi M, Molisso F, Armiento G, Chiavarini S, Rimauro J, Guida M, Siciliano A, Ceparano M, Aliberti F, Tosti E, Gallo A, Libralato G, Patti FP, Gorbi S, Fattorini D, Nardi A, di Carlo M, Mezzelani M, Benedetti M, Pellegrini D, Musco L, Danovaro R, Dell'Anno A, Regoli F (2020) Integrated characterization and risk management of marine sediments: The case study of the industrialized Bagnoli area (Naples, Italy). Marine environmental research 160:104984. https://doi.org/10.1016/J.MARENVRES.2020.104984

Mulitza S, Bergmann F, Brück L, Chiessi CM, Govin A, Klann M, Kuhnert H, Lübben B, Max L, Meyer V, Morard R, Müller V, Patton G, Paul A, Poirier A, Riesen P, Schade T, Stöber

U, Völker GS, Völpel R, von Dobeneck T (2015) Cruise No. MSM39 - June 07-June 25, 2014 - St. John's (Canada) - St. John's (Canada).

Müller G Schwermetalle in den Sedimenten des Rheins-Veranderungen seit 1971. Umschau, 79, 778-783.

Murali A, Bhargava A, Wright ES (2018) IDTAXA: A novel approach for accurate taxonomic classification of microbiome sequences. Microbiome 6:1–14. https://doi.org/10.1186/S40168-018-0521-5

Murchie TJ, Kuch M, Duggan AT, Ledger ML, Roche K, Klunk J, Karpinski E, Hackenberger D, Sadoway T, MacPhee R, Froese D, Poinar H (2021) Optimizing extraction and targeted capture of ancient environmental DNA for reconstructing past environments using the PalaeoChip Arctic-1.0 bait-set. Quaternary Research 99:305–328. https://doi.org/10.1017/QUA.2020.59

Murray JW (2007) Biodiversity of living benthic foraminifera : How many species are there ? Marine Micropaleontology 64:163–176. https://doi.org/10.1016/j.marmicro.2007.04.002

Nascimento FJA, Lallias D, Bik HM, Creer S (2018) Sample size effects on the assessment of eukaryotic diversity and community structure in aquatic sediments using high-throughput sequencing. Scientific Reports 2018 8:1 8:1–12. https://doi.org/10.1038/s41598-018-30179-1

Newmaster SG, Fazekas AJ, Ragupathy S (2006) DNA barcoding in land plants: Evaluation of rbcL in a multigene tiered approach. Canadian Journal of Botany 84:335–341. https://doi.org/10.1139/B06-047

Nguyen, Ngoc-Loi, et al. "Metabarcoding reveals high diversity of benthic foraminifera linked to water masses circulation at coastal Svalbard." Geobiology (2022). https://doi.org/10.1111/gbi.12530

Nota K, Orlando L, Marchesini A, Girardi M, Vernesi C, Parducci L (2021) Shotgun barcode baiting: capturing barcoding genes from environmental samples for species identification. EGU21. https://doi.org/10.5194/EGUSPHERE-EGU21-13501

Nota K, Klaminder J, Milesi P, Bindler R, Nobile A, van Steijn T, Bertilsson S, Svensson B, Hirota SK, Matsuo A, Gunnarsson U, Seppä H, Väliranta MM, Wohlfarth B, Suyama Y, Parducci L (2022) Norway spruce postglacial recolonization of Fennoscandia. Nature Communications 2022 13:1 13:1–9. https://doi.org/10.1038/s41467-022-28976-4

Obiol A, Giner CR, Sánchez P, Duarte CM, Acinas SG, Massana R (2020) A metagenomic assessment of microbial eukaryotic diversity in the global ocean. Molecular Ecology Resources 20:718–731. https://doi.org/10.1111/1755-0998.13147

Obura D, Gudka M, Samoilys M, Osuka K, Mbugua J, Keith DA, Porter S, Roche R, van Hooidonk R, Ahamada S, Araman A, Karisa J, Komakoma J, Madi M, Ravinia I, Razafindrainibe H, Yahya S, Zivane F (2021) Vulnerability to collapse of coral reef ecosystems in the Western Indian Ocean. Nature Sustainability 2021 5:2 5:104–113. https://doi.org/10.1038/s41893-021-00817-0

Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, Suggests M (2007) The vegan package. Community ecology package 10:631–637.

Parada AE, Needham DM, Fuhrman JA (2016) Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environmental Microbiology 18:1403–1414. https://doi.org/10.1111/1462-2920.13023

Paradis E, Schliep K (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526–528. https://doi.org/10.1093/BIOINFORMATICS/BTY633

Parducci L, Jørgensen T, Tollefsrud MM, Elverland E, Alm T, Fontana SL, Bennett KD, Haile J, Matetovici I, Suyama Y, others (2012) Glacial survival of boreal trees in northern Scandinavia. Science (1979) 335:1083–1086. https://doi.org/10.1126/science.1216043

Parducci L, Väliranta M, Sakari Salonen J, Ronkainen T, Matetovici I, Fontana SL, Eskola T, Sarala P, Suyama Y (2015) Proxy comparison in ancient peat sediments: pollen, macrofossil and plant DNA. Philosophical Transactions of the Royal Society B: Biological Sciences. https://doi.org/10.1098/RSTB.2013.0382

Parducci L, Bennett KD, Ficetola GF, Alsos IG, Suyama Y, Wood JR, Pedersen MW (2017) Ancient plant DNA in lake sediments. New Phytologist 214:924–942. https://doi.org/10.1111/nph.14470

Parent B, Hyams-Kaphzan O, Barras C, Lubinevsky H, Jorissen F (2021) Testing foraminiferal environmental quality indices along a well-defined organic matter gradient in the Eastern Mediterranean. Ecological Indicators 125:107498. https://doi.org/10.1016/J.ECOLIND.2021.107498

Passaro S, Gherardi S, Romano E, Ausili A, Sesta G, Pierfranceschi G, Tamburrino S, Sprovieri M (2020) Coupled geophysics and geochemistry to record recent coastal changes of

contaminated sites of the Bagnoli industrial area, Southern Italy. Estuarine, Coastal and Shelf Science 246:107036. https://doi.org/10.1016/J.ECSS.2020.107036

Patrício J, Adão H, Neto JM, Alves AS, Traunspurger W, Marques JC (2012) Do nematode and macrofauna assemblages provide similar ecological assessment information? Ecological Indicators 14:124–137. https://doi.org/10.1016/J.ECOLIND.2011.06.027

Paulson JN, Colin Stine O, Bravo HC, Pop M (2013a) Differential abundance analysis for microbial marker-gene surveys. Nature Methods 10:1200–1202. https://doi.org/10.1038/NMETH.2658

Paulson JN, Pop M, Bravo HC (2013b) metagenomeSeq: Statistical analysis for sparse high-throughput sequencing.

Paulus E (2021) Shedding Light on Deep-Sea Biodiversity—A Highly Vulnerable Habitat in the Face of Anthropogenic Change. Frontiers in Marine Science 8:281. https://doi.org/10.3389/FMARS.2021.667048

Pawlowska J, Lejzerowicz F, Esling P, Szczucinski W, Zajkaczkowski M, Pawlowski J (2014) Ancient DNA sheds new light on the Svalbard foraminiferal fossil record of the last millennium. Geobiology 12:277–288. https://doi.org/10.1111/gbi.12087

Pawlowska J, Zajaczkowski M, Lacka M, Lejzerowicz F, Esling P, Pawlowski J (2016) Palaeoceanographic changes in Hornsund Fjord (Spitsbergen, Svalbard) over the last millennium: New insights from ancient DNA. Climate of the Past 12:1459–1472. https://doi.org/10.5194/cp-12-1459-2016

Pawłowska J, Wollenburg JE, Zajączkowski M, Pawlowski J (2020) Planktonic foraminifera genomic variations reflect paleoceanographic changes in the Arctic: evidence from sedimentary ancient DNA. Scientific Reports 2020 10:1 10:1–10. https://doi.org/10.1038/s41598-020-72146-9

Pawlowski J, Holzmann M (2014) A plea for DNA barcoding of Foraminifera. Article in The Journal of Foraminiferal Research. https://doi.org/10.2113/gsjfr.44.1.62

Pawlowski J, Lecroq B (2010) Short rDNA barcodes for species identification in foraminifera. Journal of Eukaryotic Microbiology 57:197–205. https://doi.org/10.1111/j.1550-7408.2009.00468.x

Pawlowski J, Lee JJ (1991) Taxonomic notes on some tiny, shallow water foraminifera from the northern Gulf of Elat (Red Sea). Micropaleontology 37:149–162. https://doi.org/10.2307/1485555

Pawlowski J, Bolivar I, Fahrni F, de Vargas C, Gouy M, Zaninetti L (1997) Extreme differences in rates of molecular evolution of foraminifera revealed by comparison of ribosomal DNA sequences and the fossil record. Molecular biology and evolution 14:498–505. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A025786

Pawlowski J, Holzmann M, Berney C, Fahrni J, Gooday AJ, Cedhagen T, Habura A, Bowser SS (2003) The evolution of early Foraminifera. Proceedings of the National Academy of Sciences 100:11494–11498. https://doi.org/10.1073/PNAS.2035132100

Pawlowski J, Fahrni J, Lecroq B, Longet D, Cornelius N, Excoffier L, Cedhagen T, Gooday AJ (2007) Bipolar gene flow in deep-sea benthic foraminifera. Molecular Ecology 16:4089–4096. https://doi.org/10.1111/J.1365-294X.2007.03465.X

Pawlowski J, Christen R, Lecroq B, Bachar D, Shahbazkia HR, Amaral-Zettler L, Guillou L (2011) Eukaryotic Richness in the Abyss: Insights from Pyrotag Sequencing. PLoS One 6:e18169. https://doi.org/10.1371/JOURNAL.PONE.0018169

Pawlowski J, Fontaine D, da Silva AA, Guiard J (2011) Novel lineages of Southern Ocean deep-sea foraminifera revealed by environmental DNA sequencing. Deep Sea Research Part II: Topical Studies in Oceanography 58:1996–2003. https://doi.org/10.1016/J.DSR2.2011.01.009

Pawlowski J, Holzmann M, Tyszka J (2013) New supraordinal classification of Foraminifera: Molecules meet morphology. Marine Micropaleontology100:1–10. https://doi.org/10.1016/j.marmicro.2013.04.002

Pawlowski J, Esling P, Lejzerowicz F, Cedhagen T, Wilding TA (2014a) Environmental monitoring through protist next-generation sequencing metabarcoding: assessing the impact of fish farming on benthic foraminifera communities. Molecular Ecology Resources 14:1129–1140. https://doi.org/10.1111/1755-0998.12261

Pawlowski J, Lejzerowicz F, Esling P (2014b) Next-generation environmental diversity surveys of foraminifera: Preparing the future. Biological Bulletin 227:93–106. https://doi.org/10.1086/BBLV227N2P93

Pawlowski J, Kelly-Quinn M, Altermatt F, Apothéloz-Perret-Gentil L, Beja P, Boggero A, Borja A, Bouchez A, Cordier T, Domaizon I, Feio MJ, Filipe AF, Fornaroli R, Graf W, Herder J, van der Hoorn B, Iwan Jones J, Sagova-Mareckova M, Moritz C, Barquín J, Piggott JJ, Pinna M, Rimet F, Rinkevich B, Sousa-Santos C, Specchia V, Trobajo R, Vasselon V, Vitecek S, Zimmerman J, Weigand A, Leese F, Kahlert M (2018) The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological

assessment of aquatic ecosystems. Science of The Total Environment 637–638:1295–1310. https://doi.org/10.1016/J.SCITOTENV.2018.05.002

Pawlowski J, Apothéloz-Perret-Gentil L, Altermatt F (2020) Environmental DNA: What's behind the term? Clarifying the terminology and recommendations for its future use in biomonitoring. Molecular Ecology 29:4258–4264. https://doi.org/10.1111/MEC.15643

Pawlowski J, Bruce K, Panksep K, Aguirre FI, Amalfitano S, Apothéloz-Perret-Gentil L, Baussant T, Bouchez A, Carugati L, Cermakova K, Cordier T, Corinaldesi C, Costa FO, Danovaro R, Dell'Anno A, Duarte S, Eisendle U, Ferrari BJD, Frontalini F, Frühe L, Haegerbaeumer A, Kisand V, Krolicka A, Lanzén A, Leese F, Lejzerowicz F, Lyautey E, Maček I, Sagova-Marečková M, Pearman JK, Pochon X, Stoeck T, Vivien R, Weigand A, Fazi S (2021a) Environmental DNA metabarcoding for benthic monitoring: A review of sediment sampling and DNA extraction methods. Science of The Total Environment 151783. https://doi.org/10.1016/J.SCITOTENV.2021.151783

Pawlowski J, Apothéloz-Perret-Gentil L, Altermatt F (2021b) Environmental versus extra-organismal DNA. Molecular Ecology 30:4606–4607. https://doi.org/10.1111/MEC.16144

Paz G, Rinkevich B (2021) Gap analysis of DNA barcoding in ERMS reference libraries for ascidians and cnidarians. Environmental Sciences Europe 33:1–8. https://doi.org/10.1186/S12302-020-00449-9

Pearson PN, Penny L (2021) Coiling directions in the planktonic foraminifer Pulleniatina: A complex eco-evolutionary dynamic spanning millions of years. PLoS One 16:e0249113. https://doi.org/10.1371/JOURNAL.PONE.0249113

Pedersen MW, Overballe-Petersen S, Ermini L, der Sarkissian C, Haile J, Hellstrom M, Spens J, Thomsen PF, Bohmann K, Cappellini E, Schnell IB, Wales NA, Carøe C, Campos PF, Schmidt AMZ, Gilbert MTP, Hansen AJ, Orlando L, Willerslev E (2015) Ancient and modern environmental DNA. Philosophical Transactions of the Royal Society B: Biological Sciences 370:20130383. https://doi.org/10.1098/RSTB.2013.0383

Pedersen MW, Ruter A, Schweger C, Friebe H, Staff RA, Kjeldsen KK, Mendoza MLZ, Beaudoin AB, Zutter C, Larsen NK, Potter BA, Nielsen R, Rainville RA, Orlando L, Meltzer DJ, Kjær KH, Willerslev E (2016) Postglacial viability and colonization in North America's ice-free corridor. Nature 2016 537:7618 537:45–49. https://doi.org/10.1038/nature19085

Penton CR, Gupta VVSR, Yu J, Tiedje JM (2016) Size matters: Assessing optimum soil sample size for fungal and bacterial community structure analyses using high throughput

sequencing of rRNA gene amplicons. Frontiers in microbiology 7:824. https://doi.org/10.3389/FMICB.2016.00824

Pernice MC, Giner CR, Logares R, Perera-Bel J, Acinas SG, Duarte CM, Gasol JM, Massana R (2015) Large variability of bathypelagic microbial eukaryotic communities across the world's oceans. The ISME Journal 2016 10:4 10:945–958. https://doi.org/10.1038/ismej.2015.170

Phillippy AM, Mason JA, Ayanbule K, Sommer DD, Taviani E, Huq A, Colwell RR, Knight IT, Salzberg SL (2007) Comprehensive DNA Signature Discovery and Validation. PLoS Computational Biology 3:e98. https://doi.org/10.1371/JOURNAL.PCBI.0030098

Pietramellara G, Ascher J, Borgogni F, Ceccherini MT, Guerri G, Nannipieri P (2009) Extracellular DNA in soil and sediment: Fate and ecological relevance. Biology and Fertility of Soils 45:219–235. https://doi.org/10.1007/S00374-008-0345-8

Pillet L, Fontaine D, Pawlowski J (2012) Intra-Genomic Ribosomal RNA Polymorphism and Morphological Variation in Elphidium macellum Suggests Inter-Specific Hybridization in Foraminifera. PLoS One 7:e32373. https://doi.org/10.1371/JOURNAL.PONE.0032373

Pochon X, Wood SA, Keeley NB, Lejzerowicz F, Esling P, Drew J, Pawlowski J (2015) Accurate assessment of the impact of salmon farming on benthic sediment enrichment using foraminiferal metabarcoding. Marine Pollution Bulletin 100:370–382. https://doi.org/10.1016/J.MARPOLBUL.2015.08.022

Poff KE, Leu AO, Eppley JM, Karl DM, DeLong EF (2021) Microbial dynamics of elevated carbon flux in the open ocean's abyss. Proceedings of the National Academy of Sciences 118:e2018269118. https://doi.org/10.1073/PNAS.2018269118

Porter TM, Hajibabaei M (2020) Putting COI Metabarcoding in Context: The Utility of Exact Sequence Variants (ESVs) in Biodiversity Analysis. Frontiers in Ecology and Evolution 8:248. https://doi.org/10.3389/FEVO.2020.00248

Porzio L, Grech D, Buia MC (2020) Long-term changes (1800–2019) in marine vegetational habitats: Insights from a historic industrialised coastal area. Marine environmental research 161:105003. https://doi.org/10.1016/J.MARENVRES.2020.105003

Prentice ML, Hope GS, Maryunani K, Peterson JA (2005) An evaluation of snowline data across New Guinea during the last major glaciation, and area-based glacier snowlines in the Mt. Jaya region of Papua, Indonesia, during the Last Glacial Maximum. Quaternary International 138–139:93–117. https://doi.org/10.1016/J.QUAINT.2005.02.008

Preston CM, Durkin CA, Yamahara KM (2020) DNA metabarcoding reveals organisms contributing to particulate matter flux to abyssal depths in the North East Pacific ocean. Deep Sea Research Part II: Topical Studies in Oceanography 173:104708. https://doi.org/10.1016/J.DSR2.2019.104708

Pusceddu A, Bianchelli S, Martín J, Puig P, Palanques A, Masqué P, Danovaro R (2014) Chronic and intensive bottom trawling impairs deep-sea biodiversity and ecosystem functioning. Proceedings of the National Academy of Sciences 111:8861–8866. https://doi.org/10.1073/PNAS.1405454111

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research . https://doi.org/10.1093/NAR/GKS1219

R Development Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing. http://www.r-project.org/

Ramirez-Llodra E, Brandt A, Danovaro R, de Mol B, Escobar E, German CR, Levin LA, Martinez Arbizu P, Menot L, Buhl-Mortensen P, Narayanaswamy BE, Smith CR, Tittensor DP, Tyler PA, Vanreusel A, Vecchione M (2010) Deep, diverse and definitely different: unique attributes of the world's largest ecosystem. Biogeosciences 7:2851–2899. https://doi.org/10.5194/bg-7-2851-2010

Ratnasingham S, Hebert PDN (2007) bold: The Barcode of Life Data System (http://www.barcodinglife.org). Molecular Ecology Notes 7:355. https://doi.org/10.1111/J.1471-8286.2007.01678.X

Rex MA, Etter RJ (2010) Deep-sea biodiversity: pattern and scale. Choice Reviews Online 47:47-6850-47–6850. https://doi.org/10.5860/choice.47-6850

Rijal DP, Heintzman PD, Lammers Y, Yoccoz NG, Lorberau KE, Pitelkova I, Goslar T, Murguzur FJA, Salonen JS, Helmens KF, Bakke J, Edwards ME, Alm T, Bråthen KA, Brown AG, Alsos IG (2021) Sedimentary ancient DNA shows terrestrial plant richness continuously increased over the Holocene in northern Fennoscandia. Science Advances. https://doi.org/10.1126/SCIADV.ABF9557

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu WT, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T (2013) Insights into the phylogeny and coding potential of microbial dark matter. Nature 499:431–437. https://doi.org/10.1038/nature12352

Rodriguez-Ezpeleta N, Morissette O, Bean CW, Manu S, Banerjee P, Lacoursière-Roussel A, Beng KC, Alter SE, Roger F, Holman LE, Stewart KA, Monaghan MT, Mauvisseau Q, Mirimin L, Wangensteen OS, Antognazza CM, Helyar SJ, de Boer H, Monchamp ME, Nijland R, Abbott CL, Doi H, Barnes MA, Leray M, Hablützel PI, Deiner K (2021) Trade-offs between reducing complex terminology and producing accurate interpretations from environmental DNA: Comment on "Environmental DNA: What's behind the term?" by Pawlowski et al., (2020). Molecular Ecology 30:4601–4605. https://doi.org/10.1111/MEC.15942

Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016a) VSEARCH: A versatile open source tool for metagenomics. PeerJ 2016:e2584. https://doi.org/10.7717/PEERJ.2584

Rohart F, Gautier B, Singh A, Lê Cao KA (2017) mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Comput Biol 13:e1005752. https://doi.org/10.1371/JOURNAL.PCBI.1005752

Romano E, Ausili A, Zharova N, Celia Magno M, Pavoni B, Gabellini M (2004) Marine sediment contamination of an industrial site at Port of Bagnoli, Gulf of Naples, Southern Italy. Marine Pollution Bulletin 49:487–495. https://doi.org/10.1016/J.MARPOLBUL.2004.03.014

Romano E, Bergamin L, Finoia MG, Carboni MG, Ausili A, Gabellini M (2008) Industrial pollution at Bagnoli (Naples, Italy): Benthic foraminifera as a tool in integrated programs of environmental characterisation. Marine Pollution Bulletin 56:439–457. https://doi.org/10.1016/J.MARPOLBUL.2007.11.003

Romano E, Bergamin L, Ausili A, Pierfranceschi G, Maggi C, Sesta G, Gabellini M (2009) The impact of the Bagnoli industrial site (Naples, Italy) on sea-bottom environment. Chemical and textural features of sediments and the related response of benthic foraminifera. Marine Pollution Bulletin 59:245–256. https://doi.org/10.1016/J.MARPOLBUL.2009.09.017

Romano E, Bergamin L, Celia Magno M, Pierfranceschi G, Ausili A (2018) Temporal changes of metal and trace element contamination in marine sediments due to a steel plant: The case study of Bagnoli (Naples, Italy). Applied Geochemistry 88:85–94. https://doi.org/10.1016/J.APGEOCHEM.2017.05.012

Rutherford S, D'Hondt S, Prell W (1999) Environmental controls on the geographic distribution of zooplankton diversity. Nature 400:749–753. https://doi.org/10.1038/23449

S. dos S. de Jesus M, Frontalini F, Bouchet VMP, Yamashita C, Sartoretto JR, Figueira RCL, de Mello e Sousa SH (2020) Reconstruction of the palaeo-ecological quality status in an

impacted estuary using benthic foraminifera: The Santos Estuary (São Paulo state, SE Brazil). Marine environmental research 162:105121. https://doi.org/10.1016/J.MARENVRES.2020.105121

Savranskaia T, Egli R, Valet JP, Bassinot F, Meynadier L, Bourlès DL, Simon Q, Thouveny N (2021) Disentangling magnetic and environmental signatures of sedimentary 10Be/9Be records. Quaternary Science Reviews. https://doi.org/10.1016/j.quascirev.2021.106809

Savvides C, Papadopoulos A, Haralambous KJ, Loizidou M (1995) Sea sediments contaminated with heavy metals: metal speciation and removal. Water Science and Technology 32:65–73. https://doi.org/10.2166/WST.1995.0671

Sayre RG, Wright DJ, Breyer SP, Butler KA, van Graafeiland K, Costello MJ, Harris PT, Goodin KL, Guinotte JM, Basher Z, Kavanaugh MT, Halpin PN, Monaco ME, Cressie NA, Aniello P, Frye CE, Stephens D (2017) A three-dimensional mapping of the ocean based on environmental data. Oceanography 30:90–103. https://doi.org/10.5670/OCEANOG.2017.116

Scandone R, Giacomelli L, Speranza FF (2008) Persistent activity and violent strombolian eruptions at Vesuvius between 1631 and 1944. Journal of Volcanology and Geothermal Research 170:167–180. https://doi.org/10.1016/J.JVOLGEORES.2007.09.014

Schiebel R, Hemleben C (2017) Planktic Foraminifers in the Modern Ocean. Springer Berlin Heidelberg, Berlin, Heidelberg

Schiebel R, Spielhagen RF, Garnier J, Hagemann J, Howa H, Jentzen A, Martínez-garcia A, Meilland J, Michel E, Repschläger J (2017) Modern planktic foraminifers in the high-latitude ocean Marine Micropaleontology Modern planktic foraminifers in the high-latitude ocean. Marine Micropaleontology136:1–13. https://doi.org/10.1016/j.marmicro.2017.08.004

Schiebel R, Smart SM, Jentzen A, Jonkers L, Morard R, Meilland J, Michel E, Coxall HK, Hull PM, de Garidel-Thoron T, Aze T, Quillévéré F, Ren H, Sigman DM, Vonhof HB, Martínez-García A, Kučera M, Bijma J, Spero HJ, Haug GH (2018) Advances in planktonic foraminifer research: New perspectives for paleoceanography. Revue de Micropaleontologie

Schmiedl G (2019) Use of Foraminifera in Climate Science. Oxford Research Encyclopedia of Climate Science. https://doi.org/10.1093/ACREFORE/9780190228620.013.735

Schönfeld J, Alve E, Geslin E, Jorissen F, Korsun S, Spezzaferri S, Abramovich S, Almogi-Labin A, du Chatelet EA, Barras C, Bergamin L, Bicchi E, Bouchet V, Cearreta A, di Bella

L, Dijkstra N, Disaro ST, Ferraro L, Frontalini F, Gennari G, Golikova E, Haynert K, Hess S, Husum K, Martins V, McGann M, Oron S, Romano E, Sousa SM, Tsujimoto A (2012) The FOBIMO (FOraminiferal BIo-MOnitoring) initiative—Towards a standardised protocol for soft-bottom benthic foraminiferal monitoring studies. Marine Micropaleontology 94–95:1–13. https://doi.org/10.1016/J.MARMICRO.2012.06.001

Schulte L, Bernhardt N, Stoof-Leichsenring K, Zimmermann HH, Pestryakova LA, Epp LS, Herzschuh U, Kühn- J (2021) Hybridization capture of larch (Larix Mill.) chloroplast genomes from sedimentary ancient DNA reveals past changes of Siberian forest. Molecular Ecology Resources 21:801–815. https://doi.org/10.1111/1755-0998.13311

Selway CA, Armbrecht L, Thornalley D (2022) An Outlook for the Acquisition of Marine Sedimentary Ancient DNA (sedaDNA) From North Atlantic Ocean Archive Material. Paleoceanogr Paleoclimatol 37:e2021PA004372. https://doi.org/10.1029/2021PA004372

Sgarella F, Zei M (1993) Benthic foraminifera of the Gulf of Naples (Italy): systematics and autoecology.

Sheldon CM, Seidenkrantz M-S, Frandsen P, Jacobsen H V., Van Nieuwenhove N, Solignac S, Pearce C, Palitzsch MG, Kuijpers A (2015) Variable influx of West Greenland Current water into the Labrador Current through the last 7200 years: a multiproxy record from Trinity Bay (NE Newfoundland). Arktos 1:8. https://doi.org/10.1007/s41063-015-0010-z

Sheldon CM, Seidenkrantz M-S, Pearce C, Kuijpers A, Hansen MJ, Christensen EZ (2016) Holocene oceanographic changes in SW Labrador Sea, off Newfoundland. The Holocene 26:274–289. https://doi.org/10.1177/09596836156086

Shulse CN, Maillot B, Smith CR, Church MJ (2017) Polymetallic nodules, sediments, and deep waters in the equatorial North Pacific exhibit highly diverse and distinct bacterial, archaeal, and microeukaryotic communities. Microbiologyopen 6:e00428. https://doi.org/10.1002/MBO3.428

Siano R, Lassudrie M, Cuzin P, Briant N, Loizeau V, Schmidt S, Ehrhold A, Mertens KN, Lambert C, Quintric L, Noël C, Latimier M, Quéré J, Durand P, Penaud A (2021) Sediment archives reveal irreversible shifts in plankton communities after World War II and agricultural pollution. Current Biology 31:2682-2689.e7. https://doi.org/10.1016/j.cub.2021.03.079

Siccha M, Kucera M (2017) ForCenS , a curated database of planktonic foraminifera census counts in marine surface sediment samples. Scientific Data 4:170109:1–12. https://doi.org/10.1038/sdata.2017.109

Sicre M-A, Weckström K, Seidenkrantz M-S, Kuijpers A, Benetti M, Massé G, Ezat U, Schmidt S, Bouloubassi I, Olsen J, others (2014) Labrador current variability over the last 2000 years. Earth and Planetary Science Letters 400:26–32. https://doi.org/10.1016/j.epsl.2014.05.016

Silver MW, Gowing MM (1991) The "particle" flux: Origins and biological components. Progress in Oceanography, 26(1), 75–113 | 10.1016/0079-6611(91)90007-9. Progress in Oceanography 75–113. https://doi.org/10.1016/0079-6611(91)90007-9

Simpson Gavin, Oksanen Jari, Maechler Martin (2021) Package Analogue and Weighted Averaging Methods for Palaeoecology.

Singer GAC, Fahner NA, Barnes JG, McCarthy A, Hajibabaei M (2019a) Comprehensive biodiversity analysis via ultra-deep patterned flow cell technology: a case study of eDNA metabarcoding seawater. Scientific Reports 2019 9:1 9:1–12. https://doi.org/10.1038/s41598-019-42455-9

Sinniger F, Pawlowski J, Harii S, Gooday AJ, Yamamoto H, Chevaldonné P, Cedhagen T, Carvalho G, Creer S (2016a) Worldwide analysis of sedimentary DNA reveals major gaps in taxonomic knowledge of deep-sea benthos. Frontiers in Marine Science 3:92. https://doi.org/10.3389/FMARS.2016.00092

Sinniger F, Pawlowski J, Harii S, Gooday AJ, Yamamoto H, Chevaldonné P, Cedhagen T, Carvalho G, Creer S (2016b) Worldwide analysis of sedimentary DNA reveals major gaps in taxonomic knowledge of deep-sea benthos. Frontiers in Marine Science 3:92. https://doi.org/10.3389/FMARS.2016.00092

Sloan WT, Woodcock S, Lunn M, Head IM, Curtis TP (2007) Modeling taxa-abundance distributions in microbial communities using environmental sequence data. Microbial ecology 53:443–455. https://doi.org/10.1007/S00248-006-9141-X

Smith CR, de Leo FC, Bernardino AF, Sweetman AK, Arbizu PM (2008) Abyssal food limitation, ecosystem structure and climate change. Trends in Ecology & Evolution 23:518–528. https://doi.org/10.1016/J.TREE.2008.05.002

Smith CR, Tunnicliffe V, Colaço A, Drazen JC, Gollner S, Levin LA, Mestre NC, Metaxas A, Molodtsova TN, Morato T, Sweetman AK, Washburn T, Amon DJ (2020) Deep-Sea Misconceptions Cause Underestimation of Seabed-Mining Impacts. Trends in Ecology & Evolution 35:853–857. https://doi.org/10.1016/j.tree.2020.07.002

Smol JP (2009) A Review of "Pollution of Lakes and Rivers: A Paleoenvironmental Perspective." Blackwell Pub

Soininen J, McDonald R, Hillebrand H (2007) The distance decay of similarity in ecological communities. Ecography 30:3–12. https://doi.org/10.1111/J.0906-7590.2007.04817.X

Sprovieri M, Passaro S, Ausili A, Bergamin L, Finoia MG, Gherardi S, Molisso F, Quinci EM, Sacchi M, Sesta G, Trincardi F, Romano E (2020) Integrated approach of multiple environmental datasets for the assessment of sediment contamination in marine areas affected by long-lasting industrial activity: the case study of Bagnoli (southern Italy). J Soils Sediments 20:1692–1705. https://doi.org/10.1007/S11368-019-02530-0

Steinberg DK, Landry MR (2017) Zooplankton and the Ocean Carbon Cycle. Annual review of marine science 9:413–444. https://doi.org/10.1146/annurev-marine-010814-015924

Stoof-Leichsenring KR, Epp LS, Trauth MH, Tiedemann R (2012) Hidden diversity in diatoms of Kenyan Lake Naivasha: A genetic approach detects temporal variation. Molecular Ecology 21:1918–1930. https://doi.org/10.1111/J.1365-294X.2011.05412.X

Sweetman AK, Thurber AR, Smith CR, Levin LA, Mora C, Wei CL, Gooday AJ, Jones DOB, Rex M, Yasuhara M, Ingels J, Ruhl HA, Frieder CA, Danovaro R, Würzberg L, Baco A, Grupe BM, Pasulka A, Meyer KS, Dunlop KM, Henry LA, Roberts JM (2017) Major impacts of climate change on deep-sea benthic ecosystems. Elementa. https://doi.org/10.1525/ELEMENTA.203/112418

Szczucinski W, Pawlowska J, Lejzerowicz F, Nishimura Y, Kokocinski M, Majewski W, Nakamura Y, Pawlowski J (2016) Ancient sedimentary DNA reveals past tsunami deposits. Marine Geology 381:29–33. https://doi.org/10.1016/j.margeo.2016.08.006

Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, Vermat T, Corthier G, Brochmann C, Willerslev E (2007) Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic acids research 35:e14. https://doi.org/10.1093/NAR/GKL938

Taberlet P, Coissac E, Hajibabaei M, Riesenberg LH (2012) Environmental DNA. Molecular Ecology 21:1789–1793. https://doi.org/10.1111/j.1365-294x.2012.05542.x

Taberlet P, Bonin A, Zinger L, Coissac E (2018) Environmental DNA: For biodiversity research and monitoring. Environmental DNA: For Biodiversity Research and Monitoring 1–253. https://doi.org/10.1093/oso/9780198767220.001.0001

Tachikawa K, Cartapanis O, Vidal L, Beaufort L, Barlyaeva T, Bard E (2011) The precession phase of hydrological variability in the Western Pacific Warm Pool during the past 400 ka. Quaternary Science Reviews 30:3716–3727. https://doi.org/10.1016/J.QUASCIREV.2011.09.016

Takahashi K, Be AWH (1984) Planktonic foraminifera: factors controlling sinking speeds. Deep Sea Research Part A Oceanographic Research Papers 31:1477–1500. https://doi.org/10.1016/0198-0149(84)90083-9

Tang CQ, Leasi F, Obertegger U, Kieneke A, Barraclough TG, Fontaneto D (2012) The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. Proceedings of the National Academy of Sciences 109:16208–16212. https://doi.org/10.1073/PNAS.1209160109

Tangherlini M, Corinaldesi C, Rastelli E, Musco L, Armiento G, Danovaro R, Dell'Anno A (2020) Chemical contamination can promote turnover diversity of benthic prokaryotic assemblages: The case study of the Bagnoli-Coroglio bay (southern Tyrrhenian Sea). Marine environmental research 160:105040. https://doi.org/10.1016/J.MARENVRES.2020.105040

Tennant RK, Power AL, Burton SK, Sinclair N, Parker DA, Jones RT, Lee R, Love J (2022a) In-situ sequencing reveals the effect of storage on lacustrine sediment microbiome demographics and functionality. Environmental Microbiomes 17:1–17. https://doi.org/10.1186/S40793-022-00400-W

Thiel H (1988) Phytodetritus on the deep-sea floor in a central oceanic region of the northeast Atlantic. Biological Oceanography 6:203–239. https://doi.org/10.1080/01965581.1988.10749527

Thomas E, Gooday AJ (1996) Cenozoic deep-sea benthic foraminifers: Tracers for changes in oceanic productivity? Geology 24:355–358. https://doi.org/10.1130/0091-7613(1996)024<0355:CDSBFT>2.3.CO;2

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

Thurber AR, Sweetman AK, Narayanaswamy BE, Jones DOB, Ingels J, Hansman RL (2014) Ecosystem function and services provided by the deep sea. Biogeosciences 11:3941–3963. https://doi.org/10.5194/BG-11-3941-2014

Tollefsrud MM, Kissling R, Gugerli F, Johnsen Ø, Skrøppa T, Cheddadi R, van der Knaap WO, Latałowa M, Terhürne-Berson R, Litt T, Geburek T, Brochmann C, Sperisen C (2008) Genetic consequences of glacial survival and postglacial colonization in Norway spruce: combined analysis of mitochondrial DNA and fossil pollen. Molecular Ecology 17:4134–4150. https://doi.org/10.1111/J.1365-294X.2008.03893.X

Torti A, Lever MA, Jørgensen BB (2015) Origin, dynamics, and implications of extracellular DNA pools in marine sediments. Marine Genomics 24:185–196. https://doi.org/10.1016/J.MARGEN.2015.08.007

Torti A, Jørgensen BB, Lever MA (2018) Preservation of microbial DNA in marine sediments: insights from extracellular DNA pools. Environmental Microbiology 20:4526–4542. https://doi.org/10.1111/1462-2920.14401

Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic acids research 44:W232. https://doi.org/10.1093/NAR/GKW256

Trifuoggi M, Donadio C, Mangoni O, Ferrara L, Bolinesi F, Nastro RA, Stanislao C, Toscanesi M, di Natale G, Arienzo M (2017) Distribution and enrichment of trace metals in surface marine sediments in the Gulf of Pozzuoli and off the coast of the brownfield metallurgical site of Ilva of Bagnoli (Campania, Italy). Marine Pollution Bulletin 124:502–511. https://doi.org/10.1016/J.MARPOLBUL.2017.07.033

Triki HZ, Laabir M, Lafabrie C, Malouche D, Bancon-Montigny C, Gonzalez C, Deidun A, Pringault O, Daly-Yahia OK (2017) Do the levels of industrial pollutants influence the distribution and abundance of dinoflagellate cysts in the recently-deposited sediment of a Mediterranean coastal ecosystem? Science of The Total Environment 595:380–392. https://doi.org/10.1016/J.SCITOTENV.2017.03.183

Trubovitz S, Lazarus D, Renaudie J, Noble PJ (2020) Marine plankton show threshold extinction response to Neogene climate change. Nature Communications 2020 11:1 11:1–10. https://doi.org/10.1038/s41467-020-18879-7

Tsuda Y, Chen J, Stocks M, Källman T, Sønstebø JH, Parducci L, Semerikov V, Sperisen C, Politov D, Ronkainen T, Väliranta M, Vendramin GG, Tollefsrud MM, Lascoux M (2016) The extent and meaning of hybridization and introgression between Siberian spruce (Picea obovata) and Norway spruce (Picea abies): cryptic refugia as stepping stones to the west? Molecular Ecology 25:2773–2789. https://doi.org/10.1111/MEC.13654

Tzedakis PC, Andrieu V, de Beaulieu JL, Crowhurst S, Follieri M, Hooghiemstra H, Magri D, Reille M, Sadori L, Shackleton NJ, Wijmstra TA (1997) Comparison of terrestrial and marine records of changing climate of the last 500,000 years. Earth and Planetary Science Letters 150:171–176. https://doi.org/10.1016/S0012-821X(97)00078-2

Ujiié Y, Ishitani Y (2016) Evolution of a Planktonic Foraminifer during Environmental Changes in the Tropical Oceans. PLoS One 11:e0148847. https://doi.org/10.1371/JOURNAL.PONE.0148847

Vacher C, Tamaddoni-Nezhad A, Kamenova S, Peyrard N, Moalic Y, Sabbadin R, Schwaller L, Chiquet J, Smith MA, Vallance J, Fievet V, Jakuschkin B, Bohan DA (2016) Learning Ecological Networks from Next-Generation Sequencing Data. Advances in ecological research 54:1–39. https://doi.org/10.1016/BS.AECR.2015.10.004

van den Bulcke L, de Backer A, Ampe B, Maes S, Wittoeck J, Waegeman W, Hostens K, Derycke S (2021) Towards harmonization of DNA metabarcoding for monitoring marine macrobenthos: the effect of technical replicates and pooled DNA extractions on species detection. Metabarcoding and Metagenomics 5: e71107 5:e71107-. https://doi.org/10.3897/MBMG.5.71107

van den Heuvel-Greve MJ, van den Brink AM, Glorius ST, de Groot GA, Laros I, Renaud PE, Pettersen R, Węsławski JM, Kuklinski P, Murk AJ (2021) Early detection of marine non-indigenous species on Svalbard by DNA metabarcoding of sediment. Polar Biol 44:653–665. https://doi.org/10.1007/S00300-021-02822-7

Vargas C de, Pollina T, Romac S, Bescot N le, Henry N, Berger C, Colin S, Haëntjens N, Carmichael M, Guen D le, Decelle J, Mahé F, Malpot E, Beaumont C, Hardy M, the planktonauts the PP team, Guiffant D, Probert I, Gruber DF, Allen A, Gorsky G, Follows M, Cael BB, Pochon X, Troublé R, Lombard F, Boss E, Prakash M (2020) Plankton Planet: 'seatizen' oceanography to assess open ocean life at the planetary scale. bioRxiv 2020.08.31.263442. https://doi.org/10.1101/2020.08.31.263442

Voltaggio M, Branca M, Tedesco D, Tuccimei P, di Pietro L, Voltaggio M, Branca M, Tedesco D, Tuccimei P, di Pietro L (2004) ^{226}Ra-excess during the 1631-1944 activity period of Vesuvius (Italy) . A model of alpha-recoil enrichment in a metasomatized mantle and implications on the current state of the magmatic system. Geochimica et Cosmochimica Acta 68:167–181. https://doi.org/10.1016/S0016-7037(03)00236-9

Washburn TW, Menot L, Bonifácio P, Pape E, Błażewicz M, Bribiesca-Contreras G, Dahlgren TG, Fukushima T, Glover AG, Ju SJ, Kaiser S, Yu OH, Smith CR (2021) Patterns of Macrofaunal Biodiversity Across the Clarion-Clipperton Zone: An Area Targeted for Seabed Mining. Frontiers in Marine Science 8:250. https://doi.org/10.3389/FMARS.2021.626571

Watling L, Guinotte J, Clark MR, Smith CR (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 AA-T, Pawlowski J (2013) Can Abundance of Protists Be Inferred from Sequence Data: A Case Study of Foraminifera. PLoS One 8:e56739. https://doi.org/10.1371/journal.pone.0056739

Weber AAT, Pawlowski J (2014) Wide Occurrence of SSU rDNA Intragenomic Polymorphism in Foraminifera and its Implications for Molecular Species Identification. Protist 165:645–661. https://doi.org/10.1016/J.PROTIS.2014.07.006

Webster G, Newberry CJ, Fry JC, Weightman AJ (2003) Assessment of bacterial community structure in the deep sub-seafloor biosphere by 16S rDNA-based techniques: a cautionary tale. Journal of Microbiological Methods 55:155–164. https://doi.org/10.1016/S0167-7012(03)00140-4

Weigand H, Beermann AJ, Čiampor F, Costa FO, Csabai Z, Duarte S, Geiger MF, Grabowski M, Rimet F, Rulik B, Strand M, Szucsich N, Weigand AM, Willassen E, Wyler SA, Bouchez A, Borja A, Čiamporová-Zaťovičová Z, Ferreira S, Dijkstra KDB, Eisendle U, Freyhof J, Gadawski P, Graf W, Haegerbaeumer A, van der Hoorn BB, Japoshvili B, Keresztes L, Keskin E, Leese F, Macher JN, Mamos T, Paz G, Pešić V, Pfannkuchen DM, Pfannkuchen MA, Price BW, Rinkevich B, Teixeira MAL, Várbíró G, Ekrem T (2019) DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. Science of The Total Environment 678:499–524. https://doi.org/10.1016/J.SCITOTENV.2019.04.247

Weiner AKM, Morard R, Weinkauf MF, Darling KF, André A, Quillévéré F, Ujiie Y, Douady CJ, de Vargas C, Kucera M (2016) Methodology for single-cell genetic analysis of planktonic foraminifera for studies of protist diversity and evolution. Frontiers in Marine Science 3:1–15. https://doi.org/10.3389/fmars.2016.00255

Westerhold T, Marwan N, Drury AJ, Liebrand D, Agnini C, Anagnostou E, Barnet JSK, Bohaty SM, de Vleeschouwer D, Florindo F, Frederichs T, Hodell DA, Holbourn AE, Kroon D, Lauretano V, Littler K, Lourens LJ, Lyle M, Pälike H, Röhl U, Tian J, Wilkens RH, Wilson PA, Zachos JC (2020) An astronomically dated record of Earth's climate and its predictability over the last 66 million years. Science (1979) 369:1383–1388. https://doi.org/10.1126/SCIENCE.ABA6853

Wölfl AC, Snaith H, Amirebrahimi S, Devey CW, Dorschel B, Ferrini V, Huvenne VAI, Jakobsson M, Jencks J, Johnston G, Lamarche G, Mayer L, Millar D, Pedersen TH, Picard K, Reitz A, Schmitt T, Visbeck M, Weatherall P, Wigley R (2019) Seafloor mapping - The challenge of a truly global ocean bathymetry. Frontiers in Marine Science 6:283. https://doi.org/10.3389/FMARS.2019.00283

Worm B, Barbier EB, Beaumont N, Duffy JE, Folke C, Halpern BS, Jackson JBC, Lotze HK, Micheli F, Palumbi SR, Sala E, Selkoe KA, Stachowicz JJ, Watson R (2006) Impacts of biodiversity loss on ocean ecosystem services. Science (1979) 314:787–790. https://doi.org/10.1126/SCIENCE.1132294

Wright DJ (1999) Getting to the bottom of it: Tools, techniques, and discoveries of deep ocean geography. Professional Geographer 51:426–439. https://doi.org/10.1111/0033-0124.00177

Wright MN, Ziegler A (2017) ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. J Stat Softw 77:1–17. https://doi.org/10.18637/JSS.V077.I01

Xiong W, Zhan A (2018) Testing clustering strategies for metabarcoding-based investigation of community–environment interactions. Molecular Ecology Resources 18:1326–1338. https://doi.org/10.1111/1755-0998.12922

Xu D, Li R, Hu C, Sun P, Jiao N, Warren A (2017) Microbial eukaryote diversity and activity in the water column of the South China sea based on DNA and RNA high throughput sequencing. Frontiers in microbiology 8:1121. https://doi.org/10.3389/FMICB.2017.01121

Xu Y, Vick-Majors T, Morgan-Kiss R, Priscu JC, Amaral-Zettler L (2014) Ciliate diversity, community structure, and Novel Taxa in lakes of the McMurdo Dry Valleys, Antarctica. Biological Bulletin 227:175–190. https://doi.org/10.1086/BBLV227N2P175

Yasuhara M, Tittensor DP, Hillebrand H, Worm B (2017) Combining marine macroecology and palaeoecology in understanding biodiversity: microfossils as a model. Biological Reviews 92:199–215. https://doi.org/10.1111/BRV.12223

Yu G, Smith DK, Zhu H, Guan Y, Lam TTY (2017) ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods in Ecology and Evolution 8:28–36. https://doi.org/10.1111/2041-210X.12628

Zamkovaya T, Foster JS, de Crécy-Lagard V, Conesa A (2021) A network approach to elucidate and prioritize microbial dark matter in microbial communities. ISME JOURNAL 15:228. https://doi.org/10.1038/S41396-020-00777-X

Zeppilli D, Pusceddu A, Trincardi F, Danovaro R (2016) Seafloor heterogeneity influences the biodiversity–ecosystem functioning relationships in the deep sea. Scientific Reports 2016 6:1 6:1–12. https://doi.org/10.1038/srep26352

Zhao F, Filker S, Xu K, Li J, Zhou T, Huang P (2019) Effects of intragenomic polymorphism in the SSU rRNA gene on estimating marine microeukaryotic diversity: A test for ciliates using single-cell high-throughput DNA sequencing. Limnology and oceanography Methods 17:533–543. https://doi.org/10.1002/LOM3.10330

Zheng Q, Bartow-McKenney C, Meisel JS, Grice EA (2018) HmmUFOtu: An HMM and phylogenetic placement based ultra-fast taxonomic assignment and OTU picking tool for microbiome amplicon sequencing studies. Genome Biol 19:1–20. https://doi.org/10.1186/S13059-018-1450-0

Zheng X, Liu W, Dai X, Zhu Y, Wang J, Zhu Y, Zheng H, Huang Y, Dong Z, Du W, Zhao F, Huang L (2021) Extraordinary diversity of viruses in deep-sea sediments as revealed by metagenomics without prior virion separation. Environmental Microbiology 23:728–743. https://doi.org/10.1111/1462-2920.15154

Zielske S, Haase M (2015) Molecular phylogeny and a modified approach of character-based barcoding refining the taxonomy of New Caledonian freshwater gastropods (Caenogastropoda, Truncatelloidea, Tateidae). Molecular phylogenetics and evolution 89:171–181. https://doi.org/10.1016/J.YMPEV.2015.04.020

Zimmermann HH, Stoof-Leichsenring KR, Kruse S, Nürnberg D, Tiedemann R, Herzschuh U (2021) Sedimentary Ancient DNA From the Subarctic North Pacific: How Sea Ice, Salinity, and Insolation Dynamics Have Shaped Diatom Composition and Richness Over the Past 20,000 Years. Paleoceanography and Paleoclimatology 36:e2020PA004091. https://doi.org/10.1029/2020PA004091

Zinger L, Amaral-Zettler LA, Fuhrman JA, Horner-Devine MC, Huse SM, Welch DBM, Martiny JBH, Sogin M, Boetius A, Ramette A (2011) Global Patterns of Bacterial Beta-Diversity in Seafloor and Seawater Ecosystems. PLoS One 6:e24570. https://doi.org/10.1371/JOURNAL.PONE.0024570

Appendix: Other publications/contributions

Benthic foraminiferal DNA metabarcodes significantly vary along a gradient from abyssal to hadal depths and between each side of the Kuril-Kamchatka trench

Cordier, T., Barrenechea, I., Lejzerowicz, F., Reo, E., & Pawlowski, J.

Published in: Progress in Oceanography 178, 102175 (2019).

Abstract

Foraminiferal assemblages are a ubiquitous and abundant component of the deep-sea benthos, even in the deepest ocean trenches. While their distribution seems not constrained over large geographical distance, the current knowledge of foraminifera in trench is solely based on morphological observations. In this study, we document the first DNA metabarcoding dataset from a deep-sea trench focusing specifically on benthic foraminifera. Here we show that, consistent with previous molecular studies of abyssal fauna, trench foraminifera include diverse sequences of yet unknown species captured only by their molecular traces in the sediment. The molecular assemblages of foraminifera significantly differed along a depth gradient of almost 5000 m in the Kuril-Kamchatka trench. The deepest stations at nearly 9500 m were composed of unique phylotypes that were not identified in shallower stations, which means that these assemblages are unlikely the result of a sinking effect from shallower depths. Finally, both sides of the trench harbored very different communities, which could imply that the trench constitutes a physical barrier for the dispersion of some deep-sea foraminiferal species.

Eukaryotic Biodiversity and Spatial Patterns in the Clarion-Clipperton Zone and Other Abyssal Regions: Insights From Sediment DNA and RNA Metabarcoding

Franck Lejzerowicz, Andrew John Gooday, Inés Barrenechea Angeles, Tristan Cordier, Raphaël Morard, Laure Apothéloz-Perret-Gentil, Lidia Lins, Lenaick Menot, Angelika Brandt1, Lisa Ann Levin, Pedro Martinez Arbizu, Craig Randall Smith and Jan Pawlowski

Published in: Frontiers in *Frontiers in Marine Science* 8, 671033 (2021).

Abstract

The abyssal seafloor is a mosaic of highly diverse habitats that represent the least known marine ecosystems on Earth. Some regions enriched in natural resources, such as polymetallic nodules in the Clarion-Clipperton Zone (CCZ), attract much interest because of their huge commercial potential. Since nodule mining will be destructive, baseline data are necessary to measure its impact on benthic communities. Hence, we conducted an environmental DNA and RNA metabarcoding survey of CCZ biodiversity targeting microbial and meiofaunal eukaryotes that are the least known component of the deep-sea benthos. We analyzed two 18S rRNA gene regions targeting eukaryotes with a focus on Foraminifera (37F) and metazoans (V1V2), sequenced from 310 surface-sediment samples from the CCZ and other abyssal regions. Our results confirm huge unknown deep-sea biodiversity. Over 60% of benthic foraminiferal and almost a third of eukaryotic operational taxonomic units (OTUs) could not be assigned to a known taxon. Benthic Foraminifera are more common in CCZ samples than metazoans and dominated by clades that are only known from environmental surveys. The most striking results are the uniqueness of CCZ areas, both datasets being characterized by a high number of OTUs exclusive to the CCZ, as well as greater beta diversity compared to other abyssal regions. The alpha diversity in the CCZ is high and correlated with water depth and terrain complexity. Topography was important at a local scale, with communities at CCZ stations located in depressions more diverse and heterogeneous than those located on slopes. This could result from eDNA accumulation, justifying the interim use of eRNA for more accurate biomonitoring surveys. Our descriptions not only support previous findings and consolidate our general understanding of deep-sea ecosystems, but also provide a data resource inviting further taxon-

specific and large-scale modeling studies. We foresee that metabarcoding will be useful for deep-sea biomonitoring efforts to consider the diversity of small taxa, but it must be validated based on ground truthing data or experimental studies.

Metabarcoding Reveals High Diversity of Benthic Foraminifera Driven by Atlantification of Coastal Svalbard

Nguyen, N. L., Pawłowska, J., Angeles, I. B., Zajaczkowski, M., & Pawłowski, J.

Accepted in Geobiology (2022)

Abstract

Arctic marine biodiversity is undergoing rapid changes due to global warming and modifications of oceanic water masses circulation. These changes have been demonstrated in the case of mega- and macrofauna, but much less is known about their impact on the biodiversity of smaller size organisms, such as foraminifera that represents a main component of meiofauna in the Arctic. Several studies analysed the distribution and diversity of Arctic foraminifera. However, all these studies are based exclusively on the morphological identification of specimens sorted from sediment samples. Here, we present the first assessment of Arctic foraminifera diversity based on metabarcoding of sediment DNA samples collected in fjords and open sea areas in Svalbard Archipelago. We obtained a total of 5,968,786 reads that represented 1,384 ASVs. More than half of the ASVs (51.7%) could not be assigned to any group in the reference database suggesting a high genetic novelty of Svalbard foraminifera. The sieved and unsieved samples resolved comparable communities, sharing 1023 ASVs, comprising over 97% of reads. Our analyses show that the foraminiferal assemblage differs between the localities, with communities distinctly separated between fjord and open sea stations. Each locality was characterized by a specific assemblage, with only a small overlap in the case of open sea areas. Our study demonstrates a clear pattern of the influence of water masses on the structure of foraminiferal communities. The stations situated on the western coast of Svalbard that is strongly influenced by warm and salty Atlantic Water (AW) are characterized by much higher diversity than stations in the northern and eastern part, where the impact of AW is less pronounced. This high diversity and specificity of Svalbard foraminifera associated with water mass distribution indicate that the foraminiferal metabarcoding data can be a very useful tool for inferring present and past environmental conditions in the Arctic.