Institute of Oceanology, Polish Academy of Sciences

Department of Paleoceanography

EXPLORING THE MARINE BIODIVERSITY WITH ENVIRONMENTAL DNA

Nguyen Ngoc Loi

Doctoral thesis constituting a collection of thematically related scientific publications.

This thesis was prepared under the supervision of

dr hab. ***Joanna Pawłowska-Tomkowicz***

and the co-supervision of

dr ***Jan Pawłowski***

Sopot, March 2024

Instytut Oceanologii Polskiej Akademii Nauk

Zakład Paleoceanografii

ODKRYWANIE BIORÓŻNORODNOŚCI MORSKIEJ Z WYKORZYSTANIEM ŚRODOWISKOWEGO DNA

Nguyen Ngoc Loi

Praca doktorska będąca zbiorem tematycznie powiązanych artykułów naukowych.

Praca przygotowana pod kierunkiem promotorki

dr hab. ***Joanny Pawłowskiej-Tomkowicz***

oraz współpromotora

dr ***Jana Pawłowskiego***

Sopot, Marzec 2024

tanquam ex ungue leonem

by Johann Bernoulli, 1697

TABLE OF CONTENTS

ACKNOWLEDGMENTS / PODZIĘKOWANIE

I would like to take the opportunity to thank all the people who supported and guided me throughout this entire PhD journey, but also to all my friends and colleagues who supported my personal and scientific development.

First of all, I would like to express my gratitude to my supervisors, Dr. Joanna Pawłowska-Tomkowicz and Prof. Jan Pawłowski, for giving me this wonderful opportunity in the first place. Further, I am very grateful for their scientific advice, moral support, patience, freedom, and trust to develop my own ideas.

Secondly, I would like to thank Prof. Marek Zajaczkowski and the "*PALEOpeople*" for always having their offices open and having enthusiastic discussions about the fascinating world of paleoceanography. Special thanks go to Dr. Natalia Szymańska, Dr. Inès Barrenechea Angeles and Dhanushka Devendra for their always professional and kind help when and wherever I needed it. I am thankful for their willingness to share their expertise, engage in numerous scientific discussions, provide constructive feedback, and offer assistance, particularly in their shared passion for foraminifera. Forams forever !!!

I extend my heartfelt thanks to the Institute of Oceanology, Polish Academy of Sciences staff for providing me with the resources, facilities, and technical and organizational assistance for conducting this research. The captain and entire crew of the R/V *Oceania* is sincerely thanked for their outstanding professionalism, and expertise throughout expeditions scattered over the Arctic.

I am indebted to my family for their unquestionable love, encouragement, and understanding during the challenging times of this academic pursuit. Their unwavering support has been my greatest source of strength. The biggest thank you is for my mother and her never-ending support and belief in my crazy dreams and ambitions which is an inspiration in itself.

Lastly, I would like to acknowledge my study participants for their willingness to contribute their time and insights, without whom this research would not have been possible. This thesis is a culmination of the collective efforts, support, and guidance of all those who have crossed paths with me during this academic endeavor, and for that, I am sincerely grateful.

ABBREVIATIONS / SKRÓTY

*e*DNA – environmental DNA / środowiskowe DNA

*sed*DNA – sedimentary DNA / osadowe DNA

*seda*DNA – sedimentary ancient DNA /kopalne DNA osadowe

ASV – Amplicon Sequence Variant /wariant sekwencji amplikonu

OTU – Operational Taxonomic Unit /operacyjna jednostka taksonomiczna

AW – Atlantic Water / woda atlantycka

ArW – Arctic Water / woda arktyczna

CCFZ – Clarion-Clipperton Fracture Zone / Strefa Clariona Clippertona

18S rRNA – small subunit ribosomal RNA / mała podjednostka rybosomalnego RNA

PCR – Polymerase Chain Reaction /reakcja łańcuchowa polimerazy

RA – Research Article or Review Article / artykuł naukowy lub przeglądowy

ABSTRACT (in English)

Marine sedimentary archives are an important repository of the whole marine biodiversity comprising both benthic and pelagic organisms. Environmental DNA (*e*DNA), which accumulates and preserves in marine sediments, can be used to study the taxonomic composition of living communities as well as to reconstruct past biodiversity depending on whether it concerns modern or historical *e*DNA deposits. The analysis of sedimentary DNA (*sed*DNA) or sedimentary ancient DNA (*seda*DNA) opens up entirely new possibilities for the study of short- and long-term responses of marine ecosystems to environmental changes. Recent advances in high-throughput sequencing technologies allow to rapidly sequence DNA from marine environments and led to a consistently increasing number of metabarcoding studies, especially for microbial biomes. However, various challenges and biases that affect the generation and analysis of metabarcoding data are not fully resolved. To unlock the full potential of *e*DNA metabarcoding applied to both modern and paleoceanographic studies more research is needed to better understand the relation between *e*DNA taphonomy and environmental changes in marine settings.

In order to further enhance the use of *e*DNA in present and past marine biodiversity studies, this thesis addressed the following research tasks: i) investigating the biodiversity of eukaryotes in water column and surface sediments and the preservation of planktonic *e*DNA on the seafloor; ii) investigating the biodiversity of selected eukaryotic taxa (foraminifera) and their responses to environmental parameters; and iii) summarizing the current advances in marine *seda*DNA research and discussing potential methodological pitfalls and limitations.

The first part describes the marine eukaryotic communities (RA I), from water column to surface sediment, and their *e*DNA taphonomy. The main advantage of the *e*DNA approach is the possibility of obtaining a holistic record of marine biodiversity. However, very little is known about how accurately marine biodiversity is recorded in sedimentary DNA archives, especially in terms of planktonic taxa. To address this important question, we provided a vertical and horizontal survey of eukaryotic diversity in the Nordic Seas and compared eukaryotic diversity throughout the water column to surface sediment. Our study has led to the following conclusions: i) the taxonomic composition of water and sediment *e*DNA samples differs significantly; ii) a large amount of plankton DNA is transported to the surface sediments and dominates sediment DNA data in terms of abundance but not diversity; iii) not all plankton taxa are equally archived on the sea floor, with some nano- and picoplankton taxa being underrepresented in sediment DNA samples. Overall, these results suggest that the composition

and structure of the plankton community recorded in sedimentary *e*DNA differ from what is observed in the water column. This highlights potential taxonomic and abundance biases that should be taken into account when reconstructing past marine biodiversity changes.

The second part focuses on the diversity of benthic foraminifera and their role as ecological indicators. This part comprises two studies. In the first study (RA II), we used an *e*DNA metabarcoding of surface sediments to investigate the diversity of Arctic foraminifera in fjords and open sea areas of the Svalbard Archipelago. Our analysis of metabarcoding data revealed a very high phylogenetic diversity of foraminifera compared to traditional morphology-based studies. More than half of the Amplicon Sequence Variants (ASVs) could not be assigned to any group in the reference database, suggesting a high genetic novelty of Svalbard foraminifera. The taxonomic composition of the foraminiferal community varied between sampling localities (fjords and open sea areas), influenced by different water masses. Numerous potential molecular foraminiferal indicators of water mass characteristics were identified, particularly regarding the impact of Atlantic Water in the Svalbard region. This study provided the first comprehensive metabarcoding data on foraminiferal biodiversity in the Svalbard area and contributed to a better knowledge on how the foraminiferal community responds to Arctic environmental gradients.

In the second study (RA III), we analyzed the deep-sea foraminifera, focusing on a huge unknown diversity revealed by metabarcoding data. We tackled this problem by using the specific genetic signature to classify unassigned foraminiferal sequences, which usually dominate in *e*DNA metabarcoding datasets. We applied this approach to benthic foraminifera from Clarion-Clipperton Fracture Zone biodiversity in the Eastern Pacific Ocean, comparing their diversity to available foraminiferal datasets from other deep-sea and shallow-water regions. As a result, 61 new foraminiferal lineages placed in 27 phylogenetic clades were identified by unique signatures in the 37F hypervariable region of the 18S rRNA gene. Most of these novel lineages were also found in other deep-sea areas, but only a few of them appeared in coastal datasets. This suggests that deep-sea benthic foraminifera form a unique group highly adapted to the abyssal environment and that the migration between shallow and deep-sea habitats is relatively limited. The signature-based approach provides an alternative to investigating the distribution and ecology of deep-sea foraminifera, given the limited current reference database. It could be especially useful in future applications of foraminiferal metabarcoding for environmental monitoring.

The last part of this thesis (RA IV) provides an overview of spectacular advances that have been made in reconstructing the history of marine ecosystems using the *seda*DNA approach. In

this article, we conducted a systematic literature review of 55 original studies to examine the last two decades of marine *seda*DNA research. We focus on both planktonic and benthic microbial (prokaryotes and single-cell eukaryotes) and meiofaunal organisms, whose genetic traces are deposited in marine sediments. We describe an in-depth overview of taphonomic or preservation processes, key issues related to the use of *seda*DNA, and the current state of knowledge and applications in marine *seda*DNA research. We anticipate that *seda*DNA approaches will soon be routinely included in paleoceanographic studies and will provide a unique insight into the biodiversity changes at geological timescales, recent anthropogenic impacts, and the past and present evolution of marine ecosystems. The continued development of the *seda*DNA field might also help to establish and optimize strategies for the conservation and management of marine ecosystems.

Overall, my PhD thesis presents various applications of *e*DNA metabarcoding to study past and present ecosystems and highlights its potential and limitations. The results obtained in this thesis contribute to exploring the diversity of deep-sea and polar foraminifera and provide insights into the biases associated with the *e*DNA taphonomy of marine eukaryotes. As demonstrated in this thesis, the use of *e*DNA metabarcoding is crucial to further advance the surveys of marine biodiversity across time and space.

ABSTRAKT (w języku polskim)

Osady morskie są niezwykle istotnym archiwum morskiej bioróżnorodności, obejmującej zarówno organizmy bentosowe, jak i pelagiczne. Środowiskowe DNA (*e*DNA), które zachowuje się w osadach morskich, może być wykorzystywane zarówno do badania składu taksonomicznego współczesnych zbiorowisk organizmów, jak i do rekonstrukcji bioróżnorodności w przeszłości. Analiza DNA osadowego (*sed*DNA) lub kopalnego DNA osadowego (*seda*DNA) otwiera zupełnie nowe możliwości badania krótko- i długoterminowych zmian ekosystemów morskich zachodzących w odpowiedzi na zmiany klimatyczne i środowiskowe. Ogromny postęp jaki obserwujemy w rozwoju technologii sekwencjonowania nowej generacji pozwala na szybkie sekwencjonowanie DNA ze środowisk morskich czego rezultatem jest stale rosnąca liczba badań opartych o metabarkodowanie (ang. *metabarcoding*). Liczba tych badań rośnie szczególnie intensywnie w przypadku zbiorowisk mikroorganizmów. Jednocześnie, problemy techniczne i wynikające z nich potencjalne błędy nie zostały jeszcze wyeliminowane, choć mogą wpływać na generowanie i analizę danych z metabarkodowania. Dlatego, aby w pełni wykorzystać potencjał metabarkodowania *e*DNA w badaniach współczesnych, jak i przeszłych ekosystemów, potrzebne są dalsze badania. Pozwolą one lepiej zrozumieć związek między zapisami *e*DNA a zmianami w środowisku morskim.

Aby w pełni wykorzystać potencjał *e*DNA w badaniach współczesnej i przeszłej bioróżnorodności morskiej, w niniejszej rozprawie wyznaczono następujące cele: i) zbadanie bioróżnorodności organizmów eukriotycznych w kolumnie wody i osadach powierzchniowych oraz zbadanie stopnia zachowania eDNA planktonicznego w osadach morskich; ii) zbadanie bioróżnorodności zbiorowisk wybranych organizmów eukariotycznych (foraminifera) i ich odpowiedzi na zmiany środowiskowe; oraz iii) podsumowanie obecnych postępów w badaniach nad morskim *seda*DNA oraz omówienie potencjalnych ograniczeń i problemów metodologicznych. Rozprawa doktorska składa się z czterech artykułów naukowych realizujących powyższe zadania badawcze.

Pierwsza część rozprawy, bazując na analizie *e*DNA (RA I), opisuje morskie zbiorowiska eukariotyczne, od kolumny wody po osady powierzchniowe. Główną zaletą analizy *e*DNA jest możliwość uzyskania pełnego zapisu bioróżnorodności morskiej, w tym organizmów nie uwzględnianych w tradycyjnych analizach mikroskopowych. Jednocześnie nieznana jest dokładność zapisu morskiej bioróżnorodności w DNA osadowym, szczególnie w przypadku organizmów planktonicznych. Aby zgłębić to zagadnienie przeprowadzono badanie zbiorowisk eukariotycznych w Morzach Nordyckich i porównano bioróżnorodność eukariotyczną w

kolumnie wody i w osadach powierzchniowych. Nasze badania doprowadziły do następujących wniosków: i) skład taksonomiczny próbek *e*DNA wody i osadów różni się znacząco; ii) duża ilość DNA planktonowego jest transportowana do osadów powierzchniowych i dominuje w zapisie DNA osadowego pod względem liczby sekwencji, ale nie bioróżnorodności; iii) nie wszystkie gatunki planktonowe są w równym stopniu archiwizowane na dnie morskim, a niektóre taksony nano- i pikoplanktonu są praktycznie nieobecne w DNA osadowym. Uzyskane wyniki sugerują, że skład i struktura zbiorowisk planktonowych zarejestrowanych w DNA osadowym różnią się od tego, co obserwuje się w kolumnie wody. Wyniki te sugerują, że skład taksonomiczny i struktura zbiorowisk planktonowych znacznie zmienia się wraz z głębokością wody. Jednocześnie, tylko nieliczne grupy organizmów żyjących w toni wodnej zachowują się w osadach dennych. Ma to znaczenie w kontekście interpretacji zapisów kopalnego DNA osadowego i sugeruje potencjane błędy wynikające z niekompletnego zapisu organizmów planktonowych.

Druga część, obejmująca dwa artykuły, koncentruje się na zmienności przestrzennej i bioróżnorodności otwornic bentosowych i ich roli jako wskaźników ekologicznych. W pierwszej pracy (RA II) przy pomocy metabarkodowania *e*DNA z osadów powierzchniowych zbadano różnorodność arktycznych otwornic w fiordach i rejonach otwartego morza archipelagu Svalbard. Analiza danych z metabarkodowania ujawniła bardzo wysoką różnorodność filogenetyczną otwornic w porównaniu z tradycyjnymi badaniami opartymi na analizie morfologicznej. Ponad połowa wariantów sekwencji amplikonów (ASV) nie mogła być przypisana do żadnej grupy w referencyjnej bazie danych, co sugeruje potencjalnie dużą obecność nieznanych linii genetycznych wśród otwornic wód Svalbardu. Skład taksonomiczny zbiorowisk otwornic różnił się między fiordami i obszarami otwartego morza. Wpływ na to miały przede wszystkim warunki oceanograficzne, zwłaszcza obecność poszczególnych mas wodnych, zwłaszcza ciepłej i zasolonej wody atlantyckiej. Na podstawie uzyskanych danych zidentyfikowano liczne potencjalne molekularne wskaźniki tej masy wodnej. Badanie to dostarczyło pierwszych tak kompleksowych danych na temat bioróżnorodności otwornicowej w rejonie Svalbardu i przyczyniło się do lepszego opisania odpowiedzi zbiorowisk otwornicowych na gradienty środowiskowe w Arktyce.

W drugim badaniu (RA III) przeanalizowaliśmy zbiorowiska otwornic głębokowodnych, koncentrując się na ogromnej nieznanej różnorodności gatunkowej ujawnionej przez metabarkodowanie osadów powierzchniowych. W prezentowanym artykule wykorzystano unikalne sygnatury DNA otwornic do klasyfikacji taksonomicznej do tej pory

niezidentyfikowanych linii genetycznych otwornic głębokowodnych i opisania ich rozmieszczenia w oceanach. Analizie poddano zbiorowiska bentosowych otwornic ze Strefy Clarion-Clipperton we wschodniej części Oceanu Spokojnego, porównując ich bioróżnorodność z dostępnymi danymi z innych regionów zarówno głęboko- jak i płytkowodnych. W rezultacie zidentyfikowano 61 nowych linii genetycznych otwornic należących do 27 kladów filogenetycznych na podstawie unikalnych sygnatur DNA w hiperzmiennym regionie 37F genu 18S rRNA. Większość z nowych linii została również znaleziona w innych obszarach głębinowych, ale tylko kilka z nich pojawiło się w zbiorach danych z rejonów przybrzeżnych. Sugeruje to, że głębokowodne otwornice bentosowe tworzą unikalną grupę wysoce przystosowaną do środowiska w którym żyją i że migracja między siedliskami płytkiego i głębokiego morza jest stosunkowo ograniczona. Podejście oparte na genetycznych sygnaturach stanowi nowatorską metodę badania rozmieszczenia i ekologii otwornic głębokowodnych, biorąc pod uwagę obecnie mocno ograniczoną bazę danych referencyjnych. Może to być przydatne w przyszłych badaniach wykorzystujących dane z metabarkodowania otwornic do biomonitoringu środowiska lub rekonstrukcji paleoceanograficznych.

Ostatnia część niniejszej rozprawy (RA IV) podsumowuje spektakularne postępy w badaniach nad rekonstrukcją zmian ekosystemów morskich w przeszłości geologicznej przy użyciu kopalnego DNA osadowego (*seda*DNA). Aby podsumować badania nad morskim *seda*DNA w ostatnich dwóch dekadach, przeprowadzono systematyczny przegląd literatury, obejmujący 55 oryginalnych badań. Prace te obejmują zarówno badania mikroorganizmów planktonicznych i bentosowych (prokariotów i jednokomórkowych eukariotów), jak i organizmów należących do meio- i makrofauny, których DNA jest zdeponowane w osadach morskich. Ninejszy artykuł przeglądowy opisuje procesy tafonomiczne jakim podlega DNA w środowisku morskim, kluczowe kwestie związane z wykorzystaniem *seda*DNA w badaniach nad ekosystemami morskimi w przeszłości oraz aktualny stan wiedzy i zastosowań w badaniach *seda*DNA w środowisku morskim. Przewidujemy, że analizy *seda*DNA będą wkrótce rutynowo włączane do badań paleoceanograficznych, zapewniąc unikalny wgląd w zmiany bioróżnorodności w geologicznych skalach czasowych. Pomogą także w określeniu wpływu antropopresji oraz kierunku ewolucji ekosystemów morskich. Ciągły rozwój badań nad *seda*DNA może również pomóc w ustanowieniu i optymalizacji strategii ochrony zasobów morskich i ich zarządzaniu.

Podsumowując, niniejsza praca doktorska przedstawia różnorodne zastosowania metabarkodowania *e*DNA do badania przeszłych i współczesnych ekosystemów oraz podkreśla potencjał i ograniczenia tej metody. Dodatkowo, wyniki uzyskane w toku przedstawionych badań przyczynią się do lepszego poznania różnorodności głębokowodnych i polarnych otwornic. Zapewniają również wgląd w procesy powiązane z tafonomią *e*DNA morskich eukariontów. Jest to kluczowe dla prawidłowej interpretacji zapisów DNA w osadach morskich. Jak wykazano w niniejszej rozprawie, wykorzystanie metabarkodowania *e*DNA ma kluczowe znaczenie dla dalszego rozwoju badań nad przeszłą i współczesną bioróżnorodnością morską.

SUMMARY in English

I. INTRODUCTION

1. General background

Earth's climate is constantly changing affecting the distribution of species and their evolution. Marine sediments are an important natural archive that could provide records of these changes over geological time scales. Recently, the environmental DNA (*e*DNA) in marine sediments has provided a new effective tool to investigate biodiversity, which allows multiple taxonomic groups to be analyzed simultaneously. The *e*DNA has been widely used in research focused on biomonitoring and biodiversity assessments. By capturing genetic traces preserved in marine sediments, the *e*DNA allows large-scale biodiversity investigations from present-day to millions of years (Ruppert et al., 2019; Pawlowski et al., 2022). In modern environments, sedimentary *e*DNA surveys provide a unique insight into the biodiversity and distribution patterns of hundreds of deep-sea species (Lejzerowicz et al., 2021; Cordier et al., 2022). At the longer timescales, marine sedimentary ancient DNA (*seda*DNA) research facilitates the reconstruction of ocean paleocommunities (De Schepper et al., 2019; Siano et al., 2021; Barrenechea Angeles et al., 2023).

However, with technological improvements, particularly the development of high-throughput sequencing, multiple issues have arisen that need to be considered when generating and analyzing metabarcoding data. Among these issues, it is worth mentioning the *e*DNA taphonomy in marine sediments, the incompleteness of reference databases, the optimization of analytical protocols for taxonomic assignment, and the search for *e*DNA-based ecological proxies. This thesis joins multiple research domains including molecular biology, taxonomy, protistology, ecology, and paleoceanography, in order to explore the potential and reliability of *e*DNA metabarcoding approaches.

2. Purpose of research

The aim of this thesis is to use *e*DNA metabarcoding to explore eukaryotic diversity from the surface to the bottom of the ocean. The thesis comprises four chapters, whose specific objectives are:

1) to investigate the eukaryotic diversity of Nordic Seas and to evaluate how well plankton diversity is recorded in sedimentary DNA, through:

- analysis of spatial and temporal variability of planktonic and benthic eukaryotic assemblages in water and sediment samples;
- analysis of the richness and abundance of planktonic DNA in marine sediments and identifying the potential biases influencing their representation in sedimentary records.

2) to investigate the foraminiferal diversity and its response to environmental changes in the fjords and open-water areas of Svalbard, by:

- comparing foraminiferal biodiversity in sieved and unsieved sediment samples;
- assessing foraminiferal communities' responses to environmental gradients in fjord and open-water environments;
- identifying potential foraminiferal bio-indicators of water masses in coastal Svalbard.

3) to classify the unknown deep-sea foraminiferal lineages and investigate their distribution, by:

- identifying the unique DNA signatures of unassigned foraminifera metabarcodes from the Clarion-Clipperton Fracture Zone area;
- investigating the distribution of novel foraminiferal lineages in other deep-sea and shallow-water regions.

4) to review the potential of marine sedimentary ancient DNA for reconstructing past biodiversity and environmental changes.

II. METHODOLOGY

1. Study sites and sample collection

The study area (RA I and II) covers the Nordic Seas and stretches from Svalbard, over the east Greenland shelf, to the north Norwegian shelf. The Nordic Seas are one of the key regions for studying the impact of climate changes on marine ecosystems. Our sampling stations cover a range of marine environments, from shelf to coastal open water and fjord areas. The oceanography of the Nordic Seas is shaped mainly by the interplay between warm and saline Atlantic Water (AW) and cold Arctic Water (ArW) (Cottier et al., 2005; Hop et al., 2019). Because of these specific settings, marine ecosystems in the Nordic Seas have represented one of the most important environmental settings for ecological and paleoclimatic studies.

We also applied DNA metabarcoding for deep-sea sediments from the Clarion-Clipperton Fracture Zone (CCFZ, RA III). The CCFZ is a vast swath of the abyssal Pacific Ocean seabed with water depths ranging from 4000 m in the east to 6000 m in the west (Rabone et al., 2023). This region is well-known for the abundance of polymetallic nodules that are subject to potential

industrial exploration and seabed mining (Lodge et al., 2014). However, the area also hosts numerous organisms adapted to extreme depths and particular conditions (Gooday et al., 2020; Rabone et al., 2023). Efforts to understand the unique biodiversity of the CCFZ are crucial for safeguarding this unique and unexplored ecosystem.

Overall, we collected samples of seawater and sediment from the Nordic Sea and deep-sea sediments from CCFZ to perform a metabarcoding analysis targeting micro-eukaryotic and foraminiferal communities. In the Nordic Seas, the samples were collected during cruises with R/V *Kronprins Haakon* in November 2020 (CAGE20-8), July 2021 (KH21-234), and with R/V *Oceania* in August 2016, 2021 (AREX). The Pacific sediment samples were collected with RESOURCE Cruise-01 in March 2020 from the Ocean Mineral Singapore areas of the CCFZ. In each sampling station, CTD profiles were obtained using a Mini CTD Sensordata SD202 or CTD Rosette, if applicable.

In brief, seawater samples were retrieved with the CTD Rosette or Niskin bottles from three water depths, i.e., the surface layer (~5 m), ~100 m water depth, and the near-bottom water (approximately 10 m above the sea floor) or from 1000 m at stations with water depth greater than 1000 m. Seawater (from 2 to 5 liters) was filtered directly on board using a sterile Sterivex filter unit (0.22 μm, Millipore, USA), with three units for each water depth. Negative controls were prepared by filtering 1 L of Milli-Q water or pressing air through the filter. Each filter was placed in an individual plastic bag and frozen at -20 °C. Surface sediment samples (0-1 cm) were collected using a multicore or box-core and stored in 50 ml sterile Falcon tubes, with three replicates (approximately 10 g each) for each station, respectively. Sediment samples were stored at -20 °C prior to further analyses.

2. DNA extraction, amplification, and sequencing

Total DNA was extracted from seawater samples using the DNeasy PowerWater Kit (Qiagen, Germany) and the QIAvac Vacuum Systems (Qiagen, Germany). The sediment DNA was extracted using DNeasy PowerSoil (Qiagen, Germany) for small samples (0.5 g – 1 g) and DNeasy PowerMax Soil (Qiagen, Germany) for samples up to 10 g. 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 reagents and to control contamination.

The small subunit ribosomal gene (SSU rRNA) or 18S rRNA in eukaryotes was amplified with different combinations of primers depending on the targeted group of organisms and the research question. The V9 region of the 18S rRNA gene (~130 bp length) was amplified by

Polymerase Chain Reaction (PCR) using universal primers 1389F/1510R to obtain eukaryotic amplicons (Amaral-Zettler et al., 2009). The 37f hypervariable region of the 18S rRNA gene (~68 - 196 bp) was amplified with foraminifera-specific primers s14F1/s15 to obtain foraminifera amplicons (Barrenechea Angeles et al., 2020). The primers were tagged with a unique 8-nucleotide sequence at the 5′ ends. Each sample was amplified in triplicate and each PCR reaction was performed in a total volume of 25 µL using Taq DNA polymerase and reaction buffer containing $MgCl_2$ solution, deoxynucleotide triphosphates, bovine serum albumin, reverse and forward primer, DNA-free water, and DNA from each sample as template. The primer sequences and PCR conditions for the amplification are summarized in Table 1. A PCR-negative control for each unique combination of tag-encoded primers was verified by agarose gel electrophoresis. The PCR products were purified and pooled in equimolar concentration within each multiplexed library. The preparation and quantification were then performed using commercial kits and quantitative PCR, following the manufacturer's instructions. The libraries were sequenced on a NovaSeq 6000 or MiSeq instrument (Illumina, USA) in paired-end reading mode 2×150 cycles.

Table 1. Primers and PCR thermal cycling profiles were used in this thesis.

Target gene and amplicon size	18S-V9 rRNA, ~ 130 bp			18S-37f rRNA, ~ 180 bp		
Primer name and sequence	1389F (5′- XXXXXXXX TTG TAC ACA CCG CCC -3′)			s14F1 (5′- XXXXXXXX AAG GGC ACC ACA AGA ACG C-3′)		
	1510R (5'- XXXXXXXX CCT TCY GCA GGT TCA CCT AC -3')			s15 (5'- XXXXXXXX CCT ATC ACA YAA TCA TG-3')		
Thermal cycling	1 cycle	95°C	5 min	1 cycle	94°C	1 min
	35 cycles	95°C	30 sec	35 cycles	94°C	30 sec
		57°C	45 sec		52°C	30 sec
		72°C	45 sec		72°C	30 sec
	1 cycle	72°C	5 min	1 cycle	72°C	2 min

3. Data quality control and analysis

Bioinformatics analyses were mainly performed using the web application SLIM (Dufresne et al., 2019) on the IOPAN server. The reads were first demultiplexed using the *double-tag-demultiplexing* algorithm based on their unique barcode sequences. Software packages such as VSEARCH (Rognes et al., 2016), DADA2 (Callahan et al., 2016), and several modules implemented in SLIM were used for quality trimming and filtering sequences, dereplicating sequences, merging of forward and reverse sequences, detection, and removal of chimeras, and

inferring operational taxonomic units (OTUs) or amplicon sequence variants (ASVs). Subsequently, all resulting OTUs/ASVs tables were curated with the LULU algorithm (Frøslev et al., 2017) to remove erroneous OTUs/ASVs with default settings if necessary. The final fasta files contained all the OTUs/ASVs sequences and their distribution in samples. In addition, to clean up the datasets from artifacts and extraneous sequences we check for the presence of specific signatures. For the eukaryotic datasets, we keep only sequences having the "GTCG" motif at the 5′ end of the eukaryotic 18S-V9 fragment regions. In the case of foraminiferal datasets, we keep sequences having the "GACAG" motif at the 5′ end and "TAGTCCCTT"/"TAGTCCTTT" motif at the 3' end of the foraminiferal 18S-37F regions. The final quality filtering of OTUs/ASVs involved the removal of unique and rare OTUs/ASVs.

The remaining OTUs/ASVs were compared to our local database of benthic foraminiferal 18S rDNA sequences (in prep.), PFR2 - Planktonic Foraminifera Ribosomal Reference database v. 1 (Morard et al., 2015), and PR2 - Protist Ribosomal Reference database v4.11.1 (Guillou et al., 2013), and custom function annotations of PR2-V9 using VSEARCH or BLASTN (Camacho et al., 2009) for taxonomic assignment. Finally, we discarded any prokaryotic, parasitic, or other OTUs/ASVs based on the target of each study.

Data analysis and visualization were performed with the R v.4.2.2 programming language (R Core Team, 2013). Microbial community diversity analyses were performed using functions of the *vegan* package v.2.6-4 (Oksanen et al., 2019). Additionally, several R packages were used with specific parameter settings, such as *venn* (Dusa, 2018), metagenome-Seq Bioconductor (Paulson et al., 2013), *mixOmics* (Rohart et al., 2017), *ggpubr* (Kassambara, 2020), *mgcv* (Wood, 2001), *ggplot2* (Wickham, 2011), *pheatmap* (Kolde, 2019), UpsetR (Gehlenborg, 2019), and others.

4. Unassigned foraminifera identification and phylogenetic analysis

In the case of CCFZ study, we prepared a subset of the CCFZ dataset including OTUs that could not be assigned by VSEARCH as well as those that VSEARCH assigned to ENFOR (environmental sequences) or Monothalamea X. After strict filtering steps, OTUs were aligned with reference sequence using the E-INS-i iterative refinement method in MAFFT (Katoh et al., 2017) and identified specific patterns to define new lineages. A phylogenetic tree including new lineages and reference sequences was built using the IQ-TREE maximum likelihood method (Trifinopoulos et al., 2016) with 1000 bootstrap replicates. Phylogenetic tree visualization and annotation were performed using the R package *ggtree* (Yu et al., 2017).

III. MAIN RESULTS OF RESEARCH ARTICLES

1. Research Article I

Taxonomic and abundance biases affect the record of marine eukaryotic plankton communities in sediment DNA archives. Ngoc-Loi Nguyen, Joanna Pawłowska, Marek Zajaczkowski, Agnes Weiner, Tristan Cordier, Danielle Grant, Stijn De Schepper and Jan Pawłowski. Submitted to *Molecular Ecology Resources.*

Marine sediments are excellent marine biodiversity repositories, and DNA preserved in sedimentary records serves as an inexhaustible source of information about modern and past ecosystems. However, very little is known about the taphonomy of environmental DNA, i.e. the accumulation and preservation of *e*DNA in sedimentary archives. Particularly for plankton taxa, it remains unclear whether their DNA is transferred from the water column to underlying sediments without distorting its taxonomic composition and community structure. In this research article, we address these issues by comparing the eukaryotic diversity in 270 *e*DNA samples from three water depths and the sediments from the same 24 stations in the Nordic Seas. Analysis of 18S-V9 metabarcoding data reveals different taxonomic compositions between water and sediment *e*DNA. Only 40% of the ASVs detected in water were also found in sediment DNA. Remarkably, the ASVs shared between water and sediments accounted for 80% of total sequence reads, suggesting that a large amount of plankton DNA is transported to the seafloor sediment. Thus, plankton DNA dominates in sediments in terms of relative abundance but not richness. Most of the planktonic DNA originates from phytoplankton blooms, predominantly diatoms. However, there are some planktonic groups, especially of pico- and nanoplankton (Picozoa or Prymnesiophyceae) that are diverse and abundant in the water layers but have rarely been detected in sediment samples. Overall, these results suggest that the genetic composition and structure of the plankton community changes considerably throughout the water column, with only fractions of some planktonic groups reaching and accumulating in the sediments. Therefore, the potential incompleteness of plankton records will be important to consider when interpreting sedimentary DNA archives to infer current and past marine biodiversity, especially when using *seda*DNA as a proxy to reconstruct past marine environmental conditions.

2. Research Article II

Metabarcoding reveals high diversity of benthic foraminifera linked to water masses circulation at coastal Svalbard. Ngoc-Loi Nguyen, Joanna Pawłowska, Inès Barrenechea Angeles, Marek Zajaczkowski and Jan Pawłowski. Published in *Geobiology*, 21(1), 133–150 (2023).

The Arctic Ocean is experiencing significant environmental changes, including the phenomenon of "atlantification", which is the increasing influence of warm and saline Atlantic water, leading to sea ice retreat, increased sea surface temperatures, and altered entire ecosystem dynamics. The impact of these changes on biodiversity and distribution across multitrophic levels has been demonstrated, but much less is known about their impact on protist biodiversity. In this research article, we focused on benthic foraminifera, a group of protists that are commonly used as ecological indicators. We conducted DNA metabarcoding of sieved and unsieved marine sediments from fjords and open sea areas in the Svalbard Archipelago to assess foraminiferal biodiversity. The effect of sieving sediment samples prior to the extraction of DNA and metabarcoding analysis was tested for 15 sampling stations. We obtained 4,836,419 reads in a sieved dataset and 742,783 reads in an unsieved dataset, which represented 1384 ASVs. The sieved and unsieved samples shared 73.91% of ASVs, comprising over 97% of reads with a slight difference in the foraminiferal compositions. However, no significant differences in alpha (Shannon and Simpson's indices) and beta diversity were observed between the datasets. Unassigned ASVs account for more than half of the DNA sequencing results, from both sieved and unsieved sediments, revealing a high genetic novelty of Svalbard foraminifera. The taxonomic composition of foraminiferal communities varies across five sampling areas, with diversity and species richness increasing from glacier-proximal/inner to glacier-distant/outer stations. Our study highlights the influence of different water masses on the structure of foraminiferal communities, notably the impact of Atlantic water in the Svalbard region. We identified potential molecular foraminiferal indicators of the Atlantic and Arctic water masses, whose efficacy, however, needs to be confirmed by further analyses. Our study emphasizes the significance of metabarcoding studies in assessing the impact of climate warming trends and associated oceanographic changes on Arctic benthic communities, particularly for monothalamous foraminifera not included in conventional morphology-based approaches.

3. Research Article III

Assigning the unassigned: a signature-based classification of rDNA metabarcodes reveals new deep-sea diversity. Inès Barrenechea Angeles, Ngoc-Loi Nguyen, Mattia Greco, Koh Siang Tan, and Jan Pawlowski. Published in *PLoS One*, 19(2), e0298440 (2024).

The development of high-throughput sequencing reveals a huge unknown environmental diversity of marine eukaryotes. However, most of this diversity remains unassigned due to a lack of functional and/or taxonomic information in public reference databases. The incompleteness of these databases, usually limited to particular taxonomic groups, specific genetic markers, and a few geographic regions is particularly challenging for analysis of metabarcoding datasets of deep-sea meiofauna and eukaryotic microbiota, which remains undescribed globally. In this research article, we address this issue by using unique foraminiferal DNA signatures to classify unassigned deep-sea lineages and investigate their worldwide distribution. We performed the metabarcoding analysis of 36 new deep-sea samples and other available datasets from the CCFZ area in the eastern-central Pacific. We identified 61 new lineages based on specific sequence patterns present at the beginning of the 37F hypervariable region of the 18S rRNA gene. The new lineages were placed in 27 phylogenetic clades, and their phylogenetic positions generally agreed with the signature assignment. Some new lineages were found in specific groups that are highly related to other CCFZ sequences from the database or formed a group on their own, with no closest reference-related sequences. Comparison of new lineages with other foraminiferal datasets from deep-sea and shallow-water regions shows that most novel lineages are widely distributed in the deep sea, but rarely occur at shallower depths. Although signature-based classification does not fill gaps in reference databases, it provides the unassigned sequences with a label that enables them to be included in future biodiversity or biogeography analyses.

4. Review Article IV

***Sedimentary ancient DNA: a new paleo-genomic tool for reconstructing the history of marine ecosystems*.** Ngoc-Loi Nguyen, Dhanushka Devendra, Natalia Szymańska, Mattia Greco, Inès Barrenechea Angeles, Agnes K. M. Weiner, Jessica Louise Ray, Tristan Cordier, Stijn De Schepper, Jan Pawłowski and Joanna Pawłowska. Published in *Frontiers in Marine Science* 10:1075 (2023).

This review article presents spectacular advances that have been made in reconstructing the history of marine ecosystems over geological timescales using sedimentary ancient DNA (*seda*DNA). By expanding the range of studied taxa beyond those preserved in the fossil record, this new approach transforms the way past ocean biodiversity can be analyzed. We summarized and discussed the current state of knowledge and potential methodological pitfalls and limitations in marine *seda*DNA research, providing important information for future research and methodology development.

First, we discussed the taphonomy of *e*DNA in marine environments, emphasizing the complexity of DNA preservation under dynamic marine conditions. We pointed out the main factors affecting DNA degradation, such as organic matter load, temperature, pH, salinity, water depth, and light intensity. Additionally, the role of physicochemical characteristics of sediments and environmental conditions in the preservation of *e*DNA on the seafloor was discussed. The lack of knowledge on the relationship between sediment properties (e.g., clay, borate, and organic content) and *seda*DNA preservation was highlighted.

Second, we critically evaluated the reliability of *seda*DNA data and the potential biases introduced by DNA preservation and extraction, highlighting challenges and opportunities for future research. We also discuss current applications of marine *seda*DNA research, ranging from long-term reconstruction of past biodiversity change as a result of climate change to monitoring the impact of anthropogenic activities on recent biodiversity change. We emphasized the potential of marine *seda*DNA for conservation purposes and the importance of interdisciplinary collaborations to better understand the causes and effects of changes in marine biodiversity. Overall, this review presents the marine *seda*DNA as a promising approach for studying the history of marine ecosystems and provides important guidelines for researchers to fully exploit its potential as a valuable tool for paleoecological studies.

IV. CONCLUSIONS AND RECOMMENDATION FOR FUTURE WORK

The thesis addresses a number of research questions that relate to the technical and analytical aspects of *e*DNA metabarcoding, as well as to the ecological interpretation of metabarcoding data. Its contribution to the field of eukaryotic metabarcoding consists of (i) demonstrating that sedimentary DNA archives provide only fragmentary records of marine biodiversity, especially regarding planktonic taxa, affected by strong taxonomic and abundance biases, (ii) confirm the usefulness of metabarcoding data as a source of new bioindicators of water masses, and (iii) propose how to overcome the limitations of current reference databases by classifying unassigned sequences into new lineages using molecular signatures. The most spectacular findings of the thesis include (i) the poor preservation of nano- and picoplanktonic eukaryotes in sedimentary DNA records, (ii) the high genetic diversity of Arctic monothalamous foraminifera, most of which remained unknown, and (iii) the specificity and wide-spread distribution of deep-sea foraminiferal lineages.

The thesis confirms that *e*DNA metabarcoding is a powerful tool for assessing present and past biodiversity in marine environments. However, more research is necessary to improve the generation and interpretation of metabarcoding data. Further studies are needed to better understand the complex processes involved in the taphonomy of planktonic and benthic DNA recorded in sediments. Additional survey time points and sediment samples from reference areas are needed to validate the potential eukaryotic indicators of water masses and their relationship to climate change. Finally, further research must continue to complete the reference database and identify representatives of the major lineages of foraminifera and other benthic eukaryotes, which diversity is largely undercharted. Overall, the results of this thesis provide a baseline for future *e*DNA-based studies on modern marine biodiversity, as well as a reference for the interpretation of historical *seda*DNA records.

STRESZCZENIE w języku polskim

I. WSTĘP

1. Ogólne informacje

Klimat Ziemi nieustannie się zmienia, wpływając na rozmieszczenie gatunków i ich ewolucję. Ważnym naturalnym archiwum tych zmian w geologicznej skali czasu są osady morskie. W ostatnich latach badania środowiskowego DNA (*e*DNA) zachowanego w osadach morskich dostarczyło nowego skutecznego narzędzia do badania różnorodności biologicznej, które pozwala na jednoczesną analizę wielu grup taksonomicznych. *e*DNA jest szeroko stosowane w badaniach koncentrujących się na biomonitoringu i ocenie bioróżnorodności. Poprzez przechwytywanie śladów genetycznych zachowanych w osadach morskich, *e*DNA umożliwia badania bioróżnorodności na dużą skalę od czasów współczesnych do milionów lat (Ruppert i in., 2019; Pawlowski i in., 2022). We współczesnych środowiskach badania *e*DNA osadowego zapewniają unikalny wgląd w różnorodność biologiczną i wzorce rozmieszczenia gatunków głębinowych (Lejzerowicz i in., 2021; Cordier i in., 2022). W dłuższych skalach czasowych badania morskiego kopalnego DNA osadowego (*seda*DNA) ułatwiają rekonstrukcję paleozbiorowisk organizmów morskich (De Schepper i in., 2019; Siano i in., 2021; Barrenechea Angeles i in., 2023)

Jednocześnie, wraz z postępem technologicznym, w szczególności rozwojem sekwencjonowania nowej generacji, pojawiło się wiele kwestii, które należy wziąć pod uwagę podczas generowania i analizowania danych z metabarkodowania próbek środowiskowych. Przede wszystkim warto wspomnieć o procesach tafonomicznych jakim podlega *e*DNA w środowisku morskim, niekompletności referencyjnych baz danych, optymalizacji protokołów analitycznych w celu taksonomicznej identyfikacji sekwencji lub poszukiwaniu wskaźników ekologicznych opartych na *e*DNA. Niniejsza rozprawa łączy wiedzę z wielu dziedzin: biologii molekularnej, taksonomii, ekologii i paleoceanografii, w celu zbadania potencjału i możliwości metod metabarkodowania *e*DNA.

2. Cele badań

Celem niniejszej rozprawy jest wykorzystanie metabarkodowania *e*DNA do zbadania bioróżnorodności organizmów eukariotycznych od powierzchni do dna oceanu. Praca składa się z czterech rozdziałów, których szczegółowymi celami są:

1) zbadanie różnorodności eukariotycznej Mórz Nordyckich i ocena, jak dobrze różnorodność organizmów planktonowych jest odzwierciedlona w DNA osadowym, poprzez:

- analizę przestrzennej i czasowej zmienności planktonowych i bentosowych zespołów eukariotycznych w próbkach wody i osadów;
- analizę wskaźników różnorodności biologicznej DNA planktonowego w osadach morskich oraz identyfikację potencjalnych czynników wpływających na ich występowanie w zapisach osadowych.

2) zbadanie bioróżnorodności zbiorowisk otwornic (foraminifera) i ich odpowiedzi na zmiany środowiskowe w rejonie Svalbardu - w fiordach i w obszarach otwartego morza, poprzez:

- porównanie bioróżnorodności otwornicowej w przesianych i nieprzesianych próbkach osadów;
- ocenę reakcji zbiorowisk otwornicowych na gradienty środowiskowe w środowiskach fiordów i wód otwartych;
- identyfikację potencjalnych otwornicowych bioindykatorów mas wodnych w fiordach i w rejonach szelfowych Svalbardu.

3) sklasyfikowanie nieznanych genetycznych linii otwornic głębokowodnych i zbadanie ich rozmieszczenia poprzez:

- identyfikację unikalnych sygnatur DNA niezidentyfikowanych gatunków otwornic z obszaru Pacyfiku (CCFZ – Clarion Clipperton Fracture Zone);
- zbadanie rozmieszczenia nowych linii genetycznych otwornic w innych regionach głęboko- i płytkowodnych.

4) analiza i opis potencjalnego zastosowania morskiego kopalnego DNA osadowego do rekonstrukcji zmian różnorodności biologicznej i środowiska w geologicznej przeszłości.

II. MATERIAŁ I METODY

1. Rejon badań i pobór prób

Rejon badań (RA I i II) obejmuje Morza Nordyckie i rozciąga się od Svalbardu, przez szelf wschodniej Grenlandii, aż do szelfu północnej Norwegii. Stacje poboru próbek obejmują szereg środowisk morskich, od rejonów szelfowych po przybrzeżne otwarte wody i fiordy. Morza

Nordyckie są jednym z regionów kluczowych dla badań wpływu zmian klimatu na ekosystemy morskie. Warunki oceanograficzne Mórz Nordyckich są kształtowane głównie przez wzajemne oddziaływanie ciepłej i słonej wody atlantyckiej (AW) i zimnej wody arktycznej (ArW) (Cottier i in., 2005; Hop i in., 2019). Z tego względu ekosystemy Mórz Nordyckich są niezwykle wrażliwe na zmiany klimatyczne i oceanograficzne, i tym samym stanowią jedno z najważniejszych miejsc do badań ekologicznych i paleoklimatycznych.

Badania były również prowadzone w Strefie Clarion-Clipperton (CCFZ, RA III). CCFZ to rozległy obszar dna Oceanu Spokojnego o głębokości wahającej się od 4000 m na wschodzie do 6000 m na zachodzie (Rabone i in., 2023). Region ten jest znany z częstego występowania konkrecji polimetalicznych, które stanowią przedmiot badań i poszukiwań ze względu na potencjalne zastosowania przemysłowe (Lodge i in., 2014). Obszar ten jest również siedliskiem licznych organizmów przystosowanych do życia w ekstremalnych środowiskach morskich głębin (Gooday i in., 2020; Rabone i in., 2023). Wysiłki mające na celu zrozumienie wyjątkowości biosfery rejonu CCFZ mają kluczowe znaczenie dla ochrony tego wyjątkowego i niezbadanego ekosystemu.

Próbki wody morskiej i osadów powierzchniowych z rejonu Mórz Nordyckich oraz osadów głębokowodnych z CCFZ zostały zebrane w celu przeprowadzenia metabarkodowania próbek środowiskowych, ze szczególnym uwzględnieniem zbiorowisk organizmów eukariotycznych, a w szczególności otwornic. Próbki z Mórz Nordyckich zostały pobrane podczas rejsów R/*V Kronprins Haakon* w listopadzie 2020 r. (CAGE20-8), lipcu 2021 r. (KH21-234) oraz w czasie rejsu R/V *Oceania* w sierpniu 2016 r. i 2021 r. (AREX 2016 oraz AREX 2021). Próbki osadów z Pacyfiku zostały pobrane podczas rejsu RESOURCE Cruise-01 w marcu 2020 r. z obszarów Ocean Mineral Singapore w CCFZ. Na każdej stacji poboru próbek wykonano pomiary CTD w kolumnie wody za pomocą sond Mini CTD Sensordata SD202 lub CTD Rosette.

Próbki wody morskiej zostały pobrane za pomocą rozety CTD (CTD Rosette) lub próbnika typu Niskin z trzech głębokości w kolumnie wody, tj. warstwy powierzchniowej (~5 m), ~100 m głębokości i warstwy naddennej (około 10 m nad dnem morskim) lub z 1000 m na stacjach o głębokości większej niż 1000 m. Bezpośrednio po pobraniu, próbki wody morskiej zostały przefiltrowane przy użyciu sterylnych filtrów Sterivex (0,22 µm, Millipore, USA). Dla każdej warstwy wykonano 3 powtórzenia, od 2 do 5 litrów wody każde. Kontrole negatywne przygotowano poprzez przefiltrowanie 1 l wody Milli-Q oraz powietrza. Każdy filtr został umieszczony w osobnej sterylnej plastikowej torebce i zamrożony w temperaturze -20 °C. Próbki osadów powierzchniowych (0-1 cm) zebrano za pomocą próbnika skrzynkowego (box

corer) lub próbnika typu multicore i przeniesiono do sterylnych probówek typu Falcon o pojemności 50 ml. Dla każdej stacji wykonano trzy powtórzenia po około 10 g osadu. Próbki osadów przechowywano w temperaturze -20 °C do momentu wykonania dalszych analiz.

2. Ekstrakcja DNA, amplifikacja i sekwencjonowanie

DNA zostało wyizolowane z próbek wody morskiej przy użyciu zestawu DNeasy PowerWater (Qiagen, Niemcy) i systemu próżniowego QIAvac (Qiagen, Niemcy). DNA z osadów morskich zostało wyizolowane przy użyciu DNeasy PowerSoil (Qiagen, Niemcy) dla małych próbek (0,5 g - 1 g mokrej masy) i DNeasy PowerMax Soil (Qiagen, Niemcy) dla większych próbek (do 10 g mokrej masy). Dla każdej z partii ekstrakcji wykonano negatywną kontrolę, niezbędną do kontrolowania czystości pomieszczenia i odczynników oraz zidentyfikowania obecności ewentualnych zanieczyszczeń.

Wybrane fragmenty DNA zlokalizowane w małej podjednosce rybosomalnej (SSU rRNA lub 18S rRNA) zostały zamplifikowane za pomocą reakcji łańcuchowej polimerazy (ang. *Polymerase Chain Reaction* - PCR) przy użyciu różnych kombinacji starterów w zależności od docelowej grupy organizmów i pytania badawczego. Region V9 genu 18S rRNA (o długości ~130 bp) amplifikowano przy użyciu uniwersalnych starterów 1389F/1510R (Amaral-Zettler i in., 2009). Hiperzmienny region 37f genu 18S rRNA (~68 - 196 bp) amplifikowano za pomocą starterów s14F1/s15 specyficznych dla foraminifera (Barrenechea Angeles i in., 2020). Każdy ze starterów został oznaczony unikalną 8-nukleotydową sekwencją na końcu 5′. Każdą próbkę amplifikowano w trzech powtórzeniach, a każdą reakcję PCR przeprowadzono w całkowitej objętości 25 µL przy użyciu polimerazy Taq DNA i buforu reakcyjnego zawierającego roztwór $MgCl_2$, trifosforany deoksynukleotydów, albuminę surowicy bydlęcej, startery, wodę jałową (wolną od DNA) oraz DNA matrycowego pochodzącego z poszczególnych próbek. Sekwencje starterów i warunki amplifikacji PCR podsumowano w Tabeli 1. Rezultaty reakcji PCR oraz negatywnych kontroli PCR dla każdej unikalnej kombinacji starterów zostały zweryfikowane za pomocą elektroforezy w żelu agarozowym. Produkty PCR oczyszczono i połączono w równomolowym stężeniu w pule. Przygotowanie bibliotek sekwencji i analizę ilościową przeprowadzono następnie przy użyciu komercyjnych zestawów i qPCR (ilościowy PCR), zgodnie z instrukcjami producenta. Biblioteki sekwencjonowano na instrumencie NovaSeq 6000 lub MiSeq (Illumina, USA) w trybie odczytu paired-end 2×150 cykli.

3. Analiza bioinformatyczna

Analizy bioinformatyczne przeprowadzono głównie przy użyciu aplikacji SLIM (Dufresne i in., 2019) zainstalowanej na serwerze IOPAN. Odczyty sekwencji zostały najpierw zdemultipleksowane przy użyciu odpowiedniego algorytmu w oparciu o unikalne sekwencje znaczników dołączonych do każdego startera. Pakiety oprogramowania, takie jak VSEARCH (Rognes i in., 2016), DADA2 (Callahan i in., 2016), a także kilka modułów wchodzących w skład SLIM wykorzystano do analizy jakości sekwencji oraz ich filtrowania, dereplikacji, wykrywania i usuwania chimer oraz łączenia sekwencji w tzw. operacyjne jednostkach taksonomicznych (OTU; Operational Taxonomic Unit) lub warianty sekwencji amplikonów (ASV; Amplicon Sequence Variants). Następnie wszystkie bazy danych OTU/ASV zostały poddane przeanalizowane za pomocą algorytmu LULU (Frøslev i in., 2017) aby usunąć błędne OTU/ASV. Ostatecznie otrzymano pliki fasta zawierające wszystkie sekwencje OTU/ASV i ich rozkład w próbkach. Ponadto, aby oczyścić zbiory danych z artefaktów i obcych sekwencji, dane zostały sprawdzone pod kątem obecności określonych sygnatur DNA. W przypadku eukariotycznych zbiorów danych zostały zachowane tylko sekwencje posiadające motyw "GTCG" na końcu 5′ eukariotycznego regionu 18S-V9. W przypadku otwornicowych zbiorów danychzachowano sekwencje posiadające motyw "GACAG" na końcu 5′ i motyw "TAGTCCCTT"/"TAGTCCTTT" na końcu 3' regionu 18S-37F. Ostateczne filtrowanie danych obejmowało usunięcie unikalnych i rzadkich OTU/ASV.

Pozostałe OTU/ASV porównano z bazą danych sekwencji 18S rDNA otwornic bentosowych (w przygotowaniu), PFR2 - Planktonic Foraminifera Ribosomal Reference database v. 1 (Morard i in., 2015), PR2 - Protist Ribosomal Reference database v4.11.1 (Guillou i in., 2013), oraz niestandardowych funkcji PR2-V9 przy użyciu VSEARCH lub BLASTN (Camacho i in., 2009). Pozwoliło to na oznaczenie sekwencji DNA do możliwie najniższego poziomu taksonomicznego. Sekwencje należące do organizmów prokariotycznych i pasożytniczych zostały usunięte z finalnej bazy danych.

Analizę i wizualizację danych wykonano przy użyciu języka programowania R v.4.2.2 (R Core Team, 2013). Analizy bioróżnorodności zbiorowisk eukariontów przeprowadzono przy użyciu pakietu *vegan* v.2.6-4 (Oksanen i in., 2019). Dodatkowo wykorzystano pakiety R takie jak venn (Dusa, 2018), metagenome-Seq Bioconductor (Paulson i in., 2013), *mixOmics* (Rohart i in., 2017), *ggpubr* (Kassambara, 2020), *mgcv* (Wood, 2001), *ggplot2* (Wickham, 2011), *pheatmap* (Kolde, 2019), UpsetR (Gehlenborg, 2019), i inne.

4. Analiza sekwencji DNA i filogenza otwornic

W przypadku analizy próbek z CCFZ wyodrębniono podzbiór danych obejmujący OTU, których nie można było oznaczyć taksonomicznie przy użyciu VSEARCH, a także te, które VSEARCH przypisał do ENFOR (sekwencje środowiskowe) lub Monothalamea X. Po rygorystycznych etapach filtrowania, OTU zostały dopasowane do sekwencji referencyjnej przy użyciu metody E-INS-i w MAFFT (Katoh i in., 2017). Ponadto, zidentyfikowano określone charakterystyczne fragmenty sekwencji DNA otwornic w celu zdefiniowania nowych linii genetycznych. Drzewo filogenetyczne obejmujące nowe linie i sekwencje referencyjne zostało zbudowane przy użyciu metody maksymalnego prawdopodobieństwa IQ-TREE (Trifinopoulos i in., 2016). Wizualizację drzewa filogenetycznego przygotowano przy użyciu pakietu R *ggtree* (Yu i in., 2017).

III. PRZEGLĄD WYNIKÓW PRZEDSTAWIONYCH W POSZCZEGÓLNYCH PUBLIKACJACH

1. Artykuł I

Taxonomic and abundance biases affect the record of marine eukaryotic plankton communities in sediment DNA archives. Ngoc-Loi Nguyen, Joanna Pawłowska, Marek Zajaczkowski, Agnes Weiner, Tristan Cordier, Danielle Grant, Stijn De Schepper, i Jan Pawłowski. Wysłano do *Molecular Ecology Resources.*

Osady morskie stanowią doskonałe repozytoria bioróżnorodności morskiej, a DNA zachowane w osadach może służyć jako niewyczerpane źródło informacji o współczesnych i przeszłych ekosystemach. Niewiele jednak wiadomo na temat procesów tafonomicznych jakim podlega środowiskowe DNA w kolumnie wody i w osadach, a zwłaszcza akumulacji i zachowania *e*DNA w osadach w geologicznej skali czasu. Szczególnie w przypadku organizmów planktonowych pozostaje niejasne, czy ich DNA jest transportowane z kolumny wody do osadów z pełnym odzwierciedleniem składu taksonomicznego i struktury zbiorowisk. Aby odpowiedzieć na najważniejsze pytania dotyczące zachowania DNA planktonowego w osadach morskich, porównano bioróżnorodność zbiorowisk eukariotycznych w 270 próbkach *e*DNA pobranych z trzech warstw kolumny wody i osadów powierzchniowych z 24 stacji zlokalizowanych w Morzach Nordyckich. Metabarkodowanie wody i osadów z wykorzystaniem fragmentu 18S-V9 ujawniło zbiorowiska znacząco różne pod kątem składu taksonomicznego. Tylko 40% ASV wykrytych w wodzie znaleziono również w DNA

osadowym. Co ciekawe, ASV występujące zarówno w wodzie, jak i w osadach stanowiły 80% wszystkich odczytów sekwencji, co sugeruje, że duża ilość DNA planktonowego jest transportowana do osadów dna morskiego. DNA organizmów planktonowych dominuje zatem w osadach pod względem udziału procentowego sekwencji, ale nie bogactwa gatunkowego. Większość DNA planktonowego zachowanego w osadach pochodzi z zakwitów fitoplanktonu, głównie okrzemek. Istnieją jednak pewne grupy planktonu, zwłaszcza piko- i nanoplanktonu (Picozoa lub Prymnesiophyceae), które występują licznie we wszystkich warstwach wody, ale rzadko były wykrywane w próbkach osadów. Wyniki te sugerują, że skład taksonomiczny i struktura zbiorowisk planktonowych zmienia się znacznie w całej kolumnie wody, ale tylko nieliczne grupy organizmów żyjących w toni wodnej są zachowane w osadach dennych. Jest to niezwykle istotne w kontekście interpretacji zapisów kopalnego DNA osadowego i potencjalnych błędów wynikających z niekompletnego zapisu organizmów planktonowych. Dane dotyczące organizmów planktonowych powinny być traktowane z należytą ostrożnością, zwłaszcza w przypadku wnioskowania o obecnej i przeszłej bioróżnorodności morskiej.

2. Artykuł II

Metabarcoding reveals high diversity of benthic foraminifera linked to water masses circulation at coastal Svalbard. Ngoc-Loi Nguyen, Joanna Pawłowska, Inès Barrenechea Angeles, Marek Zajaczkowski, i Jan Pawłowski. Opublikowano w *Geobiology,* 21(1), 133–150 (2023).

Ocean Arktyczny podlega znaczącym przemianom klimatycznym i środowiskowym, w tym tzw. "atlantyfikacji", czyli intensywnemu napływowi ciepłej i słonej wody atlantyckiej. Prowadzi to do m.in. do zmniejszania się zasięgu lodu morskiego, wzrostu temperatury powierzchniowej morza i zmian dynamiki całego ekosystemu. Wykazano wpływ tych zmian na bioróżnorodność i rozmieszczenie wielu gatunków z różnych poziomów troficznych, jednak stosunkowo niewiele wiadomo o wpływie atlantyfikacji na bioróżnorodność mikroorganizmów, w tym protistów. W niniejszej pracy przeanalizowano zmienność przestrzenną zbiorowisk otwornic bentosowych w kontekście warunków środowiskowych w rejonie Svalbardu. Przeprowadzono metabarkodowanie DNA przesianych i nieprzesianych próbek osadów morskich z fiordów i obszarów otwartego morza w archipelagu Svalbard. Wpływ przesiewania próbek osadów na wyniki metabarkodowania został przetestowany dla 15 stacji badawczych. Uzyskano 4 836 419 odczytów sekwencji z przesiewanych próbek i 742 783 odczyty z nieprzesiewanych próbek. Znacząca większość ASV i odczytów sekwencji (odpowiednio 73,91% i 97%) została zarejestrowana zarówno w bazie danych z przesianych,

jak i nieprzesianych próbek. Nie zaobserwowano znaczących różnic w różnorodności alfa (wskaźniki Shannona i Simpsona) i beta między zestawami danych. ASV, których nie udało się przypisać do żadnej z grup otwornic stanowiły ponad 50%, co może sugerować obecność nieznanych linii genetycznych. Skład taksonomiczny zbiorowisk otwornic różnił się znacznie między miejscami poboru próbek. Jednocześnie zanotowano wspólny trend zmian: różnorodność i bogactwo gatunkowe wzrastały w miarę oddalania się od lodowców, zarówno tych zlokalizowanych wewnątrz fiordów, jak i uchodzących do otwartego morza. Przeprowadzone badania wykazały wpływ różnych mas wodnych, w szczególności wody atlantyckiej, na strukturę zbiorowisk otwornic. Zaproponowano potencjalne genetyczne wskaźniki atlantyckich i arktycznych mas wodnych, jednak wymagają one potwierdzenia w dalszych badaniach. Przeprowadzone badania podkreślają znaczenie metabarkodowania w ocenie wpływu ocieplania klimatu i związanych z nim zmian oceanograficznych na arktyczne zbiorowiska otwornic, a zwłaszcza gatunków miękkookrywowych, zwykle pomijanych w konwencjonalnych analizach opartych na morfologii skorupek.

3. Artykuł III

Assigning the unassigned: a signature-based classification of rDNA metabarcodes reveals new deep-sea diversity. Inès Barrenechea Angeles, Ngoc-Loi Nguyen, Mattia Greco, Koh Siang Tan, i Jan Pawlowski. Opublikowano w *PLoS One*, 19(2), e0298440 (2024).

Niezwykle szybki rozwój metodologii i technologii sekwencjonowania nowej generacji pozwolił na ujawnienie niezwykle bogatych i dotychczas niezbadanych zbiorowisk morskich eukariontów. Mimo ogromnej bioróżnorodności, większość tych organizmów pozostaje niezidentyfikowana z powodu ograniczonej dostępności sekwencji DNA wielu grup w referencyjnych bazach danych. Niekompletność tych baz danych, zwykle ograniczonych do określonych grup taksonomicznych lub markerów genetycznych i kilku regionów geograficznych, znacznie utrudnia analizy zbiorów danych z metabarkodowania próbek środowiskowych, zwłaszcza w odniesieniu do mikroorganizmów głębinowych. W prezentowanym artykule wykorzystano unikalne sygnatury DNA otwornic do klasyfikacji taksonomicznej do tej pory niezidentyfikowanych linii genetycznych otwornic głębokowodnych i opisania ich rozmieszczenia w oceanach. Przeprowadziliśmy analizę danych z metabarkodowania 36 próbek osadów głębinowych oraz innych dostępnych zbiorów danych z obszaru CCFZ na środkowo-wschodnim Pacyfiku. Analiza tego zbioru danych pozwoliła na zidentyfikowanie 61 nowych linii genetycznych otwornic na podstawie specyficznych wzorców sekwencji (tzw. sygnatur) obecnych na początku hiperzmiennego regionu 37F 18S rRNA.

Nowe linie zostały przypisane do 27 kladów, a ich pozycja filogenetyczna zgadzała się z identyfikacją taksonomiczną wykonaną na podstawie sygnatur. Niektóre nowe linie były wysoce spokrewnione z innymi sekwencjami CCFZ z bazy danych, jednak zidentyfikowano również linie tworzące samodzielne grupy. Porównanie nowych linii genetycznych z innymi bazami otwornicowych sekwencji DNA z rejonów głęboko- i płytkowodnych pokazuje, że większość nowych linii jest szeroko rozpowszechniona w głębinach morskich, ale rzadko występuje na mniejszych głębokościach. Chociaż klasyfikacja oparta na sygnaturach nie wypełnia luk w referencyjnych bazach danych, jednak pozwala na zidentyfikowanie nieoznaczonych sekwencji, co umożliwia włączenie ich do przyszłych analiz bioróżnorodności lub badań biogeograficznych.

4. Artykuł IV

***Sedimentary ancient DNA: a new paleo-genomic tool for reconstructing the history of marine ecosystems*.** Ngoc-Loi Nguyen, Dhanushka Devendra, Natalia Szymańska, Mattia Greco, Inès Barrenechea Angeles, Agnes K. M. Weiner, Jessica Louise Ray, Tristan Cordier, Stijn De Schepper, Jan Pawłowski, i Joanna Pawłowska. Opublikowano w *Frontiers in Marine Science* 10:1075 (2023).

Niniejszy artykuł przeglądowy podsumowuje spektakularne postępy w badaniach nad rekonstrukcją zmian ekosystemów morskich w geologicznej przeszłości przy użyciu kopalnego DNA osadowego (*seda*DNA). Włączenie do analiz również tych gatunków, które nie są zachowane w zapisie kopalnym, całkowicie zrewolucjonizowało podejście do analiz bioróżnorodności morskiej w przeszłości. W prezentowanym artykule podsumowano i omówiono obecny stan wiedzy oraz potencjalne pułapki i ograniczenia metodologiczne w badaniach nad morskim *seda*DNA, dostarczając tym samym ważnych informacji dla przyszłych badań.

Pierwsza część artykułu skupia się głównie omówieniu procesów tafonomicznych jakim podlega *e*DNA w środowiskach morskich, podkreślając złożoność procesów zachowywania DNA w dynamicznych warunkach mórz i oceanów. Zostały w niej omówione główne czynniki wpływające na degradację DNA, takie jak dopływ materii organicznej, temperatura, pH, zasolenie, głębokość wody i natężenie światła. Dodatkowo omówiono rolę właściwości fizykochemicznych osadów i warunków środowiskowych w zachowaniu *e*DNA na dnie morskim. Podkreślono brak wiedzy na temat związku między właściwościami osadów (np. zawartością iłów, boranów i substancji organicznych) a zachowaniem *seda*DNA w dłuższych skalach czasowych.

W drugiej części artykułu poddano krytycznej ocenie wiarygodność danych *seda*DNA, uwzględniając potencjalne błędy wprowadzone przez metody analizy laboratoryjnej DNA oraz podkreślając wyzwania i możliwości związane z przyszłymi badaniami. Omówiono również obecne zastosowania analiz morskiego *seda*DNA, począwszy od rekonstrukcji zmian różnorodności biologicznej w przeszłości w wyniku zmian klimatu, aż po monitorowanie wpływu działań antropogenicznych na współczesne zmiany różnorodności biologicznej. Dzięki temu został podkreślony potencjał wykorzystania zapisów morskiego *seda*DNA dla celów ochrony przyrody jak i znaczenie interdyscyplinarnej współpracy w celu lepszego zrozumienia przyczyn i skutków zmian bioróżnorodności morskiej. Niniejszy przegląd przedstawia analizy morskiego *seda*DNA jako nowego obiecującego narzędzia do badania historii ekosystemów morskich i dostarcza ważnych wskazówek dla naukowców, chcących w pełni wykorzystać jego potencjał w badaniach paleoekologicznych.

IV. WNIOSKI

Niniejsza rozprawa doktorska stawia szereg pytań badawczych, które odnoszą się do technicznych i analitycznych aspektów metabarkodowania *e*DNA, a także do ekologicznej interpretacji danych. Najważniejsze osiągnięcia będące wynikiem niniejszej pracy to (i) wykazanie, że zapis osadowego DNA przedstawia jedynie część całkowitej morskiej różnorodności biologicznej, zwłaszcza w odniesieniu do organizmów planktonowych oraz może zawierać błędy (ii) potwierdzenie przydatności danych pochodzących z metabarkodowania próbek środowiskowych jako źródła nowych bioindykatorów mas wodnych oraz (iii) zaproponowanie sposobu na usunięcie obecnych ograniczeń metod identyfikacji taksonomicznej sekwencji DNA poprzez wykorzystanie sygnatur molekularnych. Najbardziej spektakularne odkrycia przedstawione w niniejszej rozprawie obejmują (i) udowodnienie słabego zachowanie nano- i pikoplanktonowych eukariontów w zapisach DNA osadowego, (ii) wykazanie niezwykle wysokiej różnorodności genetycznej zbiorowisk arktycznych otwornic miękkookrywowych (Monothalamea), do tej pory w większości niezbadanych, oraz (iii) wykazanie specyficzności oraz szerokiego rozpowszechniania w środowisku morskim nowych linii genetycznych otwornic głębokowodnych.

Wyniki i wnioski przedstawione w rozprawie doktorskiej potwierdzają, że metabarkodowanie *e*DNA jest przydatnym narzędziem do oceny obecnej i przeszłej różnorodności biologicznej w środowiskach morskich. Konieczne są jednak dalsze badania, aby

usunąć błędy powstające w trakcie generowania i interpretacji danych. Podkreślono potrzebę dalszych badań, aby lepiej zrozumieć złożone procesy związane z tafonomią planktonicznego i bentosowego DNA zachowanego w osadach morskich. Potrzebne są dodatkowe badania wielosezonowe oraz próbki osadów z różnych obszarów referencyjnych, aby zweryfikować potencjalne eukariotyczne wskaźniki mas wodnych oraz związek ich pojawiania się w zapisach osadowych ze zmianami klimatu. Co więcej, dalsze badania muszą być kontynuowane w celu uzupełnienia referencyjnej bazy danych i zidentyfikowania przedstawicieli głównych linii genetycznych otwornic i innych eukariontów bentosowych, których różnorodność jest w dużej mierze niezbadana. Wyniki tej pracy stanowią istotny punkt odniesienia dla przyszłych badań opartych na *e*DNA, zarówno tych dotyczących współczesnej bioróżnorodności morskiej, jak również interpretacji historycznych zapisów *seda*DNA.

V. REFERENCES / BIBLIOGRAFIA

Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W., et al. (2009). A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS One* 4(7)**,** e6372. doi: 10.1371/journal.pone.0006372.

Barrenechea Angeles, I., Lejzerowicz, F., Cordier, T., et al. (2020). Planktonic foraminifera eDNA signature deposited on the seafloor remains preserved after burial in marine sediments. *Sci. Rep.* 10(1)**,** 20351. doi: 10.1038/s41598-020-77179-8.

Barrenechea Angeles, I., Romero-Martínez, M.L., Cavaliere, M., et al. (2023). Encapsulated in sediments: eDNA deciphers the ecosystem history of one of the most polluted European marine sites. *Environ. Int.* 172**,** 107738. doi: 10.1016/j.envint.2023.107738.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., et al. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13(7)**,** 581-583. doi: 10.1038/nmeth.3869.

Camacho, C., Coulouris, G., Avagyan, V., et al. (2009). BLAST+: architecture and applications. *BMC Bioinform.* 10(1)**,** 421. doi: 10.1186/1471-2105-10-421.

Cordier, T., Angeles, I.B., Henry, N., et al. (2022). Patterns of eukaryotic diversity from the surface to the deep-ocean sediment. *Sci. Adv.* 8(5)**,** eabj9309. doi: 10.1126/sciadv.abj9309.

Cottier, F., Tverberg, V., Inall, M., et al. (2005). Water mass modification in an Arctic fjord through cross-shelf exchange: The seasonal hydrography of Kongsfjorden, Svalbard. *J. Geophys. Res. Oceans* 110(C12). doi: 10.1029/2004JC002757.

De Schepper, S., Ray, J.L., Skaar, K.S., et al. (2019). The potential of sedimentary ancient DNA for reconstructing past sea ice evolution. *ISME J.* 13(10)**,** 2566-2577. doi: 10.1038/s41396-019-0457-1.

Dufresne, Y., Lejzerowicz, F., Perret-Gentil, L.A., et al. (2019). SLIM: a flexible web application for the reproducible processing of environmental DNA metabarcoding data. *BMC Bioinform.* 20(1)**,** 88. doi: 10.1186/s12859-019-2663-2.

Dusa, A. (2018). Package 'venn'. *version 1.9*.

Frøslev, T.G., Kjøller, R., Bruun, H.H., et al. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nat. Commun.* 8(1)**,** 1188. doi: 10.1038/s41467-017-01312-x.

Gehlenborg, N. (2019). UpSetR: a more scalable alternative to Venn and Euler diagrams for visualizing intersecting sets. . *R package version 1.4. 0*.

Gooday, A.J., Schoenle, A., Dolan, J.R., et al. (2020). Protist diversity and function in the dark ocean – Challenging the paradigms of deep-sea ecology with special emphasis on foraminiferans and naked protists. *Eur. J. Protistol.* 75**,** 125721. doi: 10.1016/j.ejop.2020.125721.

Guillou, L., Bachar, D., Audic, S., et al. (2013). The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res,* 41(Database issue)**,** D597-604. doi: 10.1093/nar/gks1160.

Hop, H., Assmy, P., Wold, A., et al. (2019). Pelagic Ecosystem Characteristics Across the Atlantic Water Boundary Current From Rijpfjorden, Svalbard, to the Arctic Ocean During Summer (2010–2014). *Front. Mar. Sci.* 6(181). doi: 10.3389/fmars.2019.00181.

Kassambara, A. (2020). Package 'ggpubr'. *R package version 0.1* 6(0).

Katoh, K., Rozewicki, J., and Yamada, K.D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20(4)**,** 1160-1166. doi: 10.1093/bib/bbx108

Kolde, R. (2019). pheatmap: Pretty Heatmaps. *R package version 1.0. 12*.

Lejzerowicz, F., Gooday, A.J., Barrenechea Angeles, I., et al. (2021). Eukaryotic biodiversity and spatial patterns in the Clarion-Clipperton Zone and other abyssal regions: Insights from sediment DNA and RNA metabarcoding. *Front. Mar. Sci.* 8. doi: 10.3389/fmars.2021.671033.

Lodge, M., Johnson, D., Le Gurun, G., et al. (2014). Seabed mining: International Seabed Authority environmental management plan for the Clarion–Clipperton Zone. A partnership approach. *Mar. Policy* 49**,** 66-72. doi: 10.1016/j.marpol.2014.04.006.

Morard, R., Darling, K.F., Mahé, F., et al. (2015). PFR2: a curated database of planktonic foraminifera 18S ribosomal DNA as a resource for studies of plankton ecology, biogeography and evolution. *Mol. Ecol. Resour.* 15(6)**,** 1472-1485. doi: 10.1111/1755-0998.12410.

Oksanen, J., Blanchet, F.G., Friendly, M., et al. (2019). Package 'vegan'. *Community ecology package, version 2* 2(9).

Paulson, J.N., Stine, O.C., Bravo, H.C., et al. (2013). Differential abundance analysis for microbial marker-gene surveys. *Nat. Methods* 10(12)**,** 1200-1202. doi: 10.1038/nmeth.2658.

Pawlowski, J., Bruce, K., Panksep, K., et al. (2022). Environmental DNA metabarcoding for benthic monitoring: A review of sediment sampling and DNA extraction methods. *Sci. Total Environ.* 818**,** 151783. doi: 10.1016/j.scitotenv.2021.151783.

R Core Team, R. (2013). R: A language and environment for statistical computing.

Rabone, M., Wiethase, J.H., Simon-Lledó, E., et al. (2023). How many metazoan species live in the world's largest mineral exploration region? *Curr. Biol.* 33(12)**,** 2383-2396.e2385. doi: 10.1016/j.cub.2023.04.052.

Rognes, T., Flouri, T., Nichols, B., et al. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* 4**,** e2584. doi: 10.7717/peerj.2584.

Rohart, F., Gautier, B., Singh, A., et al. (2017). mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* 13(11)**,** e1005752. doi: 10.1371/journal.pcbi.1005752.

Ruppert, K.M., Kline, R.J., and Rahman, M.S. (2019). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Glob. Ecol. Conserv.* 17**,** e00547. doi: 10.1016/j.gecco.2019.e00547.

Siano, R., Lassudrie, M., Cuzin, P., et al. (2021). Sediment archives reveal irreversible shifts in plankton communities after World War II and agricultural pollution. *Curr. Biol.* 31(12)**,** 2682-2689.e2687. doi: 10.1016/j.cub.2021.03.079.

Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., et al. (2016). W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44(W1)**,** W232-W235. doi: 10.1093/nar/gkw256.

Wickham, H. (2011). ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* 3(2)**,** 180-185.

Wood, S.N. (2001). mgcv: GAMs and generalized ridge regression for R. *R news* 1(2)**,** 20-25.

Yu, G., Smith, D.K., Zhu, H., et al. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* 8(1)**,** 28-36. doi: 10.1111/2041-210X.12628.

V. FULL ARTICLE

1. Research Article I

Taxonomic and abundance biases affect the record of marine eukaryotic plankton communities in sediment DNA archives. Ngoc-Loi Nguyen, Joanna Pawłowska, Marek Zajaczkowski, Agnes Weiner, Tristan Cordier, Danielle Grant, Stijn De Schepper and Jan Pawłowski. Submitted to *Molecular Ecology Resources.*

ABSTRACT

Environmental DNA (*e*DNA) preserved in marine sediments offers a novel tool to reconstruct past biodiversity across geological time scales. However, little is known about how exactly marine biodiversity is recorded in sedimentary *e*DNA archives, especially regarding plankton taxa. Here, we address this question by comparing the eukaryotic diversity from three water depths and the surface sediments of 24 stations in the Nordic Seas. Analysis of 18S-V9 metabarcoding data reveals distinct eukaryotic assemblages between water and sediment *e*DNA. Only 40% of amplicon sequence variants (ASVs) detected in water were also found in sediment *e*DNA. Shared ASVs accounted for 80% of total reads, suggesting that a large amount of plankton DNA is transported to the seafloor, mainly from abundant phytoplankton. However, not all plankton taxa are equally archived on the seafloor. Diatoms dominated the plankton *e*DNA deposited in the sediment, whereas certain nano- and picoplankton taxa (Picozoa or Prymnesiophyceae) were underrepresented. Overall, these results suggest that the genetic composition and structure of the plankton community vary considerably throughout the water column and differ from what accumulates in the sediment. Hence, the interpretation of sedimentary *e*DNA archives should take into account potential taxonomic and abundance biases when reconstructing past changes in marine biodiversity.

INTRODUCTION

Marine sediments are widely used as archives of biological and environmental changes over geological time scales. Paleoceanographic and paleoclimatic studies commonly use microfossils preserved in sediments as proxies for past environmental conditions (Marino et al., 2022; Pados-Dibattista et al., 2022; Devendra et al., 2023). Recent advances in sedimentary ancient DNA (*seda*DNA) research have opened new perspectives to trace non-fossilizing organisms and obtain a more holistic overview of past biodiversity changes (Capo et al., 2022;

Nguyen et al., 2023). *Seda*DNA has been used to search for new paleo proxies (De Schepper et al., 2019; Pawłowska et al., 2020; Zimmermann et al., 2020), to reconstruct ancient marine ecosystems (Armbrecht et al., 2022; Zimmermann et al., 2023), to trace the evolution of marine species (Ellegaard et al., 2020), and to reconstruct preindustrial conditions (Siano et al., 2021; Barrenechea Angeles et al., 2023). These studies acknowledged that potential biases in the *seda*DNA data could originate from the processing of sedimentary DNA samples, especially during the extraction of environmental DNA (*e*DNA) or PCR amplification (Armbrecht et al., 2019; Pawlowski et al., 2022). Other additional biases in marine settings could come from the sinking of plankton organisms through the water column to the seafloor, which is influenced by a variety of abiotic factors, such as marine surface and bottom water currents, movement of water masses, or lateral sediment transport (Pedersen et al., 2015; Torti et al., 2015; Nguyen et al., 2023). There is little known about the taphonomy of *e*DNA in marine environments, in particular, what happens to the DNA of plankton organisms after their death and whether or not all plankton DNA ends up on the sea floor (Morard et al., 2017; Barrenechea Angeles et al., 2020; Parry et al., 2020; Sogawa et al., 2022). A recent study compared the taxonomic composition and structure of plankton communities present in the water column with sediment records at the same sites and demonstrated that only a fraction of the major plankton groups were preserved in the sediments (Armbrecht et al., 2023). However, other studies suggested that plankton species make up over 50% of the diversity found in sedimentary DNA (Lecroq et al., 2011; Pawlowski et al., 2011; Cordier et al., 2022).

Here, we carried out an extensive *e*DNA metabarcoding survey of Nordic Seas biodiversity by sampling at the surface (SW), 100 m depth (100mW), and bottom (BW) layers of the water column as well as the surface sediment (SED) at the same station. *e*DNA was extracted from a total of 273 samples from 24 stations and analyzed the eukaryotic community using the 18S-V9 metabarcoding approach. We compared the taxonomic structure and relative abundances of eukaryotic communities from water and sediment *e*DNA datasets and identified potential biases in the sinking and accumulation of plankton DNA on the seafloor.

MATERIAL AND METHODS

Study sites and sample collection

We analyzed samples collected during three independent cruises with R/V *Kronprins Haakon* in November 2020 (CAGE20-8) and July 2021 (KH21-234) and with R/V *Oceania* in August 2021 (AREX). In total, twenty-four stations were sampled (Fig. 1, Supplementary Table

S1). At each station, water samples were retrieved with the CTD Rosette or Niskin bottles from the surface layer (5 m, SW), 100 m water depth (100mW), and the near-bottom water (approximately 10 m above the sea floor, BW) or from 1000 m at stations with water depth greater than 1000 m. From 2 to 5 liters of seawater per depth were filtered directly onboard using sterile Sterivex filter units (Millipore, USA) with a pore size of 0.22 µm. Three filter units were used for each water depth. Two filter units were prepared as negative controls at each station, one by filtering 1 L of Milli-Q water and the other by pressing air through the filter 10 times with a 50-ml plastic syringe, respectively. We also collected three replicates (approximately 10 g each) of surface sediment samples (0-1 cm, SED) using a multicore or box-core and stored them in 50 ml sterile Falcon tubes. Both the water and sediment samples were placed in individual plastic bags and frozen at -20 °C until molecular analysis. In each sampling station, CTD profiles were obtained using a Mini CTD Sensordata SD202 or CTD Rosette at intervals of 1s.

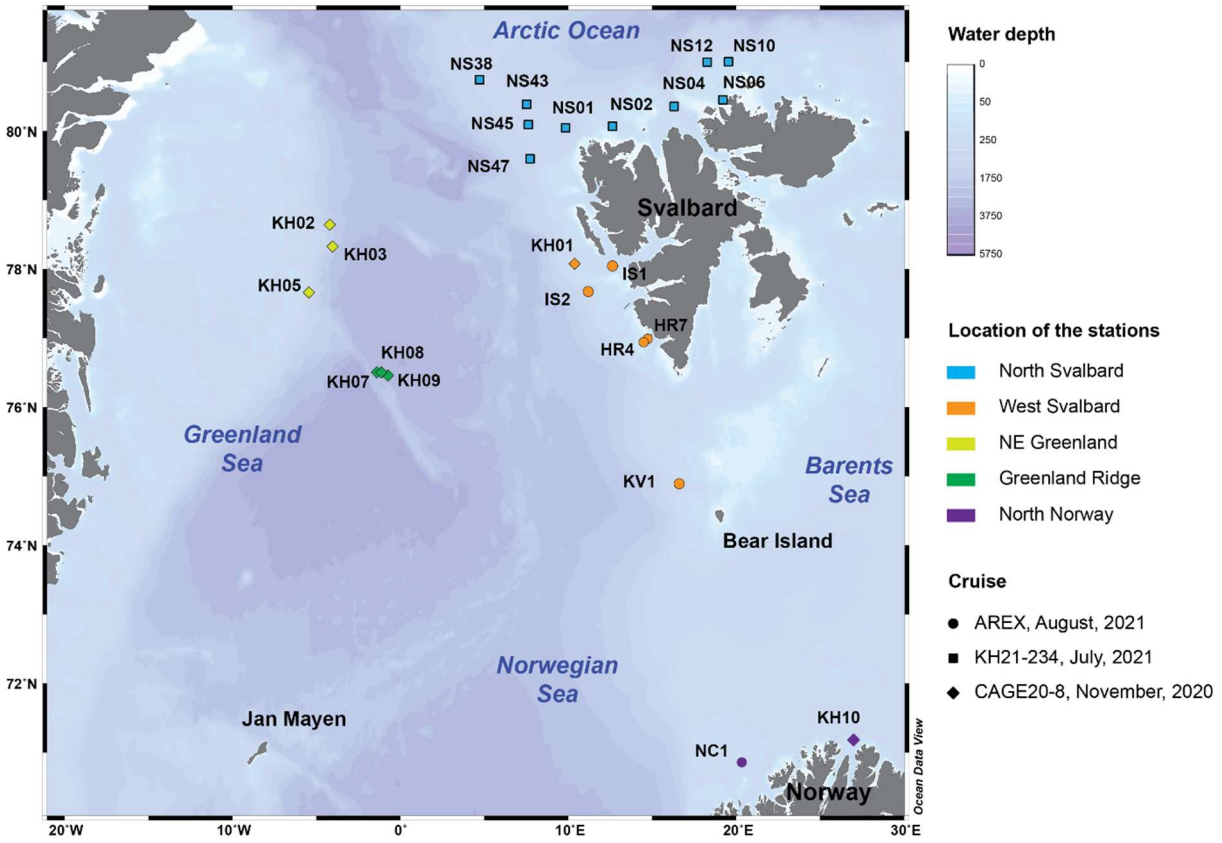


Figure 1. Map showing the location of sampling stations in the Nordic Seas. The color of the sampling stations indicated approximate correspondence with regions of the Nordic Seas, and the names and times of three different cruises are indicated by different symbols. More detail on sampling locations and obtained samples is given in Supplementary Table S1.

DNA extraction, amplification, and sequencing

DNA extractions were performed for each of the 204 seawater samples collected during three cruises, using the DNeasy PowerWater kit (Qiagen, Germany) and the QIAvac Vacuum Systems (Qiagen, Germany) according to the manufacturer's instructions. Negative controls consisting of in-laboratory and on-board blanks (i.e., filter units with air or Milli-Q water) were processed in the same procedure and in parallel with the samples. DNA was also extracted from 10 g of sediment using the DNeasy PowerMax Soil Kit (Qiagen, Germany) from a total of 69 (i.e., three replicates in 23 stations) surface sediment samples. Then all DNA extracts were stored at −20 °C until PCR amplifications.

The V9 region of the 18S rRNA gene (~130 bp length) was amplified by PCR using primers 1389F (5′-TTG TAC ACA CCG CCC-3′) and 1510R (5′-CCT TCY GCA GGT TCA CCT AC-3′) as designed in Amaral-Zettler et al. (2009), tagged with a unique 8- nucleotide sequence at the 5′ ends. Each sample was amplified in triplicate and each PCR reaction was performed in a total volume of 25 µL, which included 1.5 µL of 1.5 mM $MgCl_2$ (Applied Biosystems, USA), 2.5 µL of 10× PCR buffer II (Applied Biosystems, USA), 0.5 µL of 0.2 mM deoxynucleotide triphosphates (Promega, USA), 0.5 µL of 20 mg mL−1 bovine serum albumin (Invitrogen Ultrapure, USA), 1 µL of 10 µM of each primer, 0,2 µL of AmpliTaq Gold DNA polymerase (Applied Biosystems, USA) and 2 µL of template DNA. The conditions for the amplification consisted of a pre-denaturation step at 95 °C for 5 min, followed by 35 cycles of denaturation at 95 °C for 30 s, annealing at 57 °C for 45 s, and extension at 72 °C for 45 s, followed by a final extension step at 72 °C for 5 min. A PCR-negative control for each unique combination of tag-encoded primers was verified by agarose gel electrophoresis. PCR products were purified using the Clean-Up kit (A&A Biotechnology, Poland) and pooled in equimolar concentration within each multiplexed library. Libraries were then constructed using the KAPA HyperPrep Kit (KAPA Biosystems, USA) and the KAPA UDI Adapter Kit 15µM (KAPA Biosystems, USA), following the manufacturer's instructions. The libraries were quantified by quantitative PCR using the Kapa Library Quantification kit (Kapa Biosciences, USA) and sequenced on a NovaSeq 6000 instrument (Illumina, USA) in paired-end reading mode 2×150 cycles with Kit v1.5 (300 cycles) at the Center of New Technologies (CeNT, University of Warsaw, Poland).

Sequence data processing

Raw sequencing reads for each sample were processed using the web application SLIM (Dufresne et al., 2019). Briefly, raw reads were demultiplexed according to their unique tag in

the forward and reverse reads by using the module *demultiplexer*. Quality filtering, removal of chimera, and generation of the Amplicon Sequence Variant (ASV) table were performed using DADA2 v.1.16 (Callahan et al., 2019) with pseudo-pool parameters. Using the microDecon package with default settings (McKnight et al., 2019), the samples were decontaminated using negative control samples as reference. We then curated the ASV sequences at 97% similarity with the LULU package v.0.1.0 (Frøslev et al., 2017) with the default parameters as recommended in ref. (Brandt et al., 2021). Unique ASVs (occurring in only one sample) and rare ASVs (having <10 reads) were removed. We also excluded sequences lacking the "GTCG" motif, which is widely conserved at the 5′ end of eukaryotic 18S-V9 fragments (Cordier et al., 2022).

The assignment of the ASVs was done using VSEARCH against the taxonomically curated PR2 database v.4.14.1 (Guillou et al., 2013) with 85% similarity and customized function annotations of PR2-V9 (available at http://doi.org/10.5281/zenodo.3768951) with 95% similarity with up to three candidate reference sequences or at least 99% directly assigned to the reference sequence. We focused our analysis on the free-living eukaryotic diversity by discarding any prokaryotic and parasitic eukaryotic ASVs. Several species were grouped prior to constructing taxonomic plots to ensure consistency and ease of comparison between the taxa found in the water column and sediment samples, as shown in Supplementary Table S2.

Diversity and structural analysis

Data analysis was performed with R v.4.2.2 (R Core Team, 2013) using several R packages. All plots were created using the R package *ggplot2* v.3.4.2 (Wickham, 2011). We performed alpha and beta diversity analyses using functions of the vegan package v.2.6-4 (Oksanen et al., 2019). Before starting the analysis, we removed samples with less than 1000 reads throughout the dataset. ASV accumulation curves as a function of sampling effort were calculated with the *specaccum* function with the "random" method. The Shannon diversity for each sample was calculated and the distribution of sample diversity across both the water column and sediment samples was compared using the *stat_compare_means* function of the *ggpubr* package v.0.6.0 (Kassambara, 2020) with default settings.

For beta diversity analysis, we normalized the ASV-to-sample matrix with the cumulative sum scaling (CSS) method (Paulson et al., 2013) 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. The sampling depth variables were fitted to the NMDS as smooth surfaces using the *ordisurf* function. The dissimilarity matrix was

also used as input for the testing of the *adonis2* function for permutational multivariate analysis of variance (PERMANOVA) models and the *anosim* function for global one-way analysis of similarities (ANOSIM) for differences between eukaryotic compositional structure between sample types, sample layers, localities, and sampled times (month) using 999 permutations. The beta diversity dispersion also was measured within each sample type using the *betadisper* function and compared the distribution of the distance to group centroids between sample types.

To compare the community composition among sample layers, the unique and shared ASVs were visualized using the *venn* package (Dusa, 2018) and the UpsetR package (Gehlenborg, 2019). We calculated the normalized ASV richness and Shannon index per sample for the dataset excluding ASVs unique to the sediments and for selected plankton groups by rarefying each sample at the lowest remaining sequencing depth. The nonlinear relationship between the changes in richness and Shannon diversity along gradients of latitude, salinity, temperature, and water depth was analyzed using the *gam* function of the *mgcv* package (Wood, 2001).

RESULTS

Eukaryotic community structure and diversity

Processing of sequence data resulted in a total of 41,1 million high-quality DNA reads. After removing DNA sequences from samples with low read count (less than 1000 reads) and strict filtering, the final dataset comprised 9,987 ASVs and 37,624,065 reads (Supplementary Table S2). Among them, the SW, 100mW, BW, and SED samples comprised 26.5%, 25.5%, 20.5% and 27.4% of the sequence reads, respectively. The average number of sequences per sample was 426,093 reads. This dataset was used to analyze the alpha and beta eukaryotic diversity within and between stations and to infer taxonomic composition.

The ASV accumulation curves as a function of the sampling effort showed that the overall number of detected eukaryotic ASVs increased with depth (SW = 2385, 100mW = 2988, BW = 3853 ASVs), and was highest in the sediments (SED = 7419 ASVs) (Fig. 2A). A similar pattern was also observed for the Shannon diversity values (Fig. 2A, inset). Inter-sample similarities in ASV composition revealed a significant difference in eukaryotic communities between the water and surface sediments, illustrated by distinct clusters of water layer samples and their clear separation from sediment samples (Fig. 2B). This was confirmed by the two non-parametric tests of ANOSIM and PERMANOVA (Supplementary Table S3). The difference in the community composition of the sample types (water vs. sediment) and sample layers (SW, 100mW, BW, and SED) was greater than this between the sampling times (month) and

geographical locations (Supplementary Table S3). NMDS analyses also produced a similar biogeographical pattern of the eukaryotic communities in water layers and sediment samples, with a pronounced seasonal influence on the difference in communities of SW and 100mW at various sampling times compared to those in BW and SED (ANOMSIM, $p < 0.005$, Supplementary Fig. S1). The distance between the samples of the eukaryotic community within each group of water layers is significantly different from each other (Supplementary Fig. S2, PERMANOVA, $R2 = 0.26909$, $p < 0.001$). Taken together, these results demonstrate that the eukaryotic community present in the sediments is more diverse and has different structures and compositions than the communities present in the water samples.

The taxonomic composition of the eukaryotic communities was similar in all three water layers, with similar taxa observed for all stations (Fig. 2C), regardless of depth, as well as the combined dataset in lower taxonomic ranks (Supplementary Figs. S3-S4). In terms of ASV richness, the water samples were dominated by alveolates (primarily syndiniales ~ 22% and dinophytes ~10%), encompassing 41.8%, 45.1%, and 40.2% of all ASVs in SW, 100mW, and BW (Supplementary Fig. S3, respectively), followed by Stramenopiles (ranging from 17.1% to 13.57%) and Rhizaria (ranging from 10.4% to 9.6%). In terms of read abundance, the water samples were dominated by dinophytes and crustaceans (Supplementary Fig. S4). Both groups had a comparable relative abundance, with a higher crustacean abundance in SW (39.4% vs. 25.3%), and a higher dinophytes abundance in 100mW (18.3% vs. 34.8%) and BW (19.9% vs. 33.2%).

By comparison, the sediment samples showed clearly distinct patterns of taxonomic composition from those observed in the water column. Bacillariophyta (diatoms) and metazoans (dominated by annelids, crustaceans, bivalves, and nematodes) were the most abundant groups of eukaryotes retrieved from sediment (33.5% and 30.9%, Supplementary Fig. S4, respectively). Noticeably, the Bacillariophyta was found to dominate the Svalbard sediment samples collected in July and August (excluding KH01, Fig. 2C, and Supplementary Fig. S5), ranging from 17% to 81.7% of read abundance. A considerable amount of sediment-derived eukaryotic ASVs (41.4%, representing 12.4% of the reads) did not match with any reference sequences using the PR2 databases at a similarity cut-off of 85% (Supplementary Fig. S3). The proportion of unassigned ASVs in the sediment reached almost 50% in both richness and abundance at the deep stations of Greenland Ridge (Fig. 2C). In contrast, the proportion of unassigned ASVs in the water samples ranged from 9% to 14% (less than 3.3% of the reads).

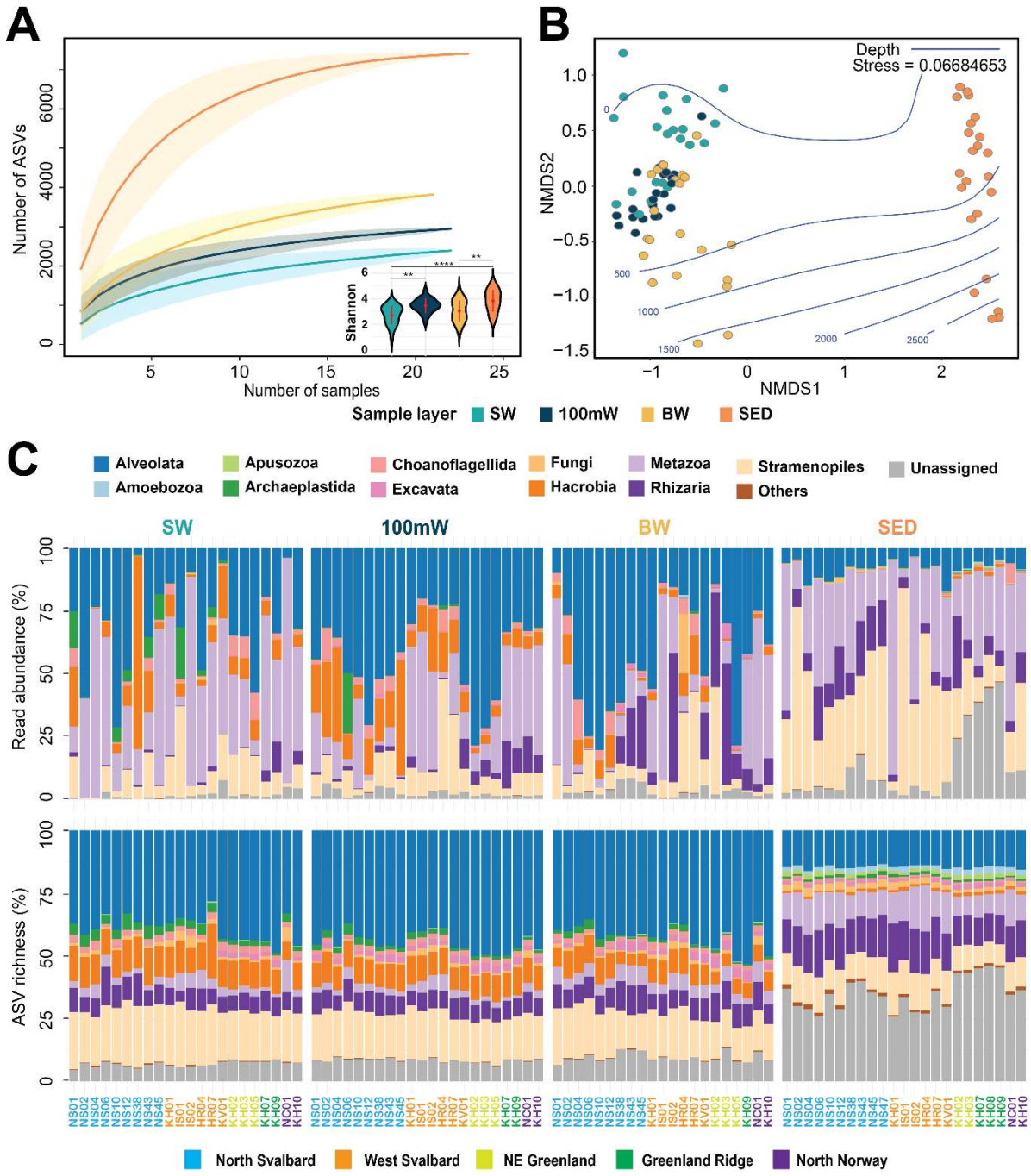


Figure 2. (A) ASV accumulation curves as a function of sampling effort. The inset shows the distribution of Shannon diversity for plankton and benthic communities. The red dots and bars within the violin plots represent means and standard deviation, and horizontal bars indicate significant differences (Wilcoxon tests, **p < 0.01 and ****p < 0.0001). (B) Nonmetric multidimensional scaling (NMDS) analysis of the Bray-Curtis dissimilarity matrix computed from whole eukaryotic datasets. The blue lines on the ordination represent the water depth as fitted smooth surfaces to the ordination. (C) Taxonomic composition of eukaryotes in terms of abundance and richness in the three water layers and sediment samples.

Patterns of eukaryotic plankton diversity in water and sediment *e*DNA

The eukaryotic plankton diversity showed substantial differences between the water column and sediments. As illustrated by the Venn diagram (Fig. 3A, and in more detail in Supplementary Table S4), only 1775 ASVs were detected in both water and sediment *e*DNA samples (representing 81.7% of the reads), while 2568 ASVs (representing 11.4% of the reads) were exclusively detected in the water and 5644 ASVs (representing 6.9% of the reads) exclusively in the sediment. Altogether, 4343 ASVs detected in water samples (43.5% of the total ASVs) were considered as plankton DNA. Some ASVs assigned to typical benthic metazoan groups (e.g., Annelida, Nemertea, Sagenista, and Urochordata) were detected in water samples, but their number is relatively low (less than 0.3%). The ASVs found exclusively in the water samples were mainly assigned to alveolates (Supplementary Figs. S6A and S6B). In contrast, the ASVs found exclusively in sediments were considered to represent benthic DNA, even if 48.2% of them remained unassigned (Fig. 3B). The 1775 ASVs shared between water and sediment samples included unassigned ASVs (19.4%) and generally belonged to typical plankton eukaryotic taxa (e.g., Syndiniales ~ 8.6%, Dinophyceae ~ 8.4%, Filosa-Thecofilosea ~ 5.4%, Supplementary Figs. S6C and S6D). We assumed that they corresponded to sinking plankton organisms. The 638 plankton ASVs were detected throughout the entire water column and sediment and generally represented the most abundant eukaryotes in the water column (Fig. 3). The taxonomic richness of the shared ASVs was roughly similar between SW and 100mW as well as BW and SED (Supplementary Fig. S6D).

Analysis of plankton ASVs from water and sediment samples revealed diversity patterns across environmental factors and geography (Fig. S7). The overall alpha diversity (normalized richness and Shannon diversity) of plankton ASVs tends to increase with increasing salinity and temperature but not latitude. In the case of water depth, the Shannon index increased, while richness showed a slight decline in deeper stations. However, this pattern is not consistent across all plankton groups, with some of them decreasing in diversity with increasing salinity (Mamiellophyceae, Prymnesionphyceae), temperature (Spirotrichea, Radiolaria), or water depth (Bacillariophyta, Prymnesiophyceae, MAST, Choanoflagellida).

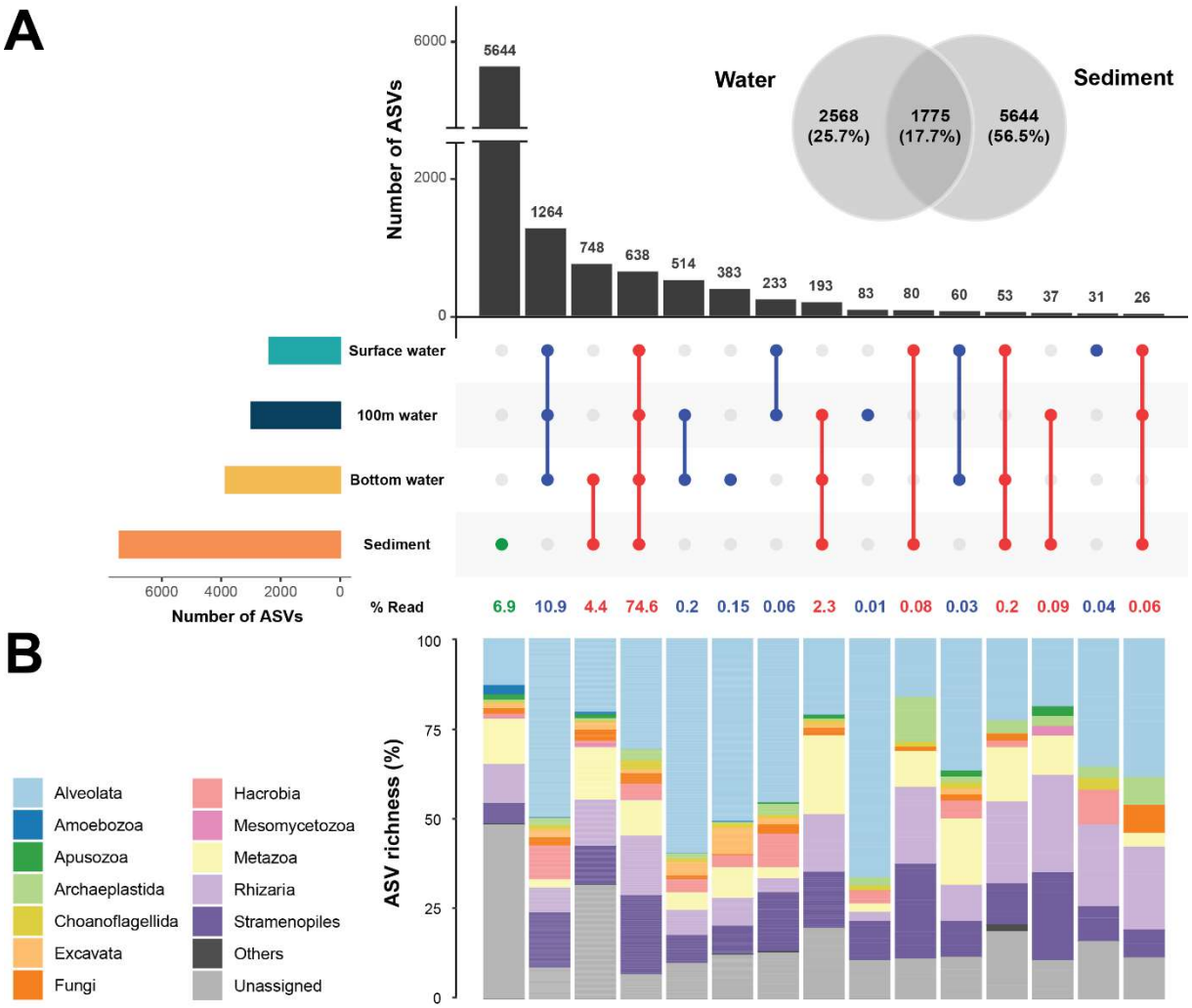


Figure 3. (A) Venn diagram representing the distribution of ASV richness and its proportions across water and sediment samples. The UpSet plot depicts intersections among water and sediment samples. The vertical bars indicated the size of the intersections (number of ASVs) on each set, sorted by size. ASVs observed exclusively in water are indicated with blue dots, exclusively in sediment with green dots, or shared between water and sediment with red dots. The read proportions of the ASVs at each intersection are indicated below the plot. (B) Bar graph indicating the taxonomic composition of ASVs of each intersection.

Taxonomic biases in sediment record of plankton taxa

To gain insight into the potential diversity loss during the sinking process, we analyzed the normalized richness and abundance of different plankton groups in the three water layers and sediments. In general, the plankton groups exhibited similar normalized richness patterns in the water column and in the sediment (Fig. 4A). Most groups exhibited a consistently high number

of ASVs in the water, with their peak richness occurring primarily in bottom waters, ranging from 0.97 to 7.3 times higher compared to sediment. The groups with the highest ASV richness in the water column, such as Dinophyceae, MAST, Bacillariophyta, and Spirotrichea, were also found in significant numbers of ASVs in the sediment (Fig. 4A). However, this pattern does not hold for other groups such as Prymnesiophyceae and Picozoa, which are diverse and abundant in the water layers but were rarely detected in our sediment samples. These two groups are represented in the water layers by 73 ASVs and 24 ASVs, respectively, while only 10 and 6 ASVs assigned to these taxa were found in the sediment (Supplementary Table S5). Regarding the relative abundance profiles, we observed a decrease in the number of reads from water to sediment samples in most plankton taxa (Maxillopoda, Dinophyceae, MAST, and Spirotrichea, as in Fig. 4B). It is noted that Prymnesiophyceae and Picozoa exhibited significantly low read abundance in sediment. In contrast, Bacillariophyta was found to dominate the sediment dataset rather than the water. We also observe an increase in both richness and abundance of Chrysophyceae, Polycystinea, Acantharea, and other radiolarians in the bottom water.

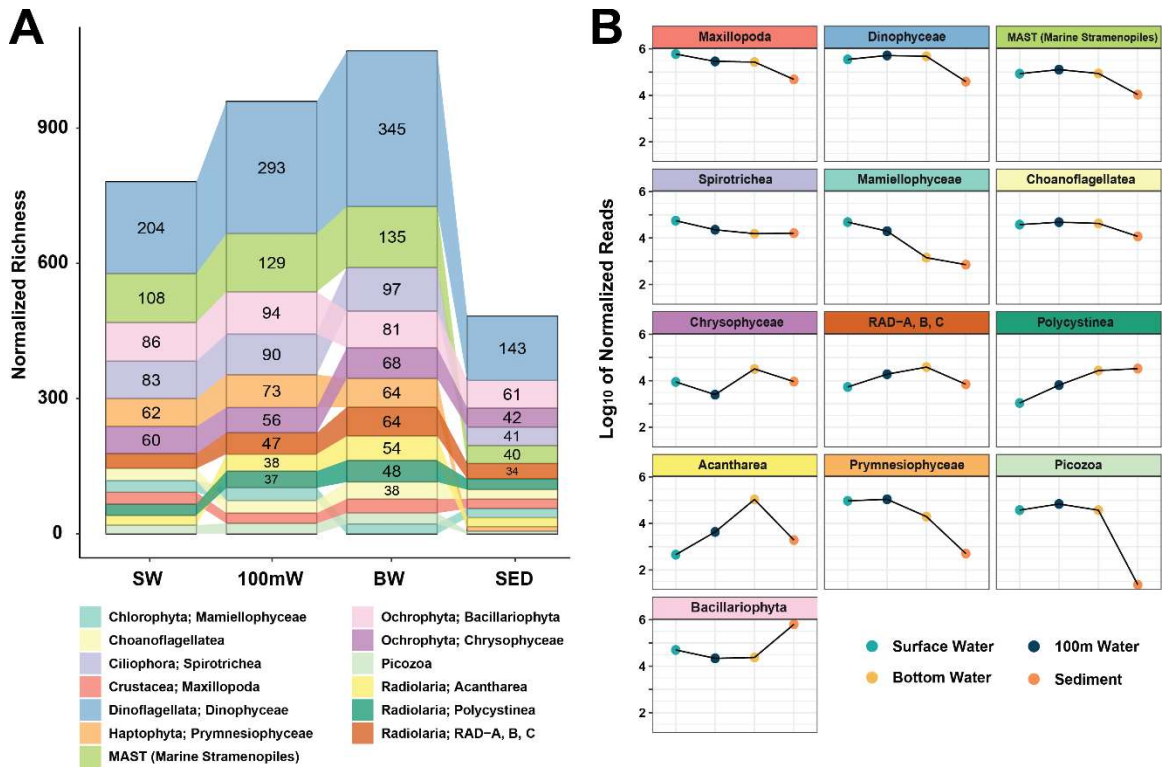


Figure 4. Vertical distribution of the plankton groups. (A) Normalized richness was obtained by rarefying the water-only ASVs matrix and (B) log10 of normalized abundance detected in three water layers and sediment.

DISCUSSION

Water and surface sediment *e*DNAs show different patterns of eukaryotic diversity

Our study confirms discrepancies in patterns of abundance and richness of eukaryotic diversity across water and surface sediment samples, in agreement with previous studies of high Arctic bacterial communities (Balmonte et al., 2018), as well as global open ocean prokaryotic and eukaryotic diversity surveys (Hoshino et al., 2020; Cordier et al., 2022). The high diversity and abundance of dinoflagellates in water samples were consistent with the observations reported throughout the year from the Svalbard fjord (Marquardt et al., 2016), and other Arctic regions such as the Siberian Arctic in summer (Kong et al., 2023), and the Canadian Arctic in autumn (Freyria et al., 2021). Other groups present in our data, including Maxillopoda (mainly copepods), haptophytes (mainly Prymnesiophyceae), radiolarians, Picozoa, and diatoms formed an assemblage typical of Arctic plankton communities (Comeau et al., 2011; Marquardt et al., 2016; Cerfonteyn et al., 2023).

Regarding sediment data, the most striking finding was the dominance of diatom DNA in Svalbard samples, possibly related to seasonal productivity (see below). In addition to diatoms, the Svalbard sediment datasets also comprised many reads assigned to benthic metazoans, as previously observed in other metabarcoding surveys of this area (van den Heuvel-Greve et al., 2021). In contrast, the sediment samples from Greenland were dominated by unassigned reads and other groups of eukaryotes (mainly Labyrinthulomycetes and MAST-12). This difference could be explained by the lower bathyal depths at which Greenland samples were collected (1900-3000 m), in agreement with a high level of unknown eukaryotic diversity generally observed on the deep-sea floor (Scheckenbach et al., 2010; Pawlowski et al., 2011; Cordier et al., 2022). According to our study, eukaryotic diversity increases from surface waters to sea bottom at a comparable sequencing effort (Fig. 2A). This is in agreement with a global-scale analysis of marine ecosystems showing that benthic communities exhibited considerably higher phylogenetic diversity than plankton or particle-associated communities, both in the case of bacteria (Zinger et al., 2011) and microbial eukaryotes (Cordier et al., 2022).

In the water column, the bottom water was found to be taxonomically richer than the overlying layers. This could be explained by the vertical accumulation of genetic material, but possibly also by the presence of many ASVs of benthic origins. Some benthic taxa are known to possess meroplankton larvae (Ershova et al., 2019), but most of these ASVs may originate from extra-organismal DNA or extracellular DNA (Barnes and Turner, 2016; Antich et al.,

2021). A previous *e*DNA metabarcoding study of metazoan and protist diversity in the deep ocean showed that meso- and bathy-plankton species abounded in water samples collected at the benthic boundary layer (BBL), but relatively few of them were found in the underlying sediments (Laroche et al., 2020). In contrast, our study showed that bottom water samples comprised the highest number of ASVs shared with the sediment. This number varied from 1.4% and 5.7% in the west and north of Svalbard, respectively, to 10% in the northeast of Greenland, most likely due to regional differences in current activities, which may enhance the exchange of eDNA between the near-bottom layers. The similarity of eukaryotic communities in sediments and BBL was already observed in different regions (Dauvin and Vallet, 2006; Parry et al., 2020; Tagliabue et al., 2023).

Phytoplankton productivity impacts the pool of sedimentary DNA

Exploring the vertical patterns of plankton *e*DNA, we found that plankton communities are influenced to varying degrees by sampling times and environmental parameters (Figs. S1 and S7). This variation corresponded to the seasonal and spatial environmental changes that are particularly high in the Arctic Ocean and its marginal seas. Compared to the seafloor, the water column is directly affected by pronounced (multi-) seasonal changes in sea ice cover (Stroeve and Notz, 2018), freshwater input (Brown et al., 2020), light availability (Berge et al., 2015), and the increased influence of warm and saline Atlantic water (Polyakov et al., 2017). In line with this, dramatic shifts in the composition of plankton communities have been demonstrated in response to seasonality and changes in environmental conditions in the Arctic Ocean (Wietz et al., 2021).

Although we could not investigate seasonal changes in our study, as the sampling of each location was done at different times of the year, the impact of seasonality was well illustrated in our data by the overabundance of diatoms (up to 81.7 % reads) in the Svalbard sediment samples. This exceptional abundance could be explained by the well-documented spring bloom of diatoms in the Arctic (Degerlund and Eilertsen, 2010; Hodal et al., 2012; Wietz et al., 2021; Shiozaki et al., 2022) and their fast sinking-mechanisms toward the bottom (Sakshaug, 2004; Agusti et al., 2015). The sinking of diatom cells may be accelerated by easily forming fast-sinking aggregates (Schewe and Soltwedel, 2003), low zooplankton grazing in the spring (Norrbin et al., 2009), and the ability of diatoms to remain intact at the sea bottom for a considerable period of time. This is supported by the resistance of some diatom species to prolonged darkness (van de Poll et al., 2020), and the formation of the resting stage (Druzhkova et al., 2018; Hegseth et al., 2019). On the other hand, the dominance of diatoms on the seafloor

could be in the form of fast-sinking fecal pellets (Van der Wal et al., 1995; Leah and Hans, 1998), which may be enhanced by the selective feeding of zooplankton with diatoms more readily grazed by copepods than, e.g., haptophytes (Wassmann, 1994; Verity, 2000).

Interestingly, diatoms were much less abundant in sediment samples from Greenland. This could be explained by the much lower productivity in Greenland waters or by the bottom depth of the Greenland stations, where the samples have been collected. Indeed, the Svalbard stations were located mainly at water depths less than 400 m, while the sediment samples were collected at water depths ranging from 400 to 1900 m for NE Greenland and from 2900 to 3000 m for the Greenland Ridge. The lower amount of diatom and other phytoplankton DNA in these samples could be due to increased DNA degradation with depth. Alternatively, it could be the effect of a large number of unassigned reads in sediment samples (more than 50% in Greenland Ridge) that significantly reduced the proportion of other reads.

It is important to note that our datasets are not time series and represent a one-time point at each location. Therefore, the abundance patterns of the plankton community in the studied water column represented a snapshot in time, while the plankton DNA accumulated in the sediment represents up to several years. To monitor the impact of seasonal phytoplankton productivity on the seafloor, time series studies of sediment *e*DNA would be needed. However, to our knowledge, only a few such studies have been conducted and all relate to coastal shallow water areas (Parry et al., 2020; Tagliabue et al., 2023). It might be difficult and certainly far more expensive to obtain a time series of sediment samples in the polar regions, where access throughout the year is limited. This issue has been solved for the survey of the plankton community by using an automated water *e*DNA sampler (Wietz et al., 2021). We expect that further development of such devices for sediment samples will be possible in the near future.

Not all plankton taxa are equally archived on the sea floor

It is well-known that the plankton DNA degrades during the sinking to the bottom due to various abiotic and biotic factors (Pedersen et al., 2015; Torti et al., 2015; Nguyen et al., 2023). According to a study of *e*DNA persistence in marine systems, its half-time in offshore waters was estimated to be about 40 hours (Collins et al., 2018). Assuming that particle-sinking velocities range from 10 to 150 m per day in polar regions (McDonnell and Buesseler, 2010), most of the plankton DNA should be degraded before reaching the sea floor at all stations that are deeper than 300 m. Our study suggests that *e*DNA may be more persistent, given the abundance of some epi-plankton taxa in sedimentary *e*DNA samples (more than 50% of reads

in some samples). This implies that *e*DNA degradation may go slower than estimated, or that *e*DNA is transported more rapidly to the seafloor (i.e. through fast-sinking aggregates or fecal pellets).

One could argue that the most abundant plankton taxa are more prone to leave their signature in the sediment. However, as shown by our data, the richness and abundance profiles differ between plankton groups (Figs. 4 and S7). For example, the richness and abundance profiles of copepods and dinoflagellates in water samples are similar to their richness and abundance profiles in the sediment, suggesting that their DNA degrades uniformly through the water column, and their community structure is preserved after arrival at the seafloor. This was not the case for radiolarians and diatoms. Both the richness and abundance patterns of radiolarians increased with increasing water depth, and the different taxa representing this group were well represented in the sediments. On the other hand, diatoms dominated metabarcoding data in Svalbard sediment samples in terms of abundance but not richness, and their presence in water *e*DNA samples is relatively limited.

Our study reinforces the idea that different taxa have different capacities to be preserved in sedimentary DNA archives. As shown in Fig. 4 and Supplementary Table S5, the highest DNA preservation is shown by diatoms, choanoflagellates, and chrysophyceae with over 60% of ASVs recorded in sediments. The lowest preservation is observed in Picozoa and Prymnesiophyceae with 21% and 13% of ASVs recorded. The good preservation of the diversity of diatoms could be easily explained by the presence of frustules that protect the cellular content, including the nucleus and DNA (Yang et al., 2023). However, in the case of chrysophyceae and choanoflagellates, protective loricas are rarely observed in plankton species. Cell size could be another factor that influences DNA preservation in sinking plankton organisms. Indeed, both Picozoa and Phaeocystis (which dominates Prymnesiophyceae) are characterized by tiny cells of less than 5 µm. Such tiny cells may settle less successfully on the seafloor or benthic predators more easily prey upon them. The other biological factors, such as the reduction in genome size in pico- and nanoeukaryotes (Derilus et al., 2020) could result in their lower abundance in sediment datasets. However, the fact that both groups are well represented in the water metabarcoding data, even the bottom water, suggests that their lower detection in the sedimentary *e*DNA data is somehow related to the processes occurring at the sediment surface.

Finally, we cannot exclude that the dominance of some groups (e.g. diatoms) in surface sediments covers the signal of other less abundant taxa. This effect of accumulation at the seafloor surface of DNA from highly productive plankton taxa might be mitigated in the

subsurface layers. Some metabarcoding studies show the difference in the taxonomic composition of plankton DNA preserved at the surface and in the subsurface layers of sediments (Morard et al., 2017; Barrenechea Angeles et al., 2020). This difference is explained by the rapid degradation of DNA during the burial process in the sediments. It would be interesting to test its effect on different planktonic taxa. Perhaps the pico- and nanoplankton groups rare in our surface sediment data will appear in the subsurface samples.

CONCLUSIONS

Our study offers the first detailed analysis of the sedimentary DNA records of plankton diversity in the Nordic Seas. We show that the taxonomic composition and structure of eukaryotic plankton communities in water and sediment DNA samples are different. Plankton DNA dominates in marine surface sediments in terms of abundance, but not diversity. Most of the DNA present on the sea floor comes from phytoplankton (diatom) blooms. Other groups of plankton eukaryotes are represented in the sediments, but their diversity is significantly lower, especially for some pico- and nanoplankton. A more extensive sampling design, encompassing additional survey time points and subsurface sampling, would be necessary to corroborate our results. However, given the number of samples analyzed here and the wide geographic area of the Nordic Seas covered by our sampling, we expect that the observed differences between water and sediment DNA are representative of this region. The potential incompleteness of sedimentary DNA archives on plankton life might need to be taken into account by DNA-based analyses of past marine ecosystems.

REFERENCE

Agusti, S., González-Gordillo, J.I., Vaqué, D., et al. (2015). Ubiquitous healthy diatoms in the deep sea confirm deep carbon injection by the biological pump. Nat. Commun. 6(1), 7608. doi: 10.1038/ncomms8608.

Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W., et al. (2009). A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. PLoS One 4(7), e6372. doi: 10.1371/journal.pone.0006372.

Antich, A., Palacín, C., Cebrian, E., et al. (2021). Marine biomonitoring with eDNA: Can metabarcoding of water samples cut it as a tool for surveying benthic communities? Mol. Ecol. 30(13), 3175-3188. doi: 10.1111/mec.15641.

Armbrecht, L., Focardi, A., Lawler, K.-A., et al. (2023). From the surface ocean to the seafloor: linking modern and paleo-genetics at the Sabrina Coast, East Antarctica (IN2017_V01). J. Geophys. Res. Biogeosci. 128(4), e2022JG007252. doi: 10.1029/2022JG007252.

Armbrecht, L., Weber, M.E., Raymo, M.E., et al. (2022). Ancient marine sediment DNA reveals diatom transition in Antarctica. Nat. Commun. 13(1), 5787. doi: 10.1038/s41467-022-33494-4.

Armbrecht, L.H., Coolen, M.J.L., Lejzerowicz, F., et al. (2019). Ancient DNA from marine sediments: Precautions and considerations for seafloor coring, sample handling and data generation. Earth Sci. Rev. 196, 102887. doi: 10.1016/j.earscirev.2019.102887.

Balmonte, J.P., Teske, A., and Arnosti, C. (2018). Structure and function of high Arctic pelagic, particle-associated and benthic bacterial communities. Environ. Microbiol. 20(8), 2941-2954. doi: 10.1111/1462-2920.14304.

Barnes, M.A., and Turner, C.R. (2016). The ecology of environmental DNA and implications for conservation genetics. Conserv. Genet. 17(1), 1-17. doi: 10.1007/s10592-015-0775-4.

Barrenechea Angeles, I., Lejzerowicz, F., Cordier, T., et al. (2020). Planktonic foraminifera eDNA signature deposited on the seafloor remains preserved after burial in marine sediments. Sci. Rep. 10(1), 20351. doi: 10.1038/s41598-020-77179-8.

Barrenechea Angeles, I., Romero-Martínez, M.L., Cavaliere, M., et al. (2023). Encapsulated in sediments: eDNA deciphers the ecosystem history of one of the most polluted European marine sites. Environ. Int. 172, 107738. doi: 10.1016/j.envint.2023.107738.

Berge, J., Renaud, P.E., Darnis, G., et al. (2015). In the dark: A review of ecosystem processes during the Arctic polar night. Prog. Oceanogr. 139, 258-271. doi: 10.1016/j.pocean.2015.08.005.

Brandt, M.I., Trouche, B., Quintric, L., et al. (2021). Bioinformatic pipelines combining denoising and clustering tools allow for more comprehensive prokaryotic and eukaryotic metabarcoding. Mol. Ecol. Resour. 21(6), 1904-1921. doi: 10.1111/1755-0998.13398.

Brown, K.A., Holding, J.M., and Carmack, E.C. (2020). Understanding Regional and Seasonal Variability Is Key to Gaining a Pan-Arctic Perspective on Arctic Ocean Freshening. Front. Mar. Sci. 7. doi: 10.3389/fmars.2020.00606.

Callahan, B.J., Wong, J., Heiner, C., et al. (2019). High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. Nucleic Acids Res. 47(18), e103-e103. doi: 10.1093/nar/gkz569.

Capo, E., Monchamp, M.-E., Coolen, M.J.L., et al. (2022). Environmental paleomicrobiology: using DNA preserved in aquatic sediments to its full potential. Environ. Microbiol. 24(5), 2201-2209. doi: 10.1111/1462-2920.15913.

Cerfonteyn, M., Groben, R., Vaulot, D., et al. (2023). The distribution and diversity of eukaryotic phytoplankton in the Icelandic marine environment. Sci. Rep. 13(1), 8519. doi: 10.1038/s41598-023-35537-2.

Collins, R.A., Wangensteen, O.S., O'Gorman, E.J., et al. (2018). Persistence of environmental DNA in marine systems. Commun. Biol. 1(1), 185. doi: 10.1038/s42003-018-0192-6.

Comeau, A.M., Li, W.K.W., Tremblay, J.-É., et al. (2011). Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. PloS One 6(11), e27492. doi: 10.1371/journal.pone.0027492.

Cordier, T., Angeles, I.B., Henry, N., et al. (2022). Patterns of eukaryotic diversity from the surface to the deep-ocean sediment. Sci. Adv. 8(5), eabj9309. doi: 10.1126/sciadv.abj9309.

Dauvin, J.-C., and Vallet, C. (2006). The near-bottom layer as an ecological boundary in marine ecosystems: Diversity, taxonomic composition and community definitions. Hydrobiologia 555(1), 49-58. doi: 10.1007/s10750-005-1105-5.

De Schepper, S., Ray, J.L., Skaar, K.S., et al. (2019). The potential of sedimentary ancient DNA for reconstructing past sea ice evolution. ISME J. 13(10), 2566-2577. doi: 10.1038/s41396-019-0457-1.

Degerlund, M., and Eilertsen, H.C. (2010). Main species characteristics of phytoplankton spring blooms in NE Atlantic and Arctic Waters (68–80° N). Estuaries Coast. 33(2), 242-269. doi: 10.1007/s12237-009-9167-7.

Derilus, D., Rahman, M.Z., Pinero, F., et al. (2020). Synergism between the Black Queen effect and the proteomic constraint on genome size reduction in the photosynthetic picoeukaryotes. Sci. Rep. 10(1), 8918. doi: 10.1038/s41598-020-65476-1.

Devendra, D., Łącka, M., Szymańska, N., et al. (2023). The development of ocean currents and the response of the cryosphere on the Southwest Svalbard shelf over the Holocene. Glob. Planet. Change 228, 104213. doi: 10.1016/j.gloplacha.2023.104213.

Druzhkova, E., Oleinik, A., and Makarevich, P. (2018). Live autochthonous benthic diatoms on the lower depths of Arctic continental shelf. Preliminary results. Oceanologia 60(1), 97-100. doi: 10.1016/j.oceano.2017.07.001.

Dufresne, Y., Lejzerowicz, F., Perret-Gentil, L.A., et al. (2019). SLIM: a flexible web application for the reproducible processing of environmental DNA metabarcoding data. BMC Bioinform. 20(1), 88. doi: 10.1186/s12859-019-2663-2.

Dusa, A. (2018). Package 'venn'. version 1.9.

Ellegaard, M., Clokie, M.R.J., Czypionka, T., et al. (2020). Dead or alive: sediment DNA archives as tools for tracking aquatic evolution and adaptation. Commun. Biol. 3(1), 169. doi: 10.1038/s42003-020-0899-z.

Ershova, E.A., Descoteaux, R., Wangensteen, O.S., et al. (2019). Diversity and distribution of meroplanktonic larvae in the Pacific Arctic and connectivity with adult benthic invertebrate communities. Front. Mar. Sci. 6. doi: 10.3389/fmars.2019.00490.

Freyria, N.J., Joli, N., and Lovejoy, C. (2021). A decadal perspective on north water microbial eukaryotes as Arctic Ocean sentinels. Sci. Rep. 11(1), 8413. doi: 10.1038/s41598-021-87906-4.

Frøslev, T.G., Kjøller, R., Bruun, H.H., et al. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nat. Commun. 8(1), 1188. doi: 10.1038/s41467-017-01312-x.

Gehlenborg, N. (2019). UpSetR: a more scalable alternative to Venn and Euler diagrams for visualizing intersecting sets. . R package version 1.4. 0.

Guillou, L., Bachar, D., Audic, S., et al. (2013). The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Res. 41(D1), D597-D604. doi: 10.1093/nar/gks1160.

Hegseth, E.N., Assmy, P., Wiktor, J.M., et al. (2019). "Phytoplankton Seasonal Dynamics in Kongsfjorden, Svalbard and the Adjacent Shelf," in The Ecosystem of Kongsfjorden, Svalbard, eds. H. Hop & C. Wiencke. (Cham: Springer International Publishing), 173-227.

Hodal, H., Falk-Petersen, S., Hop, H., et al. (2012). Spring bloom dynamics in Kongsfjorden, Svalbard: nutrients, phytoplankton, protozoans and primary production. Polar Biol. 35(2), 191-203. doi: 10.1007/s00300-011-1053-7.

Hoshino, T., Doi, H., Uramoto, G.-I., et al. (2020). Global diversity of microbial communities in marine sediment. Proc. Natl. Acad. Sci. U. S. A. 117(44), 27587-27597. doi: 10.1073/pnas.1919139117.

Kassambara, A. (2020). Package 'ggpubr'. R package version 0.1 6(0).

Kong, H., Yang, E.-J., Jiao, N., et al. (2023). RNA outperforms DNA-based metabarcoding in assessing the diversity and response of microeukaryotes to environmental variables in

the Arctic Ocean. Sci. Total Environ. 876, 162608. doi: 10.1016/j.scitotenv.2023.162608.
Laroche, O., Kersten, O., Smith, C.R., et al. (2020). From sea surface to seafloor: A benthic allochthonous eDNA survey for the abyssal ocean. Front. Mar. Sci. 7. doi: 10.3389/fmars.2020.00682.
Leah, R.F., and Hans, G.D. (1998). Effects of diet on dimensions, density and sinking rates of fecal pellets of the copepod Acartia tonsa. Mar. Ecol. Prog. Ser. 175, 87-96. doi: 10.3354/meps175087.
Lecroq, B., Lejzerowicz, F., Bachar, D., et al. (2011). Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. Proc. Natl. Acad. Sci. U. S. A. 108(32), 13177-13182. doi: 10.1073/pnas.1018426108.
Marino, M., Rodrigues, T., Quivelli, O., et al. (2022). Paleoproductivity proxies and alkenone precursors in the Western Mediterranean during the Early-Middle Pleistocene transition. Palaeogeogr. Palaeoclimatol. Palaeoecol. 601, 111104. doi: 10.1016/j.palaeo.2022.111104.
Marquardt, M., Vader, A., Stübner, E.I., et al. (2016). Strong seasonality of marine microbial eukaryotes in a high-Arctic fjord (Isfjorden, in West Spitsbergen, Norway). Appl. Environ. Microbiol. 82(6), 1868-1880. doi: 10.1128/AEM.03208-15.
McDonnell, A.M.P., and Buesseler, K.O. (2010). Variability in the average sinking velocity of marine particles. Limnol. Oceanogr. 55(5), 2085-2096. doi: 10.4319/lo.2010.55.5.2085.
McKnight, D.T., Huerlimann, R., Bower, D.S., et al. (2019). microDecon: A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies. Environ. DNA. 1(1), 14-25. doi: 10.1002/edn3.11.
Morard, R., Lejzerowicz, F., Darling, K.F., et al. (2017). Planktonic foraminifera-derived environmental DNA extracted from abyssal sediments preserves patterns of plankton macroecology. Biogeosciences 14(11), 2741-2754. doi: 10.5194/bg-14-2741-2017.
Nguyen, N.-L., Devendra, D., Szymańska, N., et al. (2023). Sedimentary ancient DNA: a new paleogenomic tool for reconstructing the history of marine ecosystems. Front. Mar. Sci. 10, 1075. doi: 10.3389/fmars.2023.1185435.
Norrbin, F., Eilertsen, H.C., and Degerlund, M. (2009). Vertical distribution of primary producers and zooplankton grazers during different phases of the Arctic spring bloom. Deep Sea Res. Part II Top. Stud. Oceanogr. 56(21), 1945-1958. doi: 10.1016/j.dsr2.2008.11.006.
Oksanen, J., Blanchet, F.G., Friendly, M., et al. (2019). Package 'vegan'. Community ecology package, version 2 2(9).
Pados-Dibattista, T., Pearce, C., Detlef, H., et al. (2022). Holocene palaeoceanography of the Northeast Greenland shelf. Clim. Past 18(1), 103-127. doi: 10.5194/cp-18-103-2022.
Parry, H.E., Atkinson, A., Somerfield, P.J., et al. (2020). A metabarcoding comparison of taxonomic richness and composition between the water column and the benthic boundary layer. ICES J. Mar. Sci. 78(9), 3333-3341. doi: 10.1093/icesjms/fsaa228.
Paulson, J.N., Stine, O.C., Bravo, H.C., et al. (2013). Differential abundance analysis for microbial marker-gene surveys. Nat. Methods 10(12), 1200-1202. doi: 10.1038/nmeth.2658.
Pawłowska, J., Wollenburg, J.E., Zajączkowski, M., et al. (2020). Planktonic foraminifera genomic variations reflect paleoceanographic changes in the Arctic: evidence from sedimentary ancient DNA. Sci. Rep. 10(1), 15102. doi: 10.1038/s41598-020-72146-9.
Pawlowski, J., Bruce, K., Panksep, K., et al. (2022). Environmental DNA metabarcoding for benthic monitoring: A review of sediment sampling and DNA extraction methods. Sci. Total Environ. 818, 151783. doi: 10.1016/j.scitotenv.2021.151783.

Pawlowski, J., Christen, R., Lecroq, B., et al. (2011). Eukaryotic richness in the Abyss: Insights from pyrotag sequencing. PLoS One 6(4), e18169. doi: 10.1371/journal.pone.0018169.
Pedersen, M.W., Overballe-Petersen, S., Ermini, L., et al. (2015). Ancient and modern environmental DNA. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 370(1660), 20130383. doi: 10.1098/rstb.2013.0383.
Polyakov, I.V., Pnyushkov, A.V., Alkire, M.B., et al. (2017). Greater role for Atlantic inflows on sea-ice loss in the Eurasian Basin of the Arctic Ocean. Science 356(6335), 285-291. doi: 10.1126/science.aai8204.
R Core Team, R. (2013). R: A language and environment for statistical computing.
Sakshaug, E. (2004). "Primary and Secondary Production in the Arctic Seas," in The Organic Carbon Cycle in the Arctic Ocean, eds. R. Stein & R.W. MacDonald. (Berlin, Heidelberg: Springer Berlin Heidelberg), 57-81.
Scheckenbach, F., Hausmann, K., Wylezich, C., et al. (2010). Large-scale patterns in biodiversity of microbial eukaryotes from the abyssal sea floor. Proc. Natl. Acad. Sci. U. S. A. 107(1), 115-120. doi: 10.1073/pnas.0908816106.
Schewe, I., and Soltwedel, T. (2003). Benthic response to ice-edge-induced particle flux in the Arctic Ocean. Polar Biol. 26(9), 610-620. doi: 10.1007/s00300-003-0526-8.
Shiozaki, T., Fujiwara, A., Sugie, K., et al. (2022). Bottom-associated phytoplankton bloom and its expansion in the Arctic Ocean. Glob. Chang. Biol. 28(24), 7286-7295. doi: 10.1111/gcb.16421.
Siano, R., Lassudrie, M., Cuzin, P., et al. (2021). Sediment archives reveal irreversible shifts in plankton communities after World War II and agricultural pollution. Curr. Biol. 31(12), 2682-2689.e2687. doi: 10.1016/j.cub.2021.03.079.
Sogawa, S., Nakamura, Y., Nagai, S., et al. (2022). DNA metabarcoding reveals vertical variation and hidden diversity of Alveolata and Rhizaria communities in the western North Pacific. Deep Sea Res. Part I Oceanogr. Res. Pap. 183, 103765. doi: 10.1016/j.dsr.2022.103765.
Stroeve, J., and Notz, D. (2018). Changing state of Arctic sea ice across all seasons. Environ. Res. Lett. 13(10), 103001. doi: 10.1088/1748-9326/aade56.
Tagliabue, A., Matterson, K.O., Ponti, M., et al. (2023). Sediment and bottom water eDNA metabarcoding to support coastal management. Ocean Coast. Manag. 244, 106785. doi: 10.1016/j.ocecoaman.2023.106785.
Torti, A., Lever, M.A., and Jørgensen, B.B. (2015). Origin, dynamics, and implications of extracellular DNA pools in marine sediments. Mar. Genomics 24, 185-196. doi: 10.1016/j.margen.2015.08.007.
van de Poll, W.H., Abdullah, E., Visser, R.J.W., et al. (2020). Taxon-specific dark survival of diatoms and flagellates affects Arctic phytoplankton composition during the polar night and early spring. Limnol. Oceanogr. 65(5), 903-914. doi: 10.1002/lno.11355.
van den Heuvel-Greve, M.J., van den Brink, A.M., Glorius, S.T., et al. (2021). Early detection of marine non-indigenous species on Svalbard by DNA metabarcoding of sediment. Polar Biol. 44(4), 653-665. doi: 10.1007/s00300-021-02822-7.
Van der Wal, P., Kempers, R.S., and Veldhuis, M.J.W. (1995). Production and downward flux of organic matter and calcite in a North Sea bloom of the coccolithophore Emiliania huxleyi. Mar. Ecol. Prog. Ser. 126, 247-265. doi: 10.3354/meps126247.
Verity, P.G. (2000). Grazing experiments and model simulations of the role of zooplankton in Phaeocystis food webs. J. Sea Res. 43(3), 317-343. doi: 10.1016/S1385-1101(00)00025-3.
Wassmann, P. (1994). Significance of sedimentation for the termination of Phaeocystis blooms. J. Mar. Syst. 5(1), 81-100. doi: 10.1016/0924-7963(94)90018-3.
Wickham, H. (2011). ggplot2. Wiley Interdiscip. Rev. Comput. Stat. 3(2), 180-185.

Wietz, M., Bienhold, C., Metfies, K., et al. (2021). The polar night shift: seasonal dynamics and drivers of Arctic Ocean microbiomes revealed by autonomous sampling. ISME Commun. 1(1), 76. doi: 10.1038/s43705-021-00074-4.

Wood, S.N. (2001). mgcv: GAMs and generalized ridge regression for R. R news 1(2), 20-25.

Yang, C., Feng, C., Li, Y., et al. (2023). Morphological and physicochemical characteristics, biological functions, and biomedical applications of diatom frustule. Algal Res. 72, 103104. doi: 10.1016/j.algal.2023.103104.

Zimmermann, H.H., Stoof-Leichsenring, K.R., Dinkel, V., et al. (2023). Marine ecosystem shifts with deglacial sea-ice loss inferred from ancient DNA shotgun sequencing. Nat. Commun. 14(1), 1650. doi: 10.1038/s41467-023-36845-x.

Zimmermann, H.H., Stoof-Leichsenring, K.R., Kruse, S., et al. (2020). Changes in the composition of marine and sea-ice diatoms derived from sedimentary ancient DNA of the eastern Fram Strait over the past 30 000 years. Ocean Sci. 16(5), 1017-1032. doi: 10.5194/os-16-1017-2020.

Zinger, L., Amaral-Zettler, L.A., Fuhrman, J.A., et al. (2011). Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. PLoS One 6(9), e24570. doi: 10.1371/journal.pone.0024570.

MOLECULAR ECOLOGY
RESOURCES

Taxonomic and abundance biases affect the record of marine eukaryotic plankton communities in sediment DNA archives

Journal:	*Molecular Ecology Resources*
Manuscript ID	MER-24-0062
Manuscript Type:	Resource Article
Date Submitted by the Author:	13-Feb-2024
Complete List of Authors:	Nguyen, Ngoc-Loi; Institute of Oceanology Polish Academy of Sciences, Pawlowska, Joanna; Institute of Oceanology Polish Academy of Sciences Zajaczkowski, Marek; Institute of Oceanology Polish Academy of Sciences Weiner, Agnes Katharina Maria; NORCE Norwegian Research Centre AS Cordier, Tristan; Nowegian Research Center Grant, Danielle; NORCE Norwegian Research Centre AS Research Area Climate, Bjerknes Centre for Climate Research De Schepper, Stijn; NORCE Norwegian Research Centre AS Pawlowski, Jan; Institute of Oceanology Polish Academy of Sciences
Keywords:	Environmental DNA, Marine biodiversity, Eukaryotes, metabarcoding, taphonomy

2. Research Article II

Metabarcoding reveals high diversity of benthic foraminifera linked to water masses circulation at coastal Svalbard. Ngoc-Loi Nguyen, Joanna Pawłowska, Inès Barrenechea Angeles, Marek Zajaczkowski and Jan Pawłowski. Published in *Geobiology*, 21(1), 133–150 (2023).

Received: 28 January 2022 | Revised: 5 July 2022 | Accepted: 19 September 2022

DOI: 10.1111/gbi.12530

ORIGINAL ARTICLE

gebiology WILEY

Metabarcoding reveals high diversity of benthic foraminifera linked to water masses circulation at coastal Svalbard

Ngoc-Loi Nguyen[1] | Joanna Pawłowska[1] | Inès Barrenechea Angeles[2,3] | Marek Zajaczkowski[1] | Jan Pawłowski[1,3]

[1]Institute of Oceanology Polish Academy of Sciences, Sopot, Poland

[2]Department of Earth Sciences, University of Geneva, Geneva, Switzerland

[3]Department of Genetics and Evolution, University of Geneva, Geneva, Switzerland

Correspondence
Ngoc-Loi Nguyen, Institute of Oceanology Polish Academy of Sciences, Powstancow Warszawy 55, Sopot 81-712, Poland.
Email: loinguyen@iopan.pl

Funding information
National Science Centre, Grant/Award Number: 2018/31/B/ST10/01616; Norwegian Financial Mechanism for 2014-2021, Grant/Award Number: 2019/34/H/ST10/00682; Swiss National Science Foundation, Grant/Award Number: 31003A_179125

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 represent a main component of meiofauna in the Arctic. Several studies analyzed 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 the Svalbard Archipelago. We obtained a total of 5,968,786 reads that represented 1384 amplicon sequence variants (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 are 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 very useful for inferring present and past environmental conditions in the Arctic.

KEYWORDS
Atlantic water, benthic foraminifera, metabarcoding, sedimentary DNA, Svalbard

1 | INTRODUCTION

The Arctic Ocean is strongly impacted by the increased influence of warm and saline Atlantic water (AW), so-called "atlantification," which causes sea ice retreat and sea surface temperature increases (Beszczynska-Möller et al., 2012; Onarheim et al., 2014; Polyakov et al., 2017), higher input of turbid melt water in summer, restricting the light availability and enhancing flocculation (Nilsen et al., 2008; Zajączkowski et al., 2010) and directly affecting the entire ecosystem of the Arctic (Csapó et al., 2021). The changing environmental conditions in this region introduces a significant impact on shaping biodiversity and the biogeography of many taxonomic groups, such as birds and mammals (Descamps et al., 2017; Vihtakari et al., 2018), fish (Fossheim et al., 2015; Frainer et al., 2017), zooplankton (Grabowski et al., 2019; Hop et al., 2019; Weydmann-Zwolicka et al., 2021), phytoplankton (Barton et al., 2016; Neukermans et al., 2018), and planktonic foraminifera (Meilland et al., 2020; Ofstad et al., 2021). Such changes in physical drivers lead to a shift in Atlantic species ranges toward the Arctic (Berge et al., 2005), an increase in productivity (Slagstad et al., 2011), and changes in the timing of spring phytoplankton bloom (Zajączkowski et al., 2010). In a marine setting, biotic interactions and physical influences (temperature and salinity) may create shifts in food webs, affecting not only planktonic but also controlling benthic community structure by vertical fluxes of mineral and organic particles or phytoplankton cells to the bottom (Kortsch et al., 2015; Zajączkowski et al., 2010). Particularly, Svalbard ecosystems are currently affected by increased heat transport from the West Spitsbergen Current (WSC) (Dai et al., 2019; Nilsen et al., 2008; Onarheim et al., 2014; Serreze & Barry, 2011), which flows northwards along western Svalbard (Figure 1).

Foraminifera is a group of protists characterized by granuloreticulopodia and belonging to the supergroup of Rhizaria (Burki et al., 2010). Benthic foraminifera are highly abundant and diverse in marine environments from coastal to deep-sea zones (Gooday & Jorissen, 2012; Murray, 2014; Schoenle et al., 2021), although freshwater forms are also known to exist (Holzmann et al., 2021). Foraminifera typically possesses an organic, agglutinated, or calcareous shell (called test), which readily enter the fossil record, where they are used as index fossils and paleoenvironmental indicators (Murray, 2006, 2014). In the modern, foraminifera are also recognized as important ecological indicators of environmental stress because they are particularly sensitive to abrupt climate change (Kawahata et al., 2019; Prazeres et al., 2017; Wittmann & Pörtner, 2013). It has been demonstrated that the abundance and diversity of benthic foraminifera were extremely variable in the eastern and western Arctic during the last interglacial and glacial climate regimes (Polyak et al., 2013; Wollenburg et al., 2007) and were directly related to changes in sea ice cover, surface productivity, sedimentation, and post-depositional processes in the Arctic (Backman et al., 2009; Hald & Korsun, 1997; Polyak et al., 2013; Sabbatini et al., 2007).

The traditional approach to analyses foraminiferal diversity consists in sorting and morphological identification of hard-shelled species belonging to either the class Tubothalamea or Globothalamea, which are commonly larger than 100 μm in size

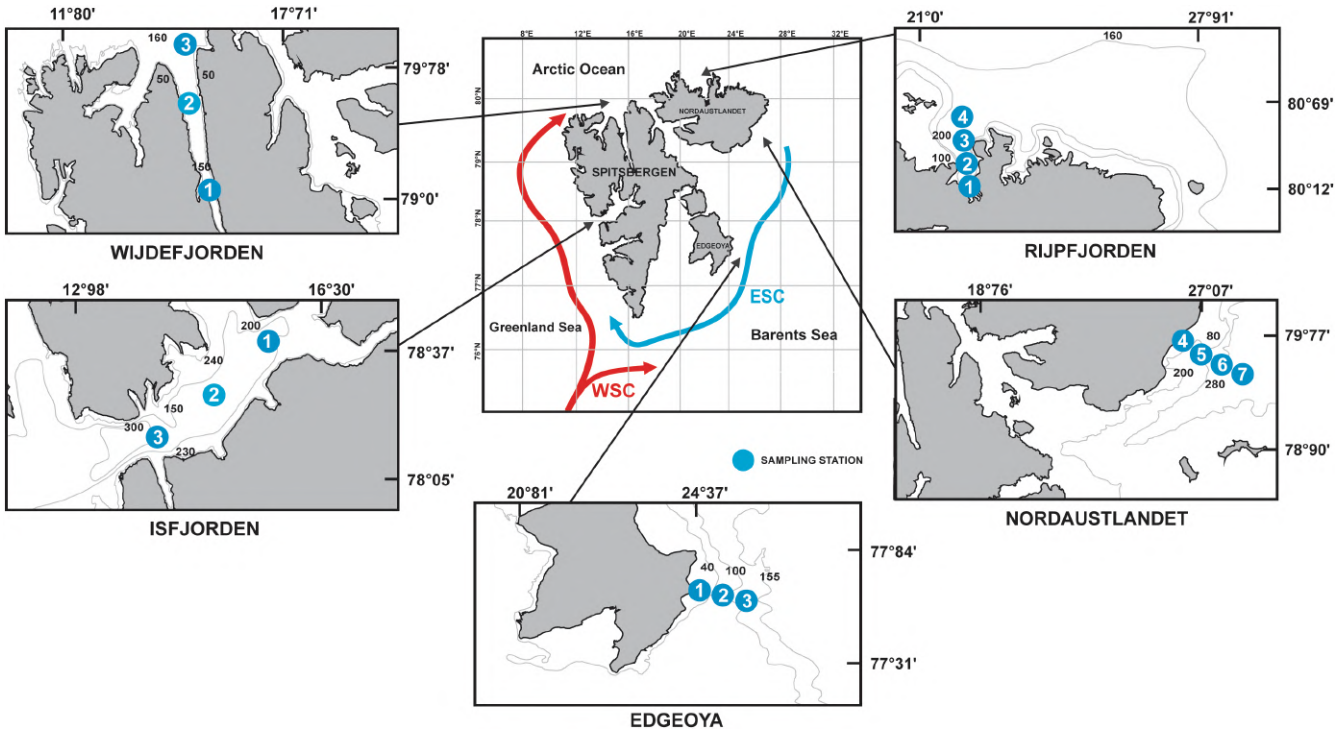


FIGURE 1 Map showing the location of sampling stations in the fjords of Svalbard Archipelago. These islands include Spitsbergen, Nordaustlandet, and Edgeøya. ESC, East Spitsbergen Current; WSC, West Spitsbergen Current.

(John W Murray, 2014). However, morphological identification is time-consuming and taxonomic expertise-demanding, making it costly and unpractical, particularly for large-scale surveys. Recently, the metabarcoding of environmental DNA (*e*DNA) samples has provided new insights into the biodiversity and ecological distribution of numerous taxonomic groups and offers an alternative to the traditional morphology-based approach (Bohmann et al., 2014; Holman et al., 2021; Pawlowski et al., 2021). Metabarcoding consists in high-throughput sequencing of short DNA barcodes that include enough information for species identification to get a comprehensive inventory of all organisms present in a given sample. For instance, short sequences derived from the 37f hypervariable region of the 18S small subunit (SSU) rRNA gene are widely used in foraminiferal metabarcoding studies (Lejzerowicz et al., 2014; Pawlowski & Lecroq, 2010). To better understand large-scale patterns of biodiversity and distribution in various groups, this method is increasingly being employed, particularly in marine environments (Armbrecht et al., 2019; Holman et al., 2021; Schoenle et al., 2021; Vargas et al., 2015). Numerous foraminiferal metabarcoding studies were conducted in various coastal areas, e.g. northern Adriatic Sea (Frontalini et al., 2020), Norwegian Sea (Pawlowski et al., 2016), west coast of Scotland (Pawlowski et al., 2014), Dutch Wadden Sea (Chronopoulou et al., 2019), and the deep sea (Cordier et al., 2019; Lecroq et al., 2011; Lejzerowicz et al., 2021; Schoenle et al., 2021). However, the application of *e*DNA metabarcoding to monitor foraminiferal diversity in the Arctic was limited to a few paleogenetic studies using foraminifera as proxies in palaeoceanographic reconstructions and investigating changes in ocean circulation patterns by targeting ancient DNA of non-fossilized foraminifera from Svalbard (Pawlowska et al., 2014, 2020).

In conventional morphology-based foraminiferal studies, the sediment samples are sieved before the specimens are sorted (Schönfeld et al., 2012). In all published foraminiferal metabarcoding studies, the DNA was extracted from unsieved sediment samples. This has some benefits, including providing a holistic view of foraminiferal diversity including small-size species and those lacking the hard shell, potentially reducing sample heterogeneity, detecting mainly small and low abundant taxa, achieving a higher number of reads, or decreasing primer bias due to the reduction in the amount of DNA template produced by the large specimens (Elbrecht et al., 2017; Leray & Knowlton, 2015). Some DNA metabarcoding studies have shown that the preprocessing of samples does not significantly alter metazoan diversity patterns (Brandt et al., 2021; Sinniger et al., 2016). However, the effectiveness of sieving versus non-sieving in the case of foraminiferal metabarcoding has not been examined yet.

The two main goals of this study are to investigate whether metabarcoding of sieved sediment is effective for the assessment of foraminiferal biodiversity and how the foraminiferal communities respond to rapid environmental shifts in Arctic marine ecosystems. Taxonomic composition, diversity, and distribution of benthic foraminifera were analyzed in fjords and open water areas in Svalbard in order to (1) compare species composition and diversity patterns inferred from sieved and unsieved sediment samples, (2) describe the spatial diversity of Svalbard foraminiferal communities, and (3) identify new potential bioindicators of water mass characteristics.

2 | STUDY AREA

The Svalbard archipelago is located north of the shallow and productive Barents Sea. The largest island is Spitsbergen, followed by Nordaustlandet and Edgeøya. Approximately 60% of the archipelago is covered by glaciers. The coastline featured numerous fjords, islets, and skerries.

The oceanography of Svalbard region is shaped mainly by the interplay between warm and saline AW and cold Arctic water (ArW), as well as locally formed water masses (Cottier et al., 2005; Hop et al., 2019). AW is transported northward along the Spitsbergen shelf edge as the WSC (Figure 1) (Blindheim & Østerhus, 2005; Loeng, 1991). WSC is one of the major heat contributors to the Arctic Ocean (Spielhagen et al., 2011), transporting heat from low latitudes into the Arctic and transferring it to the atmosphere and adjacent water masses (Saloranta & Haugan, 2004). Between 78 and 80°N, the WSC bifurcates into an eastern (Svalbard) branch and a western (Yermak) branch (Aagaard et al., 1987). The Svalbard Branch flows northeasterly, staying close to the continental margin of Svalbard (Aagaard et al., 1987). The Yermak Branch streams northwards and further recirculates southward as the Return Atlantic Current (Bourke et al., 1988). The Svalbard area is also under the influence of cold ArW that is transported from the north-eastern Barents Sea by the East Spitsbergen Current (ESC, Figure 1), also called Sørkapp Current or the Coastal Current (Sternal et al., 2014). Mixing of ArW and AW results in the formation of transformed Atlantic water (TAW) which expanded across the shelf and penetrated the fjords (Cottier et al., 2005; Nilsen et al., 2016).

Isfjorden (IS) and Wijdefjorden (WIJ) are located on the west coast of Spitsbergen, along the main pathway of AW inflow (Figure 1). Both fjords are linked directly to shelf and slope areas (Kowalewski et al., 1990; Nilsen et al., 2008) and therefore, their oceanographic conditions are shaped mainly by the inflow of AW and TAW. Isfjorden is considered to be the most AW-impacted fjord of Spitsbergen (Nilsen et al., 2016). Rijpfjorden (RIJ) is a north-facing fjord, located on the northern coast of Nordaustlandet. The oceanography of Rijpfjorden is dominated by cold ArW, with a less pronounced impact of AW. However, episodic inflows of AW may occur in ice-free periods. As such, it is considered to be a typical Arctic fjord. Most of the year, Rijpfjorden is covered by sea ice and/or drifting ice packs (Ambrose Jr. et al., 2006).

The southeastern Nordaustlandet (NAL) and the eastern Edgeøya (EDG) are strongly impacted by the presence of large ice caps, making them one of the largest glacierized areas of Svalbard (Dowdeswell et al., 1986). The tidewater cliffs supply the surrounding areas with large amounts of turbid meltwater (Julian A. Dowdeswell & Bamber, 1995). Water masses around Nordaustlandet and Edgeøya are dominated by ArW, carried by the ESC. However,

in periods of strong WSC activity, the presence of AW is also pronounced (Knies et al., 2007).

3 | MATERIAL AND METHODS

3.1 | Sampling

The samples were collected at 15 sampling stations from five localities on Western, Northern, and Eastern sides of the Svalbard Archipelago (Figure 1), including three fjord sites (Isfjorden, Wijdefjorden, Rijpfjorden) and two open marine areas in front of tidewater glaciers (Edgeøya, Nordaustlandet). Sampling station coordinates and sampling depths can be found in Table S1. Surface sediment samples were collected with the use of a box corer during the cruise of R/V *Oceania* in August 2016. The upper 2 cm of sediment has been sampled from the surface of approximately 50 cm^2. Samples for sedimentary *e*DNA analysis were split into two: one half remained unsieved and the other half has been wet sieved on 500, 100, and 63 μm sieves. A fraction smaller than 63 μm was retained. Samples were transferred to sterile containers and frozen at −20°C. In each sampling station, physical properties of the water column from a vertical conductivity-temperature-depth (CTD) profiler were obtained using a Mini CTD Sensordata SD202 at intervals of 1 s. Water temperature was reported in degrees celsius (°C), and turbidity was presented in Formazine Turbidity Units (FTU). Water masses were classified according to Cottier et al. (2005). Table S2 contains detailed information.

3.2 | Metabarcoding analyses

The genomic DNA from size fractions >500, 500–100, and 100–63 μm was extracted from 0.25 g of sediment sample with DNeasy PowerSoil Kit (Qiagen, Hilden, Germany). Approximately 10 g of unsieved part and the remained sediment fraction <63 μm were extracted using DNeasy PowerMax Soil Kit (Qiagen, Hilden, Germany). In total, five amplicon libraries per station were prepared, corresponding to the fractions >500, 500–100, 100–63, and <63 μm, as well as unsieved samples.

The foraminifera-specific 37f hypervariable region of 18S rRNA gene was PCR amplified with the primers s14F1/s15 (Barrenechea Angeles et al., 2020; Lejzerowicz et al., 2014), tagged with unique sequences of 8 nucleotides appended at 5′ ends (Esling et al., 2015). The lengths of amplified products are approximately 180 base pairs on average including the specific primers and the tags. Primer sequences and PCR conditions are detailed in Table S3. For each sample, 3 PCR replicates were obtained. PCR products were visualized by 1.5% agarose gel electrophoresis and quantified with Qubit 3.0 fluorometer (Thermo-Fisher Scientific Inc., Waltham, MA, USA). The PCR products were pooled in an equimolar mix with each duplicate located in a different pool to reach a total quantity of 100 ng of DNA. The pool was purified with High Pure PCR Cleanup Micro Kit (Roche Diagnostics GmbH, Mannheim, Germany). Library preparation was performed with TruSeq® DNA PCR-Free LT Library Prep Kit (Illumina Inc., San Diego, CA, USA) and was loaded onto a MiSeq instrument for a paired-end HTS run of 2 × 150 cycles using a v2 kit.

3.3 | Data quality control and processing

Bioinformatics analyses were performed using the web application SLIM (https://trtcrd.github.io/SLIM) (Dufresne et al., 2019). The reads were first demultiplexed using the double tag demultiplexing algorithm based on their unique barcode sequences. The software package DADA2 (Callahan et al., 2016) was used for quality trimming and filtering sequences, de-replicating sequences, inferring amplicon sequence variants (ASVs), merging of forward and reverse sequences, and detection and removal of chimeras. Subsequently, all the resulting ASVs tables were curated with the LULU algorithm (Froslev et al., 2017) to remove erroneous ASVs following the online tutorial (https://github.com/tobiasgf/lulu) with default parameters. Final quality filtering of ASVs involved the removal of unique (occurring in only one sample) and rare ASVs (having <10 reads).

The remaining ASVs were compared to the curated database of foraminiferal 18S rDNA sequences (Holzmann & Pawlowski, 2017; Pawlowski et al., 2013) and the PR2 database v4.11.1 (Guillou et al., 2013) using VSEARCH, implemented in SLIM, and BLASTN (Altschul et al., 1990) based on minimum similarity (−perc_identity 80%) and minimum coverage (−qcov_hsp 80%) for the taxonomic assignment to six taxonomic levels (phylum; class; order; family; genus; species). The representative sequences of ASVs that remained unclassified with the foraminiferal database were aligned in a stand-alone BLAST using BLAST (v2.7.1) search against the NCBI's non-redundant nucleotide database. The sequences diverging by less than 1% were considered as belonging to the same species/genus. ASVs below 99% identity were classified at the family, order, or class or as unassigned foraminifera. Finally, taxonomic compositions in terms of cluster abundance were compared among processing methods only using clusters reliably assigned at the species/genus level.

3.4 | Statistical analysis

Before statistical analyses, the ASV table was filtered to remove ASVs that were classified as planktic or non-foraminifera. For each sample, datasets of four size fractions were combined as a sieved dataset and compared to an unsieved dataset in further analysis. All statistical analyses were performed in R, version 4.1.0 (R Core Team, 2018). All formal hypothesis tests were conducted on the 5% significance level ($\alpha = .05$).

To compare the community composition among methods and size fractions, Venn diagrams were constructed using the *venn* package (Dusa, 2018). The ASVs rarefaction curves were calculated to visualize whether or when a plateau was reached based on

the number of eventually retained ASVs and reads using the iNEXT package (Hsieh et al., 2016). The species accumulation curve was also created using the function *specaccum* in the *vegan* package (Oksanen, 2007). The data of each sample were normalized using the cumulative sum scaling method available on the metagenomeSeq Bioconductor package (Paulson et al., 2013). Based on the normalized data, four alpha diversity indexes and non-metric multidimensional scaling (nMDS) on the Bray–Curtis similarity coefficient to analyze differences in the beta diversity of the community composition were calculated with the *vegan* package (Oksanen, 2007). We used the *pheatmap* package (Kolde, 2019) to create a heatmap based on Spearman's correlation. The influences of environmental factors were calculated with the *envfit* function. A global one-way analysis of similarities (ANOSIM), permutational multivariate analysis of variance (PERMANOVA), and method to fit environmental vectors onto ordination were computed using the function *anosim*, *adonis*, and *envfit* with 999 permutations and the Bray–Curtis distance matrix to test whether there were significant differences in community composition among methods and locations of sampling units.

Finally, sparse partial least squares (sPLS) regression, available in the *mixOmics* package (Le Cao et al., 2008; Rohart et al., 2017), was used for the multivariate analysis of the combined foraminiferal datasets at ASVs level to identify ASVs that were more predictive of the observed environmental response. Pairwise similarity matrices of an sPLS model with 2 components were computed and displayed by the function *cim*. This approach enabled us to identify high correlations between certain ASVs and environmental parameters.

4 | RESULTS

4.1 | CTD data

Temperature, salinity, and turbidity for all sampling stations are presented in Figures 2 and S1, respectively. AW and TAW dominated the water masses in Isfjorden (Figure 2a) and Wijdefjorden (Figure 2b), which have the highest temperatures and salinities of the investigated stations. Additionally, surface water (SW) and intermediate water (IW) were recorded at all stations in Isfjorden and station WIJ1. In Isfjorden, the highest temperature of 7.7°C was observed at the surface and progressively decreased toward the bottom to 1.4°C. Similarly, the temperature fluctuated from 5.8°C at the surface to −0.4°C near the bottom of Wijdefjorden. Water temperatures above 0°C were noted up to 111 m in depth at the station WIJ1 and in the whole water column at other stations. Salinity was the lowest at the surface, reaching 28.1 in Isfjorden and 32.5 in Wijdefjorden, respectively. The lowest salinity was recorded at inner Isfjorden and Wijdefjorden (IS1 and WIJ1) and the highest values near the mouth of these fjords (IS3 and WIJ3). The turbidity increased from the inner fjord (0.1 FTU) toward the fjord's mouth (particularly up to 12.5 FTU) in Isfjorden, reaching its maximum in the SW at the station IS3. In contrast, turbidity was the highest in the surface layer and decreased from the inner fjord toward the fjord mouth, ranging from 5.2 to 0.1 FTU.

Rijpfjorden was characterized by the lowest near-bottom temperatures (as low as −1.7°C) among the studied fjords (Figures 2c and S1). Three sites (RIJ1, RIJ2, and RIJ3) reported cold and saline winter cooled water (WCW) in addition to other water masses (AW, TAW, SW, and IW). The temperature ranged from 5.2°C at the top to −1.7°C near the bottom, with the salinity varying from 33.4 to 35.4 throughout the water mass. The lowest near-bottom temperature was noted at the station RIJ1. The water temperature of the whole water column was observed to be above 0°C at station RIJ4. Turbidity reached over 12.5 FTU at the station RIJ 3 in the near-bottom water layer and decreased to 0.1 FTU toward the mouth of the fjord.

In the region of eastern Svalbard, the Nordaustlandet stations were generally under influence of TAW, whereas AW was noted only at the glacier-distant station NAL7 (Figure 2d). NAL4 was the sole station where neither TAW nor AW was detected. SW and IW were other water masses recorded near the Nordaustlandet. Water temperature oscillated between 4.4°C at the surface to 0.2°C near the bottom. Salinity at the surface ranged from 33.6 to 35.3 and increased toward the glacier-distant stations. The water column had relatively low turbidity (<1 FTU). The only exception was glacier-proximal station NAL4, where turbidity reached 54.8 FTU, which was the highest value of all studied sites.

Edgeøya stations were the most distinct locations, with the absence of Atlantic-origin waters (Figure 2e). The IW was detected at all stations, while local water (LW) occurred only at station EGD3. Toward the bottom of the stations, the temperature in the water column varied between 4.2 and − 0.5 and salinity ranged from 33.2 to 34.9. The highest turbidity was recorded at EGD1; it increased with depth to reach 48.8 FTU near the bottom. At the other stations, turbidity values oscillated from 0.3 to 4.3 FTU.

4.2 | Metabarcoding data

We obtained a total of 5,968,786 raw paired-end reads. After bioinformatic processing, the numbers of the raw reads were reduced to 5,579,202 with 4,836,419 in a sieved dataset, and 742,783 in an unsieved dataset. The number of reads per sample is indicated in Table S4. One sample, unsieved IS2, produced a low number of reads and was not included in the analysis of foraminiferal diversity. After LULU curation step and strict filtering of ASVs, 1384 ASVs (1354 ASVs of sieved and 1053 ASVs of unsieved samples) representing 5,483,500 reads (98.28% of the total reads count) were retained for downstream analysis (Table S5). The average numbers of sequences per station were 317,306 for the sieved and 51,976 for the unsieved datasets.

The rarefaction curves were plotted at the sample level based on the number of retained ASVs and reads (Figure 3). The

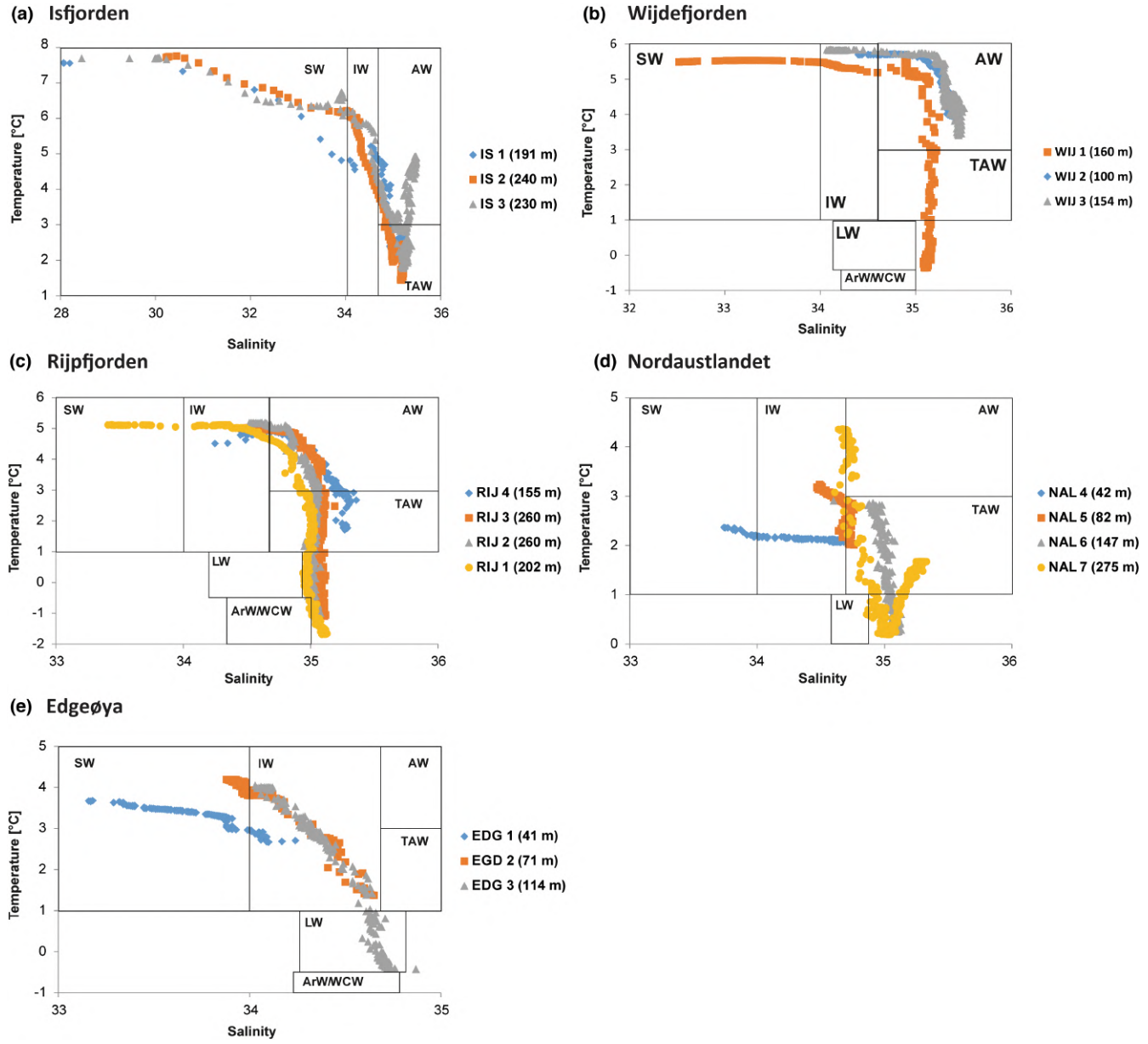


FIGURE 2 Temperature (°C) and salinity in Svalbard stations: (a) Isfjorden—IS, (b) Wijdefjorden—WIJ (c) Rijpfjorden—RIJ, (d) Nordaustlandet—NAL, and (e) Edgeøya—EDG. Values in brackets are water depth at the time of sampling. ArW, Arctic water; AW, Atlantic water; IW, intermediate water; LW, local water; SW, surface water; TAW, transformed Atlantic water; WCW, winter cooled water. Water masses are classified after Cottier et al. (2005).

rarefaction curves showed that the filtered ASV datasets reach saturation levels, indicating that most of the diversity had been captured and allowing for richness comparison among samples for all individual stations of each location (Figure 3a,b) and both methods (Figure 3c). The species accumulation curves of the samples for each location increase with the number of samples, indicating that the existing sample size could meet the needs of this study (Figure S2). Considering the accumulation of ASVs richness across two datasets, sieved datasets exhibited a higher saturation degree, respectively, and the species richness of WIJ, IS, and RIJ stations is higher than the individual station of EDG and NAL.

4.3 | Taxonomic composition of foraminiferal metabarcodes

Overall, the retained sequences were assigned to 1384 foraminiferal ASVs. Among them, 758 ASVs were assigned to the class Monothalamea, 252 ASVs were assigned to the class Globothalamea, and only 14 ASVs were assigned to the class Tubothalamea (Table S5). The 360 ASVs, classified as Foraminifera_X, had low similarity levels and could not be assigned to any existing clades. More than half of the ASVs (51.73%) were assigned with low similarity (<0.9).

The sieved and unsieved sediment DNA samples resolved comparable communities at the class level (Figure 4). Pairwise comparisons

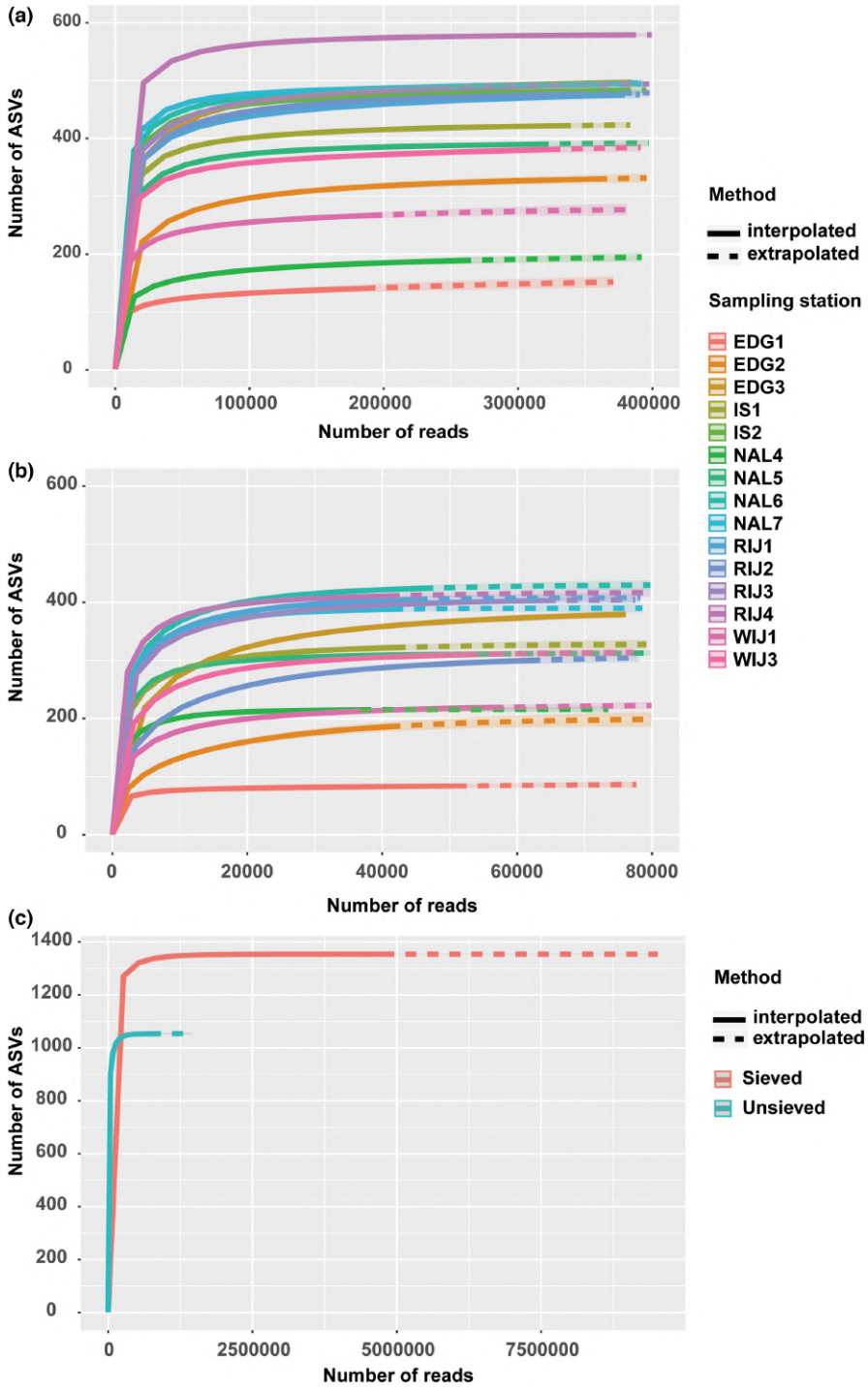


FIGURE 3 Rarefaction curves showing the relationship between sequencing depth and species richness in amplicon sequence variant (ASVs) of the 37f libraries from 15 stations of individual sieved samples (a), unsieved samples (b) and combining of sieved and unsieved samples (c). The solid line is the rarefaction curve based on the abundance of observed reads, and the dotted line is the extrapolation curve based on the abundance of extrapolated reads.

indicated no overall significant differences in community composition between sieved and unsieved datasets (ANOSIM statistic $R < 0$, $p > .05$ and PERMANOVA, Table S6). In Figure 5a, the Venn diagram showed that 1023 ASVs (corresponding to 97.23% of the reads) were shared among sieved and unsieved samples. The sieved dataset had 331 unique ASVs, while unsieved dataset comprised only 30 unique ASVs (corresponding to 0.1% and 2.67% of the reads, respectively).

Taxonomic composition in sieved samples noticeably changed between size fractions (Figures 5b,c). The <63 μm fraction comprises 1151 ASVs, corresponding to 83% of ASVs (Figure 5b). It also recovered more unique ASVs (177, corresponding to 13.07% of ASVs) than any other fractions. Shared foraminiferal ASVs among fractions including 424 ASVs (corresponding to 82.39% of the reads), mostly belonged to Monothalamea (59.72%), and Globothalamea (Rotaliida 25.17%, Textulariida 10.60%) as shown in Figure S3. In all fractions (Figure 5c), the monothalamous taxa made up from 58% to 80% of reads. Non-described monothalamids dominated in the 500–100 μm and < 63 μm fractions (30.48% and 33.85% of reads, respectively). For multichambered globothalamids, order Textulariida accounted for 15%–20% of the

FIGURE 4 Proportions of reads (a) and ASVs (b) assigned to different foraminiferal classes detected in sieved and unsieved samples at different sites.

reads in 500–63 μm fractions, while Rotaliida represented more than 20% of reads in >63 μm fractions. Interestingly, most reads of unique ASVs were assigned to specific foraminiferal groups in each fraction, e.g., Foraminifera_XX (70.69%) in >500 μm fraction, Rotaliida (66.21%) in 500–100 μm fraction, Textulariida (39.28%) in <63 um fraction, and Clade Y of Monothalamea (70.24%) in 100–63 μm fraction, see Figure S3.The taxonomic composition of benthic foraminifera also changed between the locations. At the class level (Figure 4), the monothalamous taxa were the dominant group which accounted for an average of 56.06% and 61.77% of total ASVs and reads in both datasets, respectively. The highest proportion of monothalamiids (95.80%) was observed at the station EDG1 in the sieved dataset. The contribution of monothalamiids decreased in the deeper EDG stations in favor of the class Globothalamea. Comparatively, the average proportions of ASVs and reads assigned to Globothalamea were 22.60% of the ASVs and 30.75% of reads, respectively. The highest relative abundance of Globothalamea (80.80%) occurred in the WIJ3_Sieved sample. The Tubothalamea represented only a minor part of the total community (0.06% of ASVs and 7.42% of reads on average).

The variations of foraminiferal assemblages between different sampling localities were also reflected in the taxonomic composition of foraminiferal assemblages at the lower taxonomic level. To compare species composition in each station, all ASVs which had an identity percentage with the reference database of more than 99% and no <10 reads were picked and those attributed to the same taxa were merged (Figure 6).

In the stations located at the western coast of Spitsbergen (IS, WIJ), foraminiferal communities were dominated by genera *Psammophaga* and *Micrometula*, which together made up to 57.35% of reads. The WIJ1 was the only station where the majority of sequences belonged to a textularid *Reophax* sp. and a rotalid *Stainforthia* sp. Another rotalid species *Nonionellina labradorica* was present at all IS and WIJ stations, mainly distributed in WIJ1 (sieved: 59.99%, unsieved: 39.46%), WIJ3_Sieved (11%), and IS2_Sieved (6.5%). The northern stations (RIJ) were dominated by monothalamiids assigned to Clade Y and Monothalamea_XXX. At the outermost station RIJ4, also higher percentages of *Psammophaga* sp. and *Psammosphaera* sp. sequences occurred. At the stations located at eastern Svalbard (NAL), the number of Clade Y sequences decreased toward the glacier-distant stations, while the percentage of Monothalamea_XXX increased. Also, glacier-distal stations were characterized by a higher proportion of Globothalamea, mainly *Reophax* sp. and *Stainforthia* sp. The foraminiferal community in EDG1 was dominated by a monothalamid *Psammosphaera* sp. (up to 90%), while *Hippocrepinella* sp. dominated in EDG2 and EDG3 (from 11% to 84%). Also, the percentage of *Reophax* sp. sequences increased toward the glacier-distal stations, from <1% to reaching up to 34.68% at the station EDG3_Sieved.

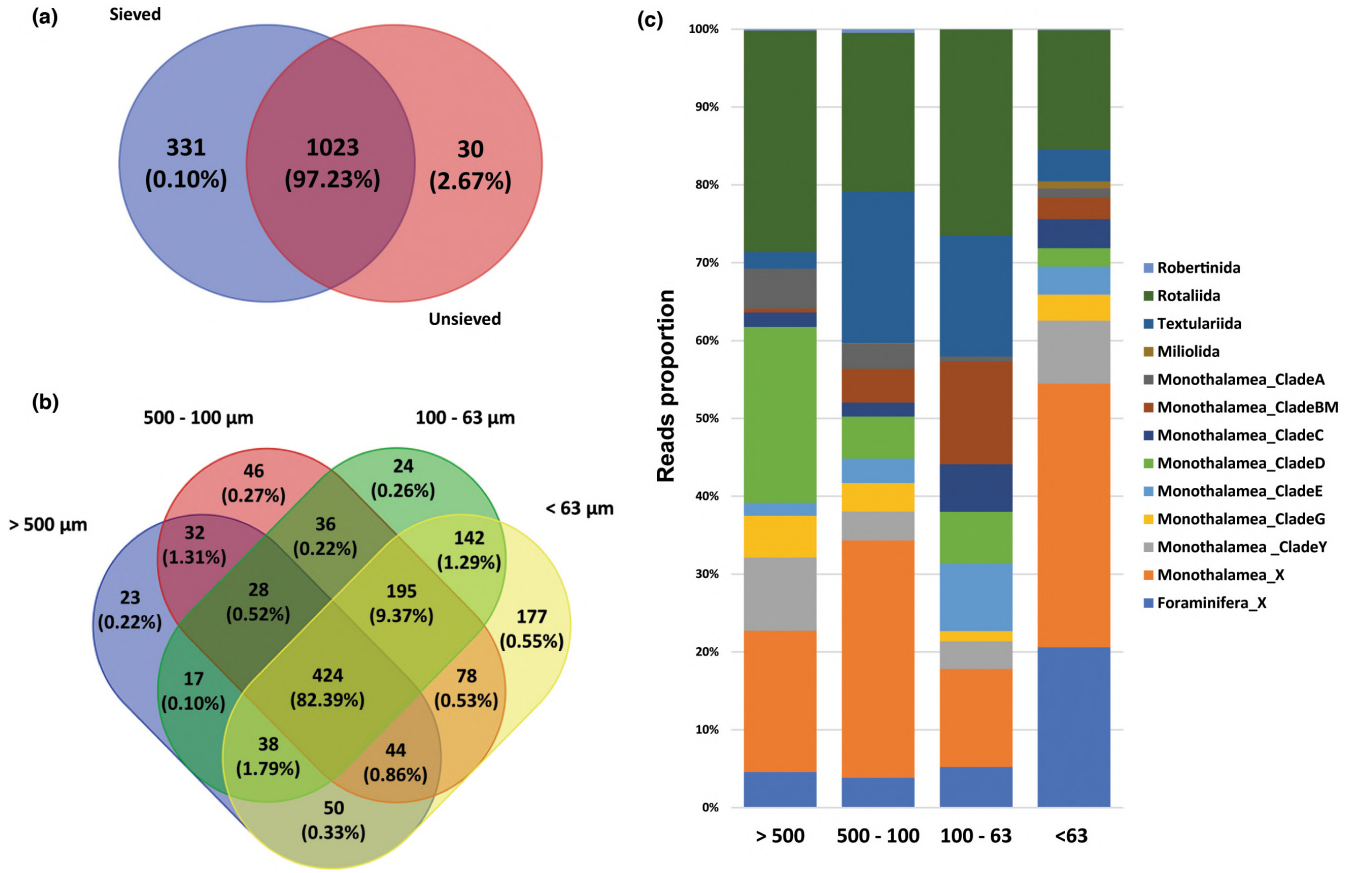
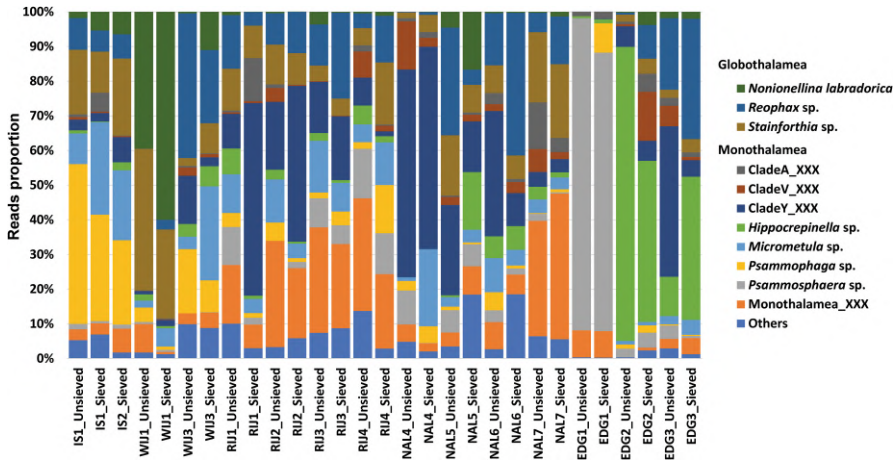


FIGURE 5 Venn diagrams showing the shared and unique numbers ASVs and proportion of reads for the sieved (combined fractions) and unsieved samples (a), and between different size fractions (b). Bar plots represent the proportion of reads assigned to the taxonomic composition of each fraction by order/clade (c).

FIGURE 6 Patterns of relative abundance of dominant genera and species in the sieved (combined fractions) and unsieved sediment samples for each location.



4.4 | Alpha and beta diversity patterns

4.4.1 | Alpha diversity

The four alpha diversity indices (Observed ASVs, Chao1, Simpson and Shannon) were measured separately for sieved and unsieved datasets (Figure 7) and showed clear variation between different locations. On the one hand, the number of ASVs varied substantially depending on sample treatments and locations. In terms of sample treatment, sieved samples recovered higher Observed ASVs and Chao1 indices (Figure 7a,b), but not Simpson or Shannon (Figure 7c,d). On the other hand, the measured alpha diversity indices tended to increase with increasing distance from the glacier. In general, the alpha diversity indices of the fjords (IS, WIJ, RIJ) were higher than those of open sea areas (NAL, EDG), indicating higher foraminiferal diversity in fjords.

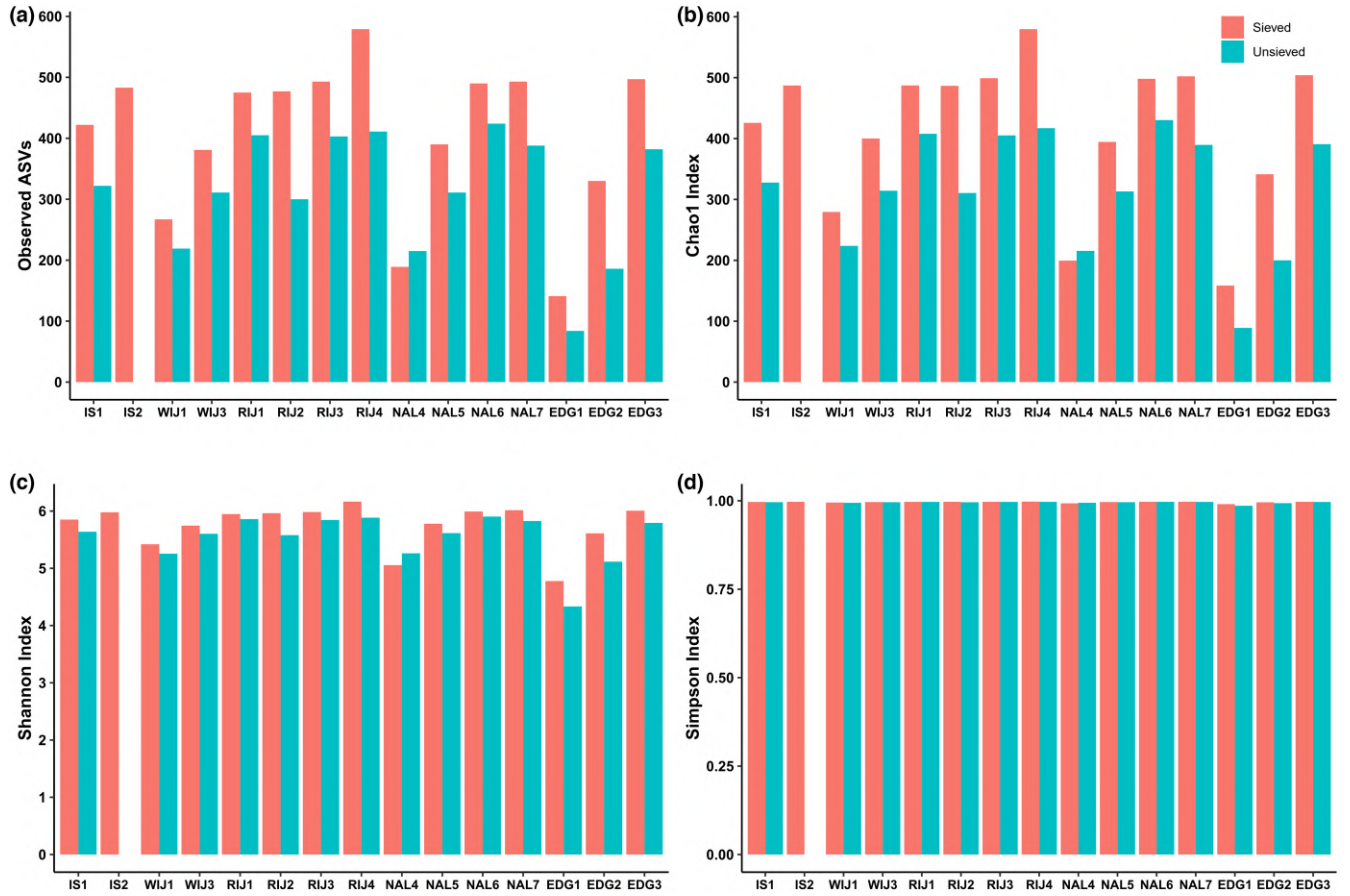


FIGURE 7 Bar plots of alpha diversity indices, including community richness (a: Observed ASVs, b: Chao1) and diversity (c: Shannon, d: Simpson) for the 15 sieved samples and 14 unsieved samples using retained ASV read abundances.

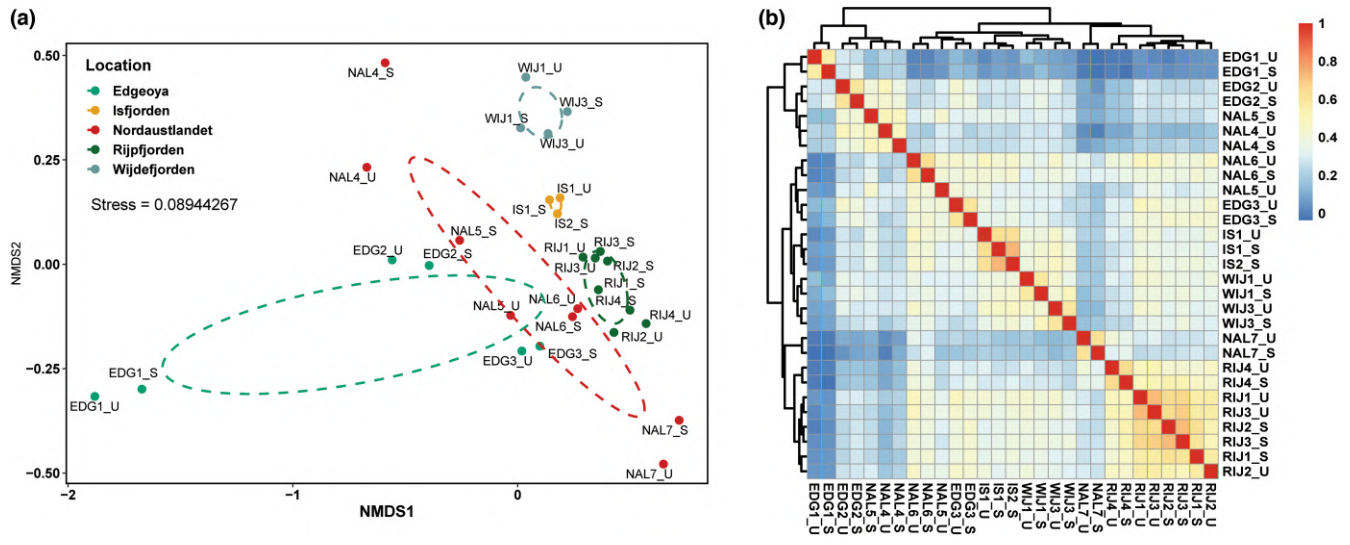


FIGURE 8 Community structuring of benthic foraminifera using nonlinear multidimensional scaling based on Bray–Curtis distance similarity coefficient (a) and heatmap based on Spearman's correlation coefficient for fractions and unsieved samples (b). Stress value is displayed on the plot.

4.4.2 | Beta diversity

The nMDS and heatmap of all stations (Figure 8) supported the findings of the ordinances, showing that foraminiferal communities detected in the sieved samples differed from those detected in the unsieved samples, but not significantly (Table S7). The nMDS and heatmap patterns also revealed the spatial distribution of the foraminiferal communities among the different localities. In Figure 8a,

the nMDS analyses produced a similar pattern with sieved and unsieved datasets, although community segregation was observed in ordinations of EDG and NAL sites. The communities were distinctly separated between fjord stations and opened sea stations, as shown by low-stress values. Although the fjord samples formed tight clusters, the samples from each fjord were not overlapping with the samples from other fjord locations. On the contrary, the communities obtained from EDG and NAL sites formed clusters with much larger internal compositional differences and have an overlap between the two sites. Heatmap further clarified the community structuration with the stations and datasets (Figure 8b), which were not visible on the nMDS (except for NAL5). The sampling sites were grouped in two main clusters: cluster 1 aggregating 3 stations (EDG1, EDG2, NAL4) and cluster 2 comprising the 12 remaining stations that were grouped into three subclusters. The stations of two fjords (IS and WIJ) had homogeneous communities and formed one separate subcluster. Two other subclusters are formed by (i) NAL6, NAL5, EDG3, and (ii) all RIJ stations and NAL7.

4.5 | sPLS prediction analysis

The results of the sPLS regression allowed the detection of several foraminiferal ASVs lineages for which relative sequence abundance was correlated with environmental parameters (Figure 9 and Table S8). The sPLS regression and subsequent hierarchical clustering suggested that the data were separated into three clusters (Figure 9). These include lineages identified as potential indicators of water mass characteristics.

In cluster I.A, the ASVs exhibited a positive correlation with turbidity and negative correlation with factors such as depth, the salinity of bottom water, and temperature of the SW, with the ASVs predominantly affiliated as members of monothalamiids: *Hippocrepinella* sp. (ASV3, ASV10, ASV21), *Psammosphaera* sp. (ASV6), *Saccamminidae* sp. (ASV12), CladeY_spallogJAP (ASV22), STICKY_ICE (ASV39), *Pelosinella fusiformis* (ASV82), and globothalamids: *Buliminella* sp. (ASV14), *Nonionellina labradorica* (ASV20), *Cibicides* sp. (ASV91). The ASVs within cluster II revealed a strong and positive correlation with temperature as well as a negative correlation with salinity in the SW masses. This cluster included globothalamids: *Stainforthia* sp. (ASV1), *Virgulinella fragilis* (ASV79), *Reophax sp*. (ASV92), *Cibicidoides fletcheri* (ASV23), and monothalamiids: *Psammophaga* sp. (ASV25, ASV34, ASV42, ASV53, ASV64), *Micrometula* sp. (ASV4), ENFOR2_EnvHabIC19 (ASV45), CladeA (ASV85, ASV32). Some ASVs belonging to cluster II also had a strong positive correlation with the depth and bottom water salinity (Table S8).

Additionally, we observed positive correlations with depth and salinity in clusters III.A and III.B. Most of ASVs belonging to these clusters were classified as undetermined Monothalamea: Monothalamea_XX (39 ASVs), ENFOR XX (20 ASVs), CladeG (19 ASVs), CladeTIN (16 ASVs). In terms of ASV abundance, the dominant ASVs included environmental monothalamiids (ASV11, ASV44, ASV81, ASV66, ASV93, ASV88, ASV69, ASV78), CladeC_spsaccam (ASV59), *Gloiogullmia* sp. (ASV63), *Cibicides* sp. (ASV38), *Nonionella auris* (ASV31), rotalid (ASV96), *Reophax* sp. (ASV41, ASV13, ASV18), *Stainforthia* sp. (ASV77). Cluster III.C had positive associations with the depth, the bottom water salinity, and the surface temperature. ENFOR2_XXX (ASV408, ASV136, ASV125) exhibited their highest abundances in this cluster.

5 | DISCUSSION

5.1 | Is pre-sieving useful for foraminiferal metabarcoding?

The methodological aim of this study was to compare the results of metabarcoding analyses based on sieved and unsieved sediment samples. Sieving is a common procedure in the conventional microscopic study of foraminiferal assemblage analyzing mainly hard-shelled, multi-chambered taxa preserved in fixed and dried sediment samples (Schönfeld et al., 2012). In contrast, metabarcoding studies of unsieved sediment samples usually provide a foraminiferal assemblage dominated by poorly known, soft-walled, or naked monothalamous taxa (Lecroq et al., 2011; Pawlowski et al., 2014). Because of this, it is difficult to compare the results of traditional morphology-based studies with those of metabarcoding analysis, which provide very different types of data (Frontalini et al., 2020).

As shown by our study, the taxonomic composition differed between the fractions. For example, the order Rotaliida was the most abundant in 500–100 and 100–63 μm size fractions. Also, another hard-shelled order Textulariida, which is microscopically studied in the 500–100 μm fraction, in metabarcoding data is present mainly in fractions 500–100 and 100–63 μm (Figure 5c). This is congruent with the rotaliids and textulariids dominating microscopic assemblage found in >63 μm sieved fraction. On the other hand, the smallest fraction (<63 μm) was dominated by monothalamiids and undetermined Foraminifera (Figure 5c), which may suggest the presence of some unknown, tiny monothalamous species. Compared to morphological approaches, DNA-based metabarcoding provides a more holistic picture of foraminiferal communities, including tiny species present in <63 μm fraction as well as those that are not preserved in dried material used in conventional surveys.

We also observed some differences between sieved and unsieved samples regarding the alpha diversity. The total number of recovered ASVs was clearly higher in sieved than in unsieved samples (approximately 30% ASVs). However, this could be explained by the difference in the number of DNA extraction, PCR amplification, and sequencing depth. In the case of sieved samples, the datasets included four DNA extractions, one for each size fraction, while only one DNA extraction was performed for non-sieved sediment samples. Further, sieving probably reduces PCR inhibitors as well as non-targeted taxa in the samples. In total, the number of sequences obtained for sieved fractions was several times higher, allowing for the detection of higher diversity in sieved compared to unsieved samples. However, no significant difference between the sieved/

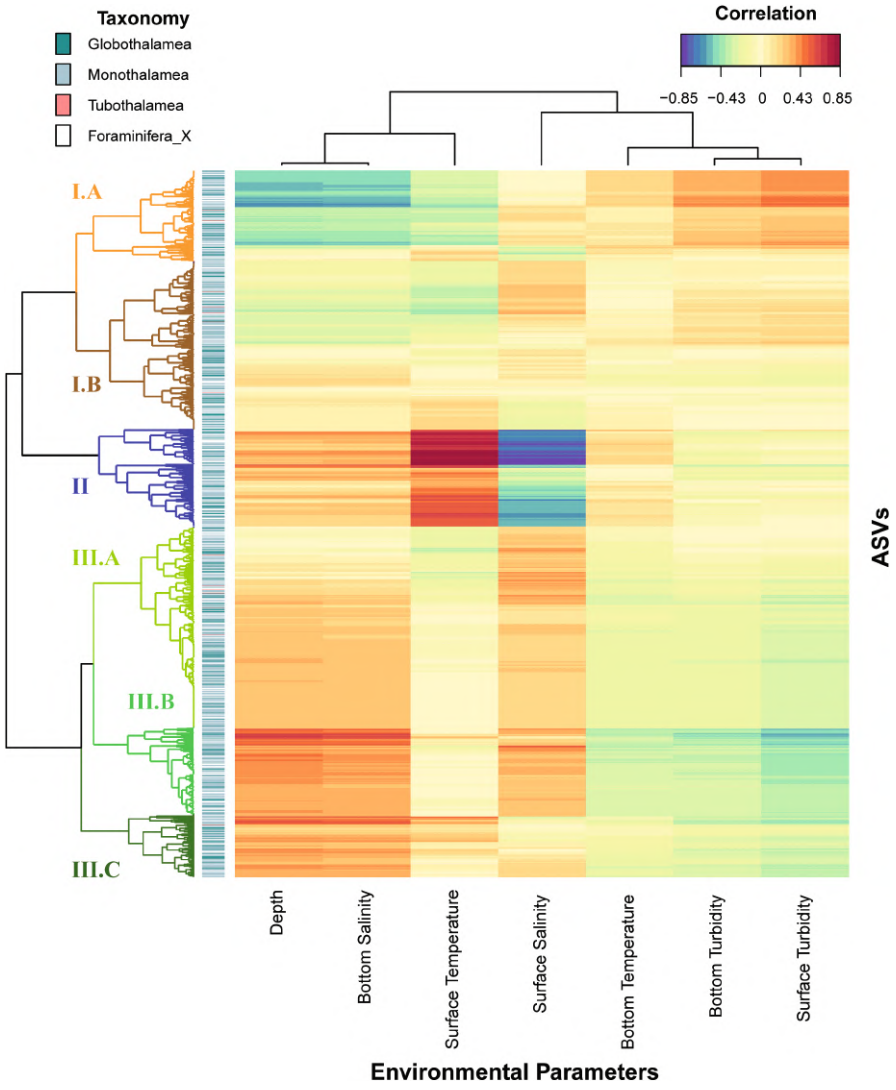


FIGURE 9 Clustered image map (CIM) of the first two sPLS dimensions, displaying pairwise correlations between foraminiferal ASVs of combined unsieved and sieved samples associated with environmental parameters. Correlations between ASVs and environmental parameters are depicted as a clustered heat map (detailed results in Table S8). Red and blue indicate positive and negative correlations, respectively.

unsieved samples was observed in alpha diversity measures such as Shannon's and Simpson's that take abundance and evenness of the sample into consideration as shown in Figure 7. Although sieving might have been predicted to lead to a reduction in alpha diversity due to the loss of microfauna and extracellular debris, this has not been observed in previous studies (Brannock & Halanych, 2015; He et al., 2020). In addition, nMDS of the beta diversity matrices and correlation test showed sieved and unsieved samples clustered together (Figure 8, Table S7), indicating that there is no significant difference in community composition inferred by the two methods.

To conclude, the decision of whether the sediment samples should be sieved or not shall be based on the type of questions one wants to answer with metabarcoding data as well as the composition and characteristics of initial samples. Sieving of samples destinated for metabarcoding analysis might be useful if particular groups of foraminifera are targeted (e.g., Rotaliida, Textulariida), for example, to compare with microscopic studies or to identify some tiny species present in fine size fractions. In general, the unsieved samples provide a more complete overview of the taxonomic composition of the foraminiferal community. However, as shown by our study, both metabarcoding datasets reveal similar trends in foraminiferal diversity. Size sieving might have some advantages; however, it also has some drawbacks, as (i) it is time-consuming, (ii) requires higher volume samples, and (iii) there is a possibility of cross-contamination between samples. Therefore, either extracting DNA directly from sediment or after sieving should be carefully considered when evaluating foraminiferal communities across metabarcoding studies.

5.2 | Distribution patterns of foraminifera in Svalbard

The most striking result of this study is the variations of foraminiferal assemblage between different sampling localities. The taxonomic composition of foraminiferal communities is generally specific to each location (Figure 7). Each fjord forms a separate cluster (IS, WIJ, and RIJ) and only some stations at open-water areas overlap with each other (EDG and NAL) as Figures 7 and 8. However, there are some similar trends documented at different locations, such as the high proportion of monothalamiids in near-glacier settings (Figure 6)

or the increase of alpha diversity from glacier proximal/inner to glacier-distant/outer stations (Figure 7), which are in agreement with the previous morphology-based studies (Hald & Korsun, 1997; Majewski et al., 2005; Sabbatini et al., 2007).

The high-Arctic settings are usually considered as a cold system influenced at different levels by ArW during summer to late autumn (Wallace et al., 2010), and covered by sea ice in winter (Ambrose Jr. et al., 2006; Dai et al., 2019; Leu et al., 2011). However, the increased influence of AW and winter sea ice loss is observed in recent years (Dahlke et al., 2020; Nilsen et al., 2008; Pavlova et al., 2019). We speculate that hydrographic conditions would lead to isolating populations from different settings and creating unique structures of the foraminiferal community.

It is well known that the unique habitats of fjords can support a high diversity and distinct biological communities (Gooday et al., 2005; Hald & Korsun, 1997; Majewski et al., 2005; Sabbatini et al., 2007; Walseng et al., 2018; Włodarska-Kowalczuk et al., 2013). Fjords create a variety of habitats suitable for specific species, where many species can converge and reach high population densities. Western Spitsbergen fjords are among the most AW-impacted areas. Both Isfjorden and Wijdefjorden are directly connected to the slope and shelf areas, which enables AW penetration into the fjords. Moreover, Isfjorden stations are located in the central basin of the fjord, which resulted in limited glacial influence. This led to the formation of foraminifera communities characterized by a relatively high proportion of globothalamids and the presence of monothalamiids community dominated by the genera *Psammophaga* and *Micrometula* (Figure 6). Similar distribution patterns were previously observed in west Spitsbergen fjords (Majewski et al., 2005; Sabbatini et al., 2007). Only station WIJ1 displayed a unique structure with a clear dominance of Rotaliida (Figures 4 and 6). Station WIJ1 is located in the inner fjord, close to the glacier termini and is influenced by turbid meltwater runoffs. The dominance of Rotaliida contradicts previous studies, indicating that glacier proximal settings are dominated by monothalamous foraminifera (Pawłowska et al., 2016; Sabbatini et al., 2007). However, this distribution pattern could be also explained by the natural patchiness of foraminiferal distribution.

In the northern site (RIJ), the dominant component of foraminiferal assemblages was undetermined monothalamiids (Figures 4 and 6). The dominance of these small, soft-walled species was previously observed in areas characterized by close to the glacier-proximal zone and influenced by freshwater inputs (Pawlowska et al., 2014; Sabbatini et al., 2007). Northern Svalbard in general and Rijpfjorden in particular are considered to be a typical Arctic setting, where sea ice forms in autumn and lasts until summer. Also, the drifting ice pack is often transported to the fjord during the summer (Ambrose Jr. et al., 2006), which leads to the formation of cold and saline WCW. This process may create a more ideal environment in inner fjords where monothalamiids thrive (Gooday et al., 2005; Korsun & Hald, 1998; Sabbatini et al., 2007).

Foraminiferal communities of open water areas (eastern Svalbard) have generally lower diversity and form different groups compared to those from western Svalbard fjords. Stations from the regions of Nordaustlandet and Edgeøya are located in front of large tidewater glaciers, releasing large amounts of turbid meltwater (Figure S1). However, only Nordaustlandet was influenced by AW and TAW, while the Edgeøya oceanographic conditions were shaped mainly by LW masses. These led to the creation of distinctly different foraminiferal communities. NAL stations were characterized by a wide range of undetermined monothalamiids, while the EDG stations were dominated by a few monothalamous species representing genera *Hipocrepinella* and *Psammosphaera* (Figure 6). Both genera were previously recorded in shallow-water assemblages of Spitsbergen fjords (Gooday et al., 2005; Majewski et al., 2005). In particular, station EDG1 exhibited a unique foraminiferal community, composed almost exclusively of *Psammosphaera* sp.

5.3 | Influence of AW on foraminifera community

The responses of benthic foraminifera to alterations in temperature and salinity in the water column are common and include expansions or retractions of distribution ranges or changes in assemblage compositions (Dong et al., 2019; Langer et al., 2013; Weinmann & Goldstein, 2016). We hypothesize that the composition of foraminiferal communities in our data resulted from water mass conditions. This hypothesis is strengthened by the clear clustering of the community in groups corresponding to different oceanographic regimes, in which stations from regions impacted by AW and/or sea ice clustered separately.

As shown by the nMDS plot and heatmap (Figure 8), the separation between two main clusters has a strong relationship with the characteristics of water masses. Cluster 1 comprises exclusively the stations of the eastern part of the archipelago (EDG1, EGD2, NAL4, NAL5), characterized by colder, and less salty water, associated with turbid glacial meltwater (Meslard et al., 2018). On the contrary, Cluster 2 includes mainly stations from the western and northern part of Svalbard where the impact of warmer and more saline AW was much more pronounced, as confirmed by our CTD profile (Figure 2). Also, subclusters that formed within cluster 2 reflected different impacts of AW. The first subcluster comprises stations (NAL6, EDG3) located in the glacial-distant regions of Nordaustlandet and Edgeøya, influenced mainly by TAW. The second subcluster (IS1, IS2, WIJ1, WIJ3) includes the most AW-impacted stations located on the western coast of the archipelago, while the third subcluster (NAL7, RIJ1, RIJ2, RIJ3, RIJ4) is composed of stations located in north-eastern Svalbard, influenced both by the inflow of AW and WCW.

The increased AW inflow, higher light availability, and the decline of sea ice around Svalbard affect the primary productivity, changing both the timing of phytoplankton bloom and phytoplankton community structure. This may have significant effects on food web dynamics, affecting higher trophic levels, including benthic communities (Csapó et al., 2021). On the other hand, recent model projections indicated low mean habitat loss of benthic macrofauna

under recent climate changes, which questions the vulnerability of Arctic benthos to atlantification (Renaud et al., 2019). This stands in clear opposition to the morphological observations that testate foraminifera communities from Svalbard fjords revealed significant changes, both in terms of abundance and species composition, related to the influence of AW (Kujawa et al., 2021). Our study confirms the impact of AW on foraminiferal communities, suggesting that AW is one of the primary factors shaping the benthic foraminifera assemblages and thus foraminifera may be potential indicators of atlantification.

Through sPLS analysis of combining datasets, we identified foraminiferal taxa that could become potential bioindicators of "atlantification." This group of species includes some monothalamiids belonging to genera *Psammophaga* and *Micrometula* as well as some undetermined monothalamous species belonging to environmental lineage ENFOR2 and Clade A. The genera *Psammophaga* and *Micrometula* are widespread in many coastal areas including polar regions (Altin-Ballero et al., 2013; Gooday et al., 2011) and are considered as bioindicator candidates in several studies (Pawlowski et al., 2014; Smith & Goldstein, 2019). However, the limited knowledge about the ecology of those taxa, as well as lacking information on their distribution in the Nordic Seas precludes making any general conclusions. Among potential bioindicators, there are also some globothalamids, such as *Stainforthia* sp., *Virgulinella fragilis*, *Reophax* sp., and *Cibicidoides fletcheri*. *Cibicidoides fletcheri* is common in the North Atlantic (Dorst & Schönfeld, 2013), but to the best of our knowledge, it was not recorded in Svalbard before. One of the major signs of increasing inflow of AW (so-called atlantification) is the northward shift of boreal species, the trend observed in the case of zooplankton, fish, and benthic organisms (Csapó et al., 2021). A recent morphological study of Svalbard foraminifera revealed the presence of boreal species *Melonis affinis* in the northern part of the archipelago (Kujawa et al., 2021). *Reophax* sp. and *Stainforthia* sp. are commonly found in Svalbard (Hald & Korsun, 1997). Also, species belonging to the genus *Reophax* are considered as indicators of AW (Majewski et al., 2009). The agreement of our results with previous morphology-based studies further proves the potential use of foraminiferal metabarcoding in biomonitoring studies.

Apart from indicators of atlantification, we identified monothalamiids that show a strong correlation with turbidity, but not with depth or salinity. These species included various monothalamids assigned to *Hippocrepinella*, *Psammosphaera*, *Saccamminidae*, as well as the CladeY_spallogJAP, and several unassigned STICKY_ICE ASVs. This correlation is particularly strong in stations closer to the coast, which is probably caused by enhanced turbidity due to sediment-laden meltwater plumes. All these monothalamids were previously recorded in Svalbard (Gooday et al., 2005; Majewski et al., 2005; Pawłowska et al., 2016) and were abundantly sequenced in the settings characterized by a high level of environmental disturbance, suggesting that they are highly resistant to environmental disturbance (Pawlowski et al., 2014). Moreover, morphological studies reported *Hippocrepinella* sp., *Psammosphaera* sp., *Saccamminidae* sp., in the shallow-water parts of the fjords, located close to meltwater outflows (Gooday et al., 2005; Majewski et al., 2005; Sabbatini et al., 2007). These findings underline how important can be to include soft-walled monothalamous foraminifera in metabarcoding studies to enhance limited knowledge about their ecology for potential use in biomonitoring.

6 | CONCLUSIONS

This study is the first to use high-throughput sequencing to comprehensively analyze the foraminiferal communities within marine sediments from Svalbard, which provide better knowledge of foraminiferal diversity and distribution patterns in the Arctic's fjords. The DNA sequencing results from sieved and unsieved sediment revealed a high diversity of the Svalbard foraminifera compared to traditional morphology-based studies and variation in the taxonomic composition of foraminiferal communities from five sampling areas. Foraminiferal diversity and species richness increased from glacier proximal/inner to glacier-distant/outer stations and were higher in the fjords than in the open water. Moreover, the structure of foraminiferal community is clearly influenced by different water masses, with a particular impact of AW in the Svalbard region. Numerous potential molecular foraminiferal bioindicators for water mass characteristics were identified. This should be confirmed by analysing more samples from reference areas in North Atlantic. With the increasing numbers of metabarcoding studies, the impact of climate warming and associated oceanographic changes on Arctic benthic communities could be better assessed and expanded to those organisms that are not covered by the conventional morphological approach.

ACKNOWLEDGMENTS

The research was financially supported by the Norwegian Financial Mechanism for 2014–2021, project no. 2019/34/H/ST10/00682. Jan P and IBA were supported by the Swiss National Science Foundation grant 31003A_179125. Joanna P was also supported by the grant no. 2018/31/B/ST10/01616 funded by the National Science Centre in Poland. We express our thanks to the captain and crew of the R/V *Oceania*.

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

All raw sequencing reads have been deposited in the NCBI Short Read Archive (SRA) database under Bioproject accession number PRJNA768352.

ORCID

Ngoc-Loi Nguyen https://orcid.org/0000-0001-9452-4830
Joanna Pawłowska https://orcid.org/0000-0001-9383-9785
Inès Barrenechea Angeles https://orcid.org/0000-0002-8051-4110

Marek Zajaczkowski https://orcid.org/0000-0002-3823-7359
Jan Pawłowski https://orcid.org/0000-0003-2421-388X

REFERENCES

Aagaard, K., Foldvik, A., & Hillman, S. R. (1987). The West Spitsbergen current: Disposition and water mass transformation. *Journal of Geophysical Research: Oceans*, *92*(C4), 3778–3784. https://doi.org/10.1029/JC092iC04p03778

Altin-Ballero, D. Z., Habura, A., & Goldstein, S. T. (2013). *Psammophaga sapela* n. sp., a new monothalamous foraminiferan from coastal Georgia, U.S.A.: Fine structure, gametogenesis, and phylogenetic placement. *Journal of Foraminiferal Research*, *43*(2), 113–126. https://doi.org/10.2113/gsjfr.43.2.113

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Ambrose, W. G., Jr., Carroll, M. L., Greenacre, M., Thorrold, S. R., & McMahon, K. W. (2006). Variation in *Serripes groenlandicus* (Bivalvia) growth in a Norwegian high-Arctic fjord: evidence for local- and large-scale climatic forcing. *Global Change Biology*, *12*(9), 1595–1607. https://doi.org/10.1111/j.1365-2486.2006.01181.x

Armbrecht, L. H., Coolen, M. J. L., Lejzerowicz, F., George, S. C., Negandhi, K., Suzuki, Y., Young, J., Foster, N. R., Armand, L. K., Cooper, A., Ostrowski, M., Focardi, A., Stat, M., Moreau, J. W., & Weyrich, L. S. (2019). Ancient DNA from marine sediments: Precautions and considerations for seafloor coring, sample handling and data generation. *Earth-Science Reviews*, *196*, 102887. https://doi.org/10.1016/j.earscirev.2019.102887

Backman, J., Fornaciari, E., & Rio, D. (2009). Biochronology and paleoceanography of late Pleistocene and Holocene calcareous nannofossil abundances across the Arctic Basin. *Marine Micropaleontology*, *72*(1), 86–98. https://doi.org/10.1016/j.marmicro.2009.04.001

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), 20351. https://doi.org/10.1038/s41598-020-77179-8

Barton, A. D., Irwin, A. J., Finkel, Z. V., & Stock, C. A. (2016). Anthropogenic climate change drives shift and shuffle in North Atlantic phytoplankton communities. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(11), 2964–2969. https://doi.org/10.1073/pnas.1519080113

Berge, J., Johnsen, G., Nilsen, F., Gulliksen, B., & Slagstad, D. (2005). Ocean temperature oscillations enable reappearance of blue mussels *Mytilus edulis* in Svalbard after a 1000 year absence. *Marine Ecology Progress Series*, *303*, 167–175. https://doi.org/10.3354/meps303167

Beszczynska-Möller, A., Fahrbach, E., Schauer, U., & Hansen, E. (2012). Variability in Atlantic water temperature and transport at the entrance to the Arctic Ocean, 1997–2010. *ICES Journal of Marine Science*, *69*(5), 852–863. https://doi.org/10.1093/icesjms/fss056

Blindheim, J., & Østerhus, S. (2005). The Nordic seas, main oceanographic features. *The Nordic seas: An integrated perspective*, *158*, 11–37. https://doi.org/10.1029/158GM03

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

Bourke, R. H., Weigel, A. M., & Paquette, R. G. (1988). The westward turning branch of the West Spitsbergen current. *Journal of Geophysical Research: Oceans*, *93*(C11), 14065–14077. https://doi.org/10.1029/JC093iC11p14065

Brandt, M. I., Pradillon, F., Trouche, B., Henry, N., Liautard-Haag, C., Cambon-Bonavita, M. A., Cueff-Gauchard, V., Wincker, P., Belser, C., Poulain, J., Arnaud-Haond, S., & Zeppilli, D. (2021). Evaluating sediment and water sampling methods for the estimation of deep-sea biodiversity using environmental DNA. *Scientific Reports*, *11*(1), 7856. https://doi.org/10.1038/s41598-021-86396-8

Brannock, P. M., & Halanych, K. M. (2015). Meiofaunal community analysis by high-throughput sequencing: Comparison of extraction, quality filtering, and clustering methods. *Marine Genomics*, *23*, 67–75. https://doi.org/10.1016/j.margen.2015.05.007

Burki, F., Kudryavtsev, A., Matz, M. V., Aglyamova, G. V., Bulman, S., Fiers, M., Keeling, P. J., & Pawlowski, J. (2010). Evolution of Rhizaria: new insights from phylogenomic analysis of uncultivated protists. *BMC Evolutionary Biology*, *10*(1), 377. https://doi.org/10.1186/1471-2148-10-377

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

Chronopoulou, P. M., Salonen, I., Bird, C., Reichart, G. J., & Koho, K. A. (2019). Metabarcoding insights into the trophic behavior and identity of intertidal benthic foraminifera. *Frontiers in Microbiology*, *10*, 1169. https://doi.org/10.3389/fmicb.2019.01169

Cordier, T., Barrenechea, I., Lejzerowicz, F., Reo, E., & Pawlowski, J. (2019). 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*, 1–18. https://doi.org/10.1016/j.pocean.2019.102175

Cottier, F., Tverberg, V., Inall, M., Svendsen, H., Nilsen, F., & Griffiths, C. (2005). Water mass modification in an Arctic fjord through cross-shelf exchange: The seasonal hydrography of Kongsfjorden, Svalbard. *Journal of Geophysical Research: Oceans*, *110*(C12). https://doi.org/10.1029/2004JC002757

Csapó, H. K., Grabowski, M., & Węsławski, J. M. (2021). Coming home—Boreal ecosystem claims Atlantic sector of the Arctic. *Science of the Total Environment*, *771*, 144817. https://doi.org/10.1016/j.scitotenv.2020.144817

Dahlke, S., Hughes, N. E., Wagner, P. M., Gerland, S., Wawrzyniak, T., Ivanov, B., & Maturilli, M. (2020). The observed recent surface air temperature development across Svalbard and concurring footprints in local sea ice cover. *International Journal of Climatology*, *40*(12), 5246–5265. https://doi.org/10.1002/joc.6517

Dai, A., Luo, D., Song, M., & Liu, J. (2019). Arctic amplification is caused by sea-ice loss under increasing CO_2. *Nature Communications*, *10*(1), 121. https://doi.org/10.1038/s41467-018-07954-9

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I., Carmichael, M., Poulain, J., Romac, S., Colin, S., Aury, J. M., Bittner, L., Chaffron, S., Dunthorn, M., Engelen, S., … Karsenti, E. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science*, *348*(6237), 1261605. https://doi.org/10.1126/science.1261605

Descamps, S., Aars, J., Fuglei, E., Kovacs, K. M., Lydersen, C., Pavlova, O., Pedersen, Å. Ø., Ravolainen, V., & Strøm, H. (2017). Climate change impacts on wildlife in a high Arctic archipelago—Svalbard, Norway. *Global Change Biology*, *23*(2), 490–502. https://doi.org/10.1111/gcb.13381

Dong, S. S., Lei, Y. L., Li, T. G., & Jian, Z. M. (2019). Responses of benthic foraminifera to changes of temperature and salinity: Results from a laboratory culture experiment. *Science China-Earth Sciences*, *62*(2), 459–472. https://doi.org/10.1007/s11430-017-9269-3

Dorst, S., & Schönfeld, J. (2013). Diversity of benthic foraminifera on the shelf and slope of the NE Atlantic: Analysis of datasets. *Journal of Foraminiferal Research*, *43*(3), 238–254. https://doi.org/10.2113/gsjfr.43.3.238

Dowdeswell, J. A., & Bamber, J. L. (1995). On the glaciology of Edgeøya and Barentsøya, Svalbard. *Polar Research*, *14*(2), 105–122. https://doi.org/10.3402/polar.v14i2.6658

Dowdeswell, J. A., Drewry, D. J., Cooper, A. P. R., Gorman, M. R., Liestøl, O., & Orheim, O. (1986). Digital mapping of the Nordaustlandet ice caps from airborne geophysical investigations. *Annals of Glaciology*, *8*, 51–58. https://doi.org/10.3189/S0260305500001130

Dufresne, Y., Lejzerowicz, F., Perret-Gentil, L. A., Pawlowski, J., & Cordier, T. (2019). SLIM: A flexible web application for the reproducible processing of environmental DNA metabarcoding data. *BMC Bioinformatics*, *20*(1), 88. https://doi.org/10.1186/s12859-019-2663-2

Dusa, A. (2018). *venn: Draw venn diagrams*. R package version 1.11. https://cran.r-project.org/package=venn

Elbrecht, V., Peinert, B., & Leese, F. (2017). Sorting things out: Assessing effects of unequal specimen biomass on DNA metabarcoding. *Ecology and Evolution*, *7*(17), 6918–6926. https://doi.org/10.1002/ece3.3192

Esling, P., Lejzerowicz, F., & Pawlowski, J. (2015). Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Research*, *43*(5), 2513–2524. https://doi.org/10.1093/nar/gkv107

Fossheim, M., Primicerio, R., Johannesen, E., Ingvaldsen, R. B., Aschan, M. M., & Dolgov, A. V. (2015). Recent warming leads to a rapid borealization of fish communities in the Arctic. *Nature Climate Change*, *5*(7), 673–677. https://doi.org/10.1038/nclimate2647

Frainer, A., Primicerio, R., Kortsch, S., Aune, M., Dolgov, A. V., Fossheim, M., & Aschan, M. M. (2017). Climate-driven changes in functional biogeography of Arctic marine fish communities. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(46), 12202–12207. https://doi.org/10.1073/pnas.1706080114

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

Froslev, T. G., Kjoller, R., Bruun, H. H., Ejrnaes, R., Brunbjerg, A. K., Pietroni, C., & Hansen, A. J. (2017). Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. *Nature Communications*, *8*(1), 1188. https://doi.org/10.1038/s41467-017-01312-x

Gooday, A. J., Anikeeva, O. V., & Pawlowski, J. (2011). New genera and species of monothalamous foraminifera from balaclava and Kazach'ya bays (Crimean peninsula, Black Sea). *Marine Biodiversity*, *41*(4), 481–494. https://doi.org/10.1007/s12526-010-0075-7

Gooday, A. J., Bowser, S. S., Cedhagen, T., Cornelius, N., Hald, M., Korsun, S., & Pawlowski, J. (2005). Monothalamous foraminiferans and gromiids (Protista) from western Svalbard: A preliminary survey published in collaboration with the University of Bergen and the Institute of Marine Research, Norway, and the Marine Biological Laboratory, university of Copenhagen, Denmark. *Marine Biology Research*, *1*(4), 290–312. https://doi.org/10.1080/17451000510019150

Gooday, A. J., & Jorissen, F. J. (2012). Benthic foraminiferal biogeography: Controls on global distribution patterns in deep-water settings. *Annual Review of Marine Science*, *4*, 237–262. https://doi.org/10.1146/annurev-marine-120709-142737

Grabowski, M., Jabłońska, A., Weydmann-Zwolicka, A., Gantsevich, M., Strelkov, P., Skazina, M., & Węsławski, J. M. (2019). Contrasting molecular diversity and demography patterns in two intertidal amphipod crustaceans reflect Atlantification of high Arctic. *Marine Biology*, *166*(12), 155. https://doi.org/10.1007/s00227-019-3603-4

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., De Vargas, C., Decelle, J., Del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H., Lara, E., Le Bescot, N., Logares, R., … Christen, R. (2013). The protist ribosomal reference database (PR2): A catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, *41*, D597–D604. https://doi.org/10.1093/nar/gks1160

Hald, M., & Korsun, S. (1997). Distribution of modern benthic foraminifera from fjords of Svalbard, European Arctic. *Journal of Foraminiferal Research*, *27*, 101–122. https://doi.org/10.2113/gsjfr.27.2.101

He, X., Sutherland, T. F., & Abbott, C. L. (2020). Improved efficiency in eDNA metabarcoding of benthic metazoans by sieving sediments prior to DNA extraction. *Environmental DNA*, *3*, 726. https://doi.org/10.1002/edn3.172

Holman, L. E., de Bruyn, M., Creer, S., Carvalho, G., Robidart, J., & Rius, M. (2021). Animals, protists and bacteria share marine biogeographic patterns. *Nature Ecology & Evolution*, *5*(6), 738–746. https://doi.org/10.1038/s41559-021-01439-7

Holzmann, M., Gooday, A. J., Siemensma, F., & Pawlowski, J. (2021). Review: Freshwater and soil foraminifera—A story of long-forgotten relatives. *Journal of Foraminiferal Research*, *51*(4), 318–331. https://doi.org/10.2113/gsjfr.51.4.318

Holzmann, M., & Pawlowski, J. (2017). An updated classification of rotaliid foraminifera based on ribosomal DNA phylogeny. *Marine Micropaleontology*, *132*, 18–34. https://doi.org/10.1016/j.marmicro.2017.04.002

Hop, H., Assmy, P., Wold, A., Sundfjord, A., Daase, M., Duarte, P., Kwasniewski, S., Gluchowska, M., Wiktor, J. M., Tatarek, A., Wiktor, J., Jr., Kristiansen, S., Fransson, A., Chierici, M., & Vihtakari, M. (2019). Pelagic ecosystem characteristics across the Atlantic water boundary current from Rijpfjorden, Svalbard, to the Arctic Ocean during summer (2010–2014). *Frontiers in Marine Science*, *6*, 181. https://doi.org/10.3389/fmars.2019.00181

Hsieh, T. C., Ma, K. H., & Chao, A. (2016). iNEXT: An R package for rarefaction and extrapolation of species diversity (Hill numbers). *Methods in Ecology and Evolution*, *7*(12), 1451–1456. https://doi.org/10.1111/2041-210X.12613

Kawahata, H., Fujita, K., Iguchi, A., Inoue, M., Iwasaki, S., Kuroyanagi, A., Maeda, A., Manaka, T., Moriya, K., Takagi, H., Toyofuku, T., Yoshimura, T., & Suzuki, A. (2019). Perspective on the response of marine calcifiers to global warming and ocean acidification—Behavior of corals and foraminifera in a high CO_2 world "hot house". *Progress in Earth and Planetary Science*, *6*(1), 5–37. https://doi.org/10.1186/s40645-018-0239-9

Knies, J., Brookes, S., & Schubert, C. J. (2007). Re-assessing the nitrogen signal in continental margin sediments: New insights from the high northern latitudes. *Earth and Planetary Science Letters*, *253*(3–4), 471–484. https://doi.org/10.1016/j.epsl.2006.11.008

Kolde, R. (2019). *pheatmap: Pretty Heatmap*. R package version 1.0.12. https://cran.r-project.org/package=pheatmap

Korsun, S., & Hald, M. (1998). Modern benthic foraminifera off Novaya Zemlya Tidewater glaciers, Russian Arctic. *Arctic and Alpine Research*, *30*(1), 61–77. https://doi.org/10.2307/1551746

Kortsch, S., Primicerio, R., Fossheim, M., Dolgov, A. V., & Aschan, M. (2015). Climate change alters the structure of arctic marine food webs due to poleward shifts of boreal generalists. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1814), 20151546. https://doi.org/10.1098/rspb.2015.1546

Kowalewski, W., Rudowski, S., & Zalewski, S. M. (1990). Seismoacoustic studies within Wijdefjorden, Spitsbergen. *Polish Polar Research*, *11*(3–4), 287–300.

Kujawa, A., Łącka, M., Szymańska, N., Telesiński, M., & Zajączkowski, M. (2021). Could Norwegian fjords serve as an analogue for the future of the Svalbard fjords? State and fate of high latitude fjords in the face of progressive "atlantification". *Polar Biology*, *44*, 2217–2233. https://doi.org/10.1007/s00300-021-02951-z

Langer, M. R., Weinmann, A. E., Lotters, S., Bernhard, J. M., & Rodder, D. (2013). Climate-driven range extension of *Amphistegina* (Protista, Foraminiferida): Models of current and predicted future ranges. *PLoS ONE*, *8*(2), e54443. https://doi.org/10.1371/journal.pone.0054443

Le Cao, K. A., Rossouw, D., Robert-Granie, C., & Besse, P. (2008). A sparse PLS for variable selection when integrating omics data. *Statistical Applications in Genetics and Molecular Biology*, *7*(1), Article 35. https://doi.org/10.2202/1544-6115.1390

Lecroq, B., Lejzerowicz, F., Bachar, D., Christen, R., Esling, P., Baerlocher, L., Østerås, 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 of the United States of America*, *108*(32), 13177–13182. https://doi.org/10.1073/pnas.1018426108

Lejzerowicz, F., Esling, P., & Pawlowski, J. (2014). Patchiness of deep-sea benthic foraminifera across the Southern Ocean: Insights from high-throughput DNA sequencing. *Deep Sea Research Part II: Topical Studies in Oceanography*, *108*, 17–26. https://doi.org/10.1016/j.dsr2.2014.07.018

Lejzerowicz, F., Gooday, A. J., Barrenechea Angeles, I., Cordier, T., Morard, R., Apothéloz-Perret-Gentil, L., Lins, L., Menot, L., Brandt, A., Levin, L. A., Arbizu, P. M., Smith, C. R., & 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), 671033. https://doi.org/10.3389/fmars.2021.671033

Leray, M., & Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(7), 2076–2081. https://doi.org/10.1073/pnas.1424997112

Leu, E., Søreide, J. E., Hessen, D. O., Falk-Petersen, S., & Berge, J. (2011). Consequences of changing sea-ice cover for primary and secondary producers in the European Arctic shelf seas: Timing, quantity, and quality. *Progress in Oceanography*, *90*(1), 18–32. https://doi.org/10.1016/j.pocean.2011.02.004

Loeng, H. (1991). Features of the physical oceanographic conditions of the Barents Sea. *Polar Research*, *10*(1), 5–18. https://doi.org/10.1111/j.1751-8369.1991.tb00630.x

Majewski, W., Pawłowski, J., & Zajączkowski, M. (2005). Monothalamous foraminifera from West Spitsbergen fjords, Svalbard: A brief overview. *PolishPolar Research*, *26*(4), 269–285.

Majewski, W., Szczuciński, W., & Zajączkowski, M. (2009). Interactions of Arctic and Atlantic water-masses and associated environmental changes during the last millennium, Hornsund (SW Svalbard). *Boreas*, *38*(3), 529–544. https://doi.org/10.1111/j.1502-3885.2009.00091.x

Meilland, J., Howa, H., Hulot, V., Demangel, I., Salaün, J., & Garlan, T. (2020). Population dynamics of modern planktonic foraminifera in the western Barents Sea. *Biogeosciences*, *17*(6), 1437–1450. https://doi.org/10.5194/bg-17-1437-2020

Meslard, F., Bourrin, F., Many, G., & Kerhervé, P. (2018). Suspended particle dynamics and fluxes in an Arctic fjord (Kongsfjorden, Svalbard). *Estuarine, Coastal and Shelf Science*, *204*, 212–224. https://doi.org/10.1016/j.ecss.2018.02.020

Murray, J. W. (2006). *Ecology and applications of benthic foraminifera*. Cambridge University Press. https://doi.org/10.1017/CBO9780511535529

Murray, J. W. (2014). *Ecology and palaeoecology of benthic foraminifera*. Routledge. https://doi.org/10.4324/9781315846101

Neukermans, G., Oziel, L., & Babin, M. (2018). Increased intrusion of warming Atlantic water leads to rapid expansion of temperate phytoplankton in the Arctic. *Global Change Biology*, *24*(6), 2545–2553. https://doi.org/10.1111/gcb.14075

Nilsen, F., Cottier, F., Skogseth, R., & Mattsson, S. (2008). Fjord–shelf exchanges controlled by ice and brine production: The interannual variation of Atlantic water in Isfjorden, Svalbard. *Continental Shelf Research*, *28*(14), 1838–1853. https://doi.org/10.1016/j.csr.2008.04.015

Nilsen, F., Skogseth, R., Vaardal-Lunde, J., & Inall, M. (2016). A simple shelf circulation model: Intrusion of Atlantic water on the West Spitsbergen shelf. *Journal of Physical Oceanography*, *46*(4), 1209–1230. https://doi.org/10.1175/jpo-d-15-0058.1

Ofstad, S., Zamelczyk, K., Kimoto, K., Chierici, M., Fransson, A., & Rasmussen, T. L. (2021). Shell density of planktonic foraminifera and pteropod species Limacina helicina in the Barents Sea: Relation to ontogeny and water chemistry. *PLoS ONE*, *16*(4), e0249178. https://doi.org/10.1371/journal.pone.0249178

Oksanen, J. A. I. (2007). *Vegan: Community ecology package*. R Package Version 1.8-5. https://cran.r-project.org/package=vegan

Onarheim, I. H., Smedsrud, L. H., Ingvaldsen, R. B., & Nilsen, F. (2014). Loss of sea ice during winter north of Svalbard. *Tellus A: Dynamic Meteorology and Oceanography*, *66*(1), 23933. https://doi.org/10.3402/tellusa.v66.23933

Paulson, J. N., Stine, O. C., Bravo, H. C., & Pop, M. (2013). Differential abundance analysis for microbial marker-gene surveys. *Nature Methods*, *10*(12), 1200–1202. https://doi.org/10.1038/nmeth.2658

Pavlova, O., Gerland, S., & Hop, H. (2019). Changes in sea-ice extent and thickness in Kongsfjorden, Svalbard (2003–2016). In H. Hop & C. Wiencke (Eds.), *The ecosystem of Kongsfjorden, Svalbard* (pp. 105–136). Springer International Publishing. https://doi.org/10.1007/978-3-319-46425-1_4

Pawlowska, J., Lejzerowicz, F., Esling, P., Szczucinski, W., Zajaczkowski, M., & Pawlowski, J. (2014). Ancient DNA sheds new light on the Svalbard foraminiferal fossil record of the last millennium. *Geobiology*, *12*(4), 277–288. https://doi.org/10.1111/gbi.12087

Pawlowska, J., Wollenburg, J. E., Zajaczkowski, M., & Pawlowski, J. (2020). Planktonic foraminifera genomic variations reflect paleoceanographic changes in the Arctic: Evidence from sedimentary ancient DNA. *Scientific Reports*, *10*(1), 15102. https://doi.org/10.1038/s41598-020-72146-9

Pawłowska, J., Zajączkowski, M., Łącka, 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*(7), 1459–1472. https://doi.org/10.5194/cp-12-1459-2016

Pawlowski, J., Bonin, A., Boyer, F., Cordier, T., & Taberlet, P. (2021). Environmental DNA for biomonitoring. *Molecular Ecology*, *30*(13), 2931–2936. https://doi.org/10.1111/mec.16023

Pawlowski, J., Esling, P., Lejzerowicz, F., Cedhagen, T., & Wilding, T. A. (2014). Environmental monitoring through protist next-generation sequencing metabarcoding: Assessing the impact of fish farming on benthic foraminifera communities. *Molecular Ecology Resources*, *14*(6), 1129–1140. https://doi.org/10.1111/1755-0998.12261

Pawlowski, J., Esling, P., Lejzerowicz, F., Cordier, T., Visco, J. A., Martins, C. I. M., Kvalvik, A., Staven, K., & Cedhagen, T. (2016). Benthic monitoring of salmon farms in Norway using foraminiferal metabarcoding. *Aquaculture Environment Interactions*, *8*, 371–386. https://doi.org/10.3354/aei00182

Pawlowski, J., Holzmann, M., & Tyszka, J. (2013). New supraordinal classification of foraminifera: Molecules meet morphology. *Marine Micropaleontology*, *100*, 1–10. https://doi.org/10.1016/j.marmicro.2013.04.002

Pawlowski, J., & Lecroq, B. (2010). Short rDNA barcodes for species identification in foraminifera. *The Journal of Eukaryotic Microbiology*, *57*(2), 197–205. https://doi.org/10.1111/j.1550-7408.2009.00468.x

Polyak, L., Best, K. M., Crawford, K. A., Council, E. A., & St-Onge, G. (2013). Quaternary history of sea ice in the western Arctic Ocean based on foraminifera. *Quaternary Science Reviews*, *79*, 145–156. https://doi.org/10.1016/j.quascirev.2012.12.018

Polyakov, I. V., Pnyushkov, A. V., Alkire, M. B., Ashik, I. M., Baumann, T. M., Carmack, E. C., Goszczko, I., Guthrie, J., Ivanov, V. V., Kanzow, T., Krishfield, R., Kwok, R., Sundfjord, A., Morison, J., Rember, R., & Yulin, A. (2017). Greater role for Atlantic inflows on sea-ice loss in the Eurasian Basin of the Arctic Ocean. *Science*, *356*(6335), 285–291. https://doi.org/10.1126/science.aai8204

Prazeres, M., Roberts, T. E., & Pandolfi, J. M. (2017). Variation in sensitivity of large benthic foraminifera to the combined effects of ocean warming and local impacts. *Scientific Reports*, *7*(1), 45227. https://doi.org/10.1038/srep45227

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Renaud, P. E., Wallhead, P., Kotta, J., Włodarska-Kowalczuk, M., Bellerby, R. G. J., Rätsep, M., Slagstad, D., & Kukliński, P. (2019). Arctic sensitivity? Suitable habitat for benthic taxa is surprisingly robust to climate change. *Frontiers in Marine Science*, *6*, 538. https://doi.org/10.3389/fmars.2019.00538

Rohart, F., Gautier, B., Singh, A., & Le Cao, K. A. (2017). mixOmics: An R package for omics feature selection and multiple data integration. *PLoS Computational Biology*, *13*(11), e1005752. https://doi.org/10.1371/journal.pcbi.1005752

Sabbatini, A., Morigi, C., Negri, A., & Gooday, A. (2007). Distribution and biodiversity of stained monothalamous foraminifera from Tempelfjord, Svalbard. *Journal of Foraminiferal Research*, *37*, 93–106. https://doi.org/10.2113/gsjfr.37.2.93

Saloranta, T. M., & Haugan, P. M. (2004). Northward cooling and freshening of the warm core of the West Spitsbergen current. *Polar Research*, *23*(1), 79–88. https://doi.org/10.3402/polar.v23i1.6268

Schoenle, A., Hohlfeld, M., Hermanns, K., Mahe, F., de Vargas, C., Nitsche, F., & Arndt, H. (2021). High and specific diversity of protists in the deep-sea basins dominated by diplonemids, kinetoplastids, ciliates and foraminiferans. *Communications Biology*, *4*(1), 501. https://doi.org/10.1038/s42003-021-02012-5

Schönfeld, J., Alve, E., Geslin, E., Jorissen, F., Korsun, S., & Spezzaferri, S. (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

Serreze, M. C., & Barry, R. G. (2011). Processes and impacts of Arctic amplification: A research synthesis. *Global and Planetary Change*, *77*(1), 85–96. https://doi.org/10.1016/j.gloplacha.2011.03.004

Sinniger, F., Pawlowski, J., Harii, S., Gooday, A. J., Yamamoto, H., Chevaldonné, P., Carvalho, G., & Creer, S. (2016). 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

Slagstad, D., Ellingsen, I. H., & Wassmann, P. (2011). Evaluating primary and secondary production in an Arctic Ocean void of summer sea ice: An experimental simulation approach. *Progress in Oceanography*, *90*(1), 117–131. https://doi.org/10.1016/j.pocean.2011.02.009

Smith, C. W., & Goldstein, S. T. (2019). The effects of selected heavy metal elements (arsenic, cadmium, nickel, zinc) on experimentally grown foraminiferal assemblages from Sapelo Island, Georgia and little duck key, Florida, U.S.A. *Journal of Foraminiferal Research*, *49*(3), 303–317. https://doi.org/10.2113/gsjfr.49.3.303

Spielhagen, R. F., Werner, K., Sørensen, S. A., Zamelczyk, K., Kandiano, E., Budeus, G., Husum, K., Marchitto, T. M., & Hald, M. (2011). Enhanced modern heat transfer to the Arctic by warm Atlantic water. *Science*, *331*(6016), 450–453. https://doi.org/10.1126/science.1197397

Sternal, B., Szczuciński, W., Forwick, M., Zajączkowski, M., Lorenc, S., & Przytarska, J. (2014). Postglacial variability in near-bottom current speed on the continental shelf off south-West Spitsbergen. *Journal of Quaternary Science*, *29*, 767–777. https://doi.org/10.1002/jqs.2748

Vihtakari, M., Welcker, J., Moe, B., Chastel, O., Tartu, S., Hop, H., Bech, C., Descamps, S., & Gabrielsen, G. W. (2018). Black-legged kittiwakes as messengers of Atlantification in the Arctic. *Scientific Reports*, *8*(1), 1178. https://doi.org/10.1038/s41598-017-19118-8

Wallace, M. I., Cottier, F. R., Berge, J., Tarling, G. A., Griffiths, C., & Brierley, A. S. (2010). Comparison of zooplankton vertical migration in an ice-free and a seasonally ice-covered Arctic fjord: An insight into the influence of sea ice cover on zooplankton behavior. *Limnology and Oceanography*, *55*(2), 831–845. https://doi.org/10.4319/lo.2010.55.2.0831

Walseng, B., Jensen, T., Dimante-Deimantovica, I., Christoffersen, K. S., Chertoprud, M., Chertoprud, E., Novichkova, A., & Hessen, D. O. (2018). Freshwater diversity in Svalbard: Providing baseline data for ecosystems in change. *Polar Biology*, *41*(10), 1995–2005. https://doi.org/10.1007/s00300-018-2340-3

Weinmann, A. E., & Goldstein, S. T. (2016). Changing structure of benthic foraminiferal communities: Implications from experimentally grown assemblages from coastal Georgia and Florida, USA. *Marine Ecology*, *37*(4), 891–906. https://doi.org/10.1111/maec.12368

Weydmann-Zwolicka, A., Pratnicka, P., Lacka, M., Majaneva, S., Cottier, F., & Berge, J. (2021). Zooplankton and sediment fluxes in two contrasting fjords reveal Atlantification of the Arctic. *Science of the Total Environment*, *773*, 145599. https://doi.org/10.1016/j.scitotenv.2021.145599

Wittmann, A. C., & Pörtner, H.-O. (2013). Sensitivities of extant animal taxa to ocean acidification. *Nature Climate Change*, *3*(11), 995–1001. https://doi.org/10.1038/nclimate1982

Włodarska-Kowalczuk, M., Pawłowska, J., & Zajączkowski, M. (2013). Do foraminifera mirror diversity and distribution patterns of macrobenthic fauna in an Arctic glacial fjord? *Marine Micropaleontology*, *103*, 30–39. https://doi.org/10.1016/j.marmicro.2013.07.002

Wollenburg, J. E., Mackensen, A., & Kuhnt, W. (2007). Benthic foraminiferal biodiversity response to a changing Arctic palaeoclimate in the last 24.000 years. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *255*(3), 195–222. https://doi.org/10.1016/j.palaeo.2007.05.007

Zajączkowski, M., Nygård, H., Hegseth, E. N., & Berge, J. (2010). Vertical flux of particulate matter in an Arctic fjord: The case of lack of the sea-ice cover in Adventfjorden 2006–2007. *Polar Biology*, *33*(2), 223–239. https://doi.org/10.1007/s00300-009-0699-x

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Nguyen, N.-L., Pawłowska, J., Angeles, I. B., Zajaczkowski, M., & Pawłowski, J. (2023). Metabarcoding reveals high diversity of benthic foraminifera linked to water masses circulation at coastal Svalbard. *Geobiology*, *21*, 133–150. https://doi.org/10.1111/gbi.12530

3. Research Article III

Assigning the unassigned: a signature-based classification of rDNA metabarcodes reveals new deep-sea diversity. Inès Barrenechea Angeles, Ngoc-Loi Nguyen, Mattia Greco, Koh Siang Tan, and Jan Pawlowski. Published in *PLoS One*, 19(2), e0298440 (2024).

PLOS ONE

RESEARCH ARTICLE

Assigning the unassigned: A signature-based classification of rDNA metabarcodes reveals new deep-sea diversity

Inès Barrenechea Angeles[1,2,3]*, **Ngoc-Loi Nguyen**[4], **Mattia Greco**[4,5], **Koh Siang Tan**[6], **Jan Pawlowski**[4,7]*

1 Department of Earth Sciences, University of Geneva, Geneva, Switzerland, **2** Department of Genetics and Evolution, University of Geneva, Geneva, Switzerland, **3** Department of Geosciences, UiT-The Arctic University of Norway, Tromsø, Norway, **4** Institute of Oceanology, Polish Academy of Sciences, Sopot, Poland, **5** Institute of Marine Sciences, Spanish National Research Council, Barcelona, Spain, **6** Tropical Marine Science Institute, National University of Singapore, Singapore, Singapore, **7** ID-Gene Ecodiagnostics Ltd., Plan-les-Ouates, Switzerland

* ines.a.angeles@uit.no (IBA); janpawlowski@iopan.pl (JP)

Citation: Barrenechea Angeles I, Nguyen N-L, Greco M, Tan KS, Pawlowski J (2024) Assigning the unassigned: A signature-based classification of rDNA metabarcodes reveals new deep-sea diversity. PLoS ONE 19(2): e0298440. https://doi.org/10.1371/journal.pone.0298440

Editor: Fabrizio Frontalini, Universita degli Studi di Urbino Carlo Bo, ITALY

Received: August 2, 2023

Accepted: January 23, 2024

Published: February 29, 2024

Data Availability Statement: The DNA sequences from the additional samples collected during RESOURCE Cruise 01 (OMS license area) are deposited in the Sequence Read Archive (SRA) public database under the accession PRJNA899048. The DNA sequences from the additional samples collected during RESOURCE Cruise 01 (OMS license area) have been deposited in the National Center for Biotechnology Information Short Reads Archive (NCBI SRA) database under Bioproject accession number

Abstract

Environmental DNA metabarcoding reveals a vast genetic diversity of marine eukaryotes. 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 eukaryotic microbiota, whose hidden diversity is largely unexplored. Here, we tackle this issue by using unique DNA signatures to classify unknown metabarcodes assigned to deep-sea foraminifera. We analyzed metabarcoding data obtained from 311 deep-sea sediment samples collected in the Clarion-Clipperton Fracture Zone, an area of potential polymetallic nodule exploitation in the Eastern Pacific Ocean. Using the signatures designed in the 37F hypervariable region of the 18S rRNA gene, we were able to classify 802 unassigned metabarcodes into 61 novel lineages, which have been placed in 27 phylogenetic clades. The comparison of new lineages with other foraminiferal datasets shows that most novel lineages are widely distributed in the deep sea. Five lineages are also present in the shallow-water datasets; however, phylogenetic analysis of these lineages separates deep-sea and shallow-water metabarcodes except in one case. While the signature-based classification does not solve the problem of gaps in reference databases, this taxonomy-free approach provides insight into the distribution and ecology of deep-sea species represented by unassigned metabarcodes, which could be useful in future applications of metabarcoding for environmental monitoring.

Introduction

The past decade has seen environmental DNA (*e*DNA) 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 [1,2]. The problem concerns generally

PRJNA899048. Additional DNA sequences analyzed in this study can be found in the NCBI SRA following Bioproject accession numbers: PRJEB44134 (CCFZ), PRJEB44134 and PRJNA554310 (deep-sea), PRJNA723313, PRJNA813562, PRJEB29469, RJNA897836 and PRJNA768352 (shallow water).

Funding: IBA and JP were supported by the Swiss National Science Foundation grant no. 31003A_179125. IBA was supported by the Norwegian Petroleum Directorate, and the project AKMA (project n° 287869). NLN, MG, and JP were supported by the Norwegian Financial Mechanism for 2014-2021, project no. 2019/34/H/ST10/00682. The authors are also grateful to Ocean Mineral Singapore, Keppel Corporation-National University of Singapore Corporate Laboratory, and the National Research Foundation of Singapore for partially supporting this work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. There was no additional external funding received for this study.

Competing interests: The authors have declared that no competing interests exist.

the surveys of prokaryotic communities, which are dominated by unknown taxa, also called "microbial dark matter" [3] especially in extreme environments [4,5], polar [6], deep-sea [5] and hydrothermal vents [4,5]. However, the unassigned sequences also prevail among protist and meiofaunal communities [7–9]. These sequences are commonly lumped into an assemblage of unassigned or unknown metabarcodes. Lacking taxonomic information, these sequences cannot be included in biodiversity or biogeography assessments, except as "unknown". Different strategies have been proposed to overcome this problem. A recent study showed that taxonomic assignment approaches based on sequence similarity and composition outperformed more complex phylogenetic and probabilistic methods [10], the accuracy of taxonomic assignment based on the percentage similarity of short metabarcodes is generally low such as in 18S rRNA gene [11] and TrnL P6 loop [12]. Alternatively, a network approach was proposed to characterize unknown species and elucidate their relationships [5].

Here, we tackle this issue by classifying the unassigned metabarcodes into novel lineages using an ultra-short nucleotide sequence that can distinguish one lineage from another, called DNA signatures or signatures character. In general, a DNA signature has focused on single genes (e.g., 16S/18S rDNA gene, mitochondrial COI gene), and could be selected by using sequence alignments in the conserved gene regions. DNA signatures of closely related species or close phylogenetic lineages are expected to be more similar to one another. The signature-based approach to detect and identify microorganisms has been proposed already earlier [13,14], yet its use in current prokaryotic taxonomy is relatively limited since number of sequenced genomes has continued to increase dramatically [15]. This approach is useful in the case of eukaryotes, whose genomic reconstruction is limited compared to prokaryotes [16]. Among eukaryotes, distinctive molecular patterns are generally used to resolve the taxonomy of closely related species [17] or to analyze geographic patterns [18]. A recent study demonstrated the usefulness of DNA signatures to facilitate the taxonomic identification of ciliated protists [18]. Therefore, the nuclear and mitochondrial genes of a microbial eukaryote may bear the signatures needed to integrate both phylogenetic and ecological information.

In our study, we applied the DNA signatures to classify deep-sea unassigned benthic foraminiferal sequences. The recent global metabarcoding analysis showed that the diversity of deep-sea benthic eukaryotes is huge and by far exceeds that of species living in surface waters [19]. However, due to the remoteness of deep-sea habitat, our knowledge about its biodiversity is limited and the majority of eukaryotic metabarcodes obtained from deep-sea sediment DNA remain unassigned. This concerns not only microbial eukaryotes but also metazoan meiofauna, which abound in deep-sea sediments [20]. Unsurprisingly, the metabarcoding surveys reporting the composition of deep-sea microbial and meiofaunal communities are dominated by unassigned taxa.

We focused on foraminifera, which comprises a significant fraction of deep-sea benthic diversity [21–23] and represents more than 50% of the total biomass in Clarion Clipperton Fracture Zone [21], Antarctic Peninsula [24], hypoxic and anoxic environments [21,25]. It has been suggested that at least some deep-sea foraminiferal species are distributed globally based on ribosomal DNA barcodes of isolated specimens [26,27]. This has been confirmed by studies reporting several cosmopolitan foraminiferal amplicon sequence variants (ASVs) or operational taxonomic units (OTUs) in deep-sea metabarcoding data [28,29]. Yet, most of these globally distributed metabarcodes could not be assigned or have only been assigned at higher levels (class, order). According to some studies, the proportion of unassigned sequences in the deep-sea foraminiferal datasets exceeds 50% [28,29].

The material for this study comes from the Eastern Pacific's Clarion-Clipperton Fracture Zone (CCFZ), an area of potential polymetallic nodule exploitation. The biological community of CCFZ was targeted by several biodiversity surveys [30–32]. The foraminiferal assemblage of

CCFZ was shown to be dominated by monothalamous taxa, most of which remained morphologically and genetically unidentified [29,33,34]. We performed a metabarcoding analysis on sediments across different areas of CCFZ 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.

Material and methods

Sediment sample collection

The sampling was carried out within the contract area assigned to Ocean Mineral Singapore by the International Seabed Authority. In this study, 36 samples were collected in 2020 using 1mx1m box cores during RESOURCE Cruise 01 (OMS license area). At each station, three replicates were taken with a 50 ml sterile syringe with the end cut off. The syringe was inserted into the sediment in order to collect at least 5 cm. As we were interested only in the surface sediments, we pushed the sediment lengthwise into a plastic cup where the last centimeters were discarded. Only the first 1–2 centimeters were placed into a tube with 10 ml of LifeGuard Preservation solution (Qiagen, Germany). Samples were frozen on board, shipped frozen to the University of Geneva, and stored at -20˚C until their extraction.

Sediment DNA extraction, amplification, and sequencing

The sediment samples were extracted using the manufacturer's guidelines of the DNeasy® PowerMax® Soil Kit (Qiagen, Germany). To target foraminifera eDNA, the 37F hypervariable region of the nuclear 18S rRNA gene (68–196 bp), was PCR amplified using specific primers [27]. To allow multiplexing of samples in one library, the forward s14F1 5′-AAGGGCACCACAAGAACGC-3′ and reverse s15 5'- CCACCTATCACAYAATCATG-3' primers were tagged with unique 8 nucleotides at the 5' end [35]. Three PCR replicates were amplified and pooled for each sample before being quantified using high-resolution capillary electrophoresis (QIAxcel System, Qiagen, Germany). The PCR products were pooled in equimolar concentration. Dimers and short amplicons (< 100 bp) were then excluded from the pool using the High Pure PCR Product Purification Kit (Roche), as the shortest amplicon including the primers and tags is 123 bp. The library was prepared using TruSeq® DNA PCR-Free Library Preparation Kit (Illumina, USA), and its concentration was quantified using Kapa Library Quantification Kit for Illumina Platforms (KAPA Biosystems, USA). Finally, the library was sequenced with a MiSeq instrument using paired-end sequencing for 300 cycles with a v.2 kit.

Bioinformatics analysis

We combined the obtained sequence with the published ones from other sites from CCFZ, and other deep-sea foraminifera datasets obtained from samples between -4000 and -9000 meters of water depth from the North Atlantic, Mid Atlantic, South Atlantic, Southern Ocean, and Northwest Pacific [29,36] (see S1 Table), and available in ENA under the following accession number PRJEB44134, PRJNA554310, and PRJNA899048. We also added the shallow water foraminifera datasets from the Tyrrhenian Sea [37], Adriatic Sea [38–41] and around Svalbard [42] (see S1 Fig), available under the following accession numbers: PRJNA723313, PRJNA897836, PRJNA813562, PRJEB29469, and PRJNA768352. Some of those datasets were

obtained using primers s14F1- s17 [43] and therefore targeting two hypervariable regions of 18S (37F and 41F), including the studied region.

The raw datasets were processed using the SLIM software [44]. First, they were demultiplexed and the primers were removed using the module *demultiplexer*. The paired fastq files from all datasets were combined and processed together (quality filtering, denoising, merging, and chimera removal on sequences) using the module DADA2 [45] implemented in SLIM. The DADA workflow was set to default parameters, without length truncation and pseudo-pooling as the pooling parameter for the inference of ASV. Then, we clustered the obtained Amplicon Sequencing Variants (ASVs) at 97% similarity into OTUs and continued with a LULU curation [46] as recommended in [47]. This curation removes erroneous clusters coming from intra-individual variability or errors during PCR or sequencing. The clustering at 97% was done using the DECIPHER R package and the curation with the LULU R package with the default parameters.

To retain only foraminifera sequences obtained with s14F1 -s15 primers, we identified conservative motifs across all foraminiferal species in the region 37 flanking the hypervariable region, i.e., before the beginning of 37F and at the end. Using *grep* command in R or bash we removed sequences not having "GACAG", adjacent to the foraminiferal-specific hypervariable region 37F [27] and at the end of the 37 conservative region "TAGTCCTTT" and "TAGTCCCTT". In some species, we noticed the presence of substitution (T > C) therefore we used these two patterns. The remaining sequences were then filtered by their size and abundance, we retained sequences with > 70 bp and > 100 reads.

Some shallow-water sequences were obtained using the primer pairs s14F1- s17 covering the 37f and 41f variable regions. For them, we retained sequences only if they contained "GACAG" in the 37 region and "GGTGGT" in the 38 conserved region.

We used three probabilistic approaches to assign the sequences taxonomically and to identify the unassigned sequences: VSEARCH [48] at 95% similarity, IDTAXA [49] at 60% of confidence, and BLAST+ [50] at 95% similarity and 100–99% of coverage. We used our local database of benthic foraminifera including selected sequences from GenBank and the planktonic foraminifera ribosomal reference database—PFR2 [51]. The resulting 4602 reference sequences cover Globothalamea, Tubothalamea, and the paraphyletic groups of monothalamids. The monothalamids comprised well-defined clades (e.g., Clade A [52]), the ENFOR (ENvironmental FORaminifera) groups consisting of environmental clades from previous metabarcoding studies obtained through cloning and Sanger sequencing (e.g., ENFOR1 [53]), and/or poorly defined clades (e.g., Monothalamids X or undetermined Monothalamids), comprising mainly the so-called squatter species [54,55].

DNA signature identification

We prepared a subset of the CCFZ dataset including 2245 OTUs that could not be assigned by VSEARCH as well as those that VSEARCH assigned to ENFOR or Monothalamids 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, as we assumed that the conserved regions should contain similar sequences across all foraminiferal OTUs. Sequences having similar molecular signatures at the beginning or the end of the 37F region were regrouped into lineages. The signatures 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 signature recognized a group already present in the database, except if they were assigned to an environmental clade or a Monothalamids X. After these restrictive filters, only 693 OTUs were used to define the unique

signature, corresponding to each lineage. The remained lineages were 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 lineages sharing most of the characters, thus obtaining sub-lineages. 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. The relative abundance was calculated using the *make_relative* function within the *funrar* package while the map was drawn using the following libraries *rnaturalearth*, *rnaturalearthdata*, and *ggspatial*.

Phylogenetic analysis

Phylogenetic tree specific to new lineages was constructed, covering the entire monothalamids to assign taxonomy and resolve undescribed clades. A total of 693 OTUs of new lineages and 388 reference sequences from well-described monothalamids were included in the phylogenetic tree construction. As an outgroup, we used two sequences from non-foraminiferal rhizarians (*Cercomonas longicauda* and *Gromia oviformis*). We aligned our sequences using the E-INS-i iterative refinement method in MAFFT v.7 [56]. Trees were built using the IQ-TREE maximum likelihood method [57,58]. Ultra-fast bootstrapping [59] was used to generate branch support values with 1000 bootstrap replicates. Phylogenetic tree visualization and annotation were done using the R package *ggtree* v.1.12.7 [60]. Default alignment parameters were used to align and generate a phylogenetic tree. Based on the phylogenetic tree, the 43 lineages were grouped into 27 higher-ranking groups (e.g., CCZ1). This provides an appropriate degree of phylogenetic specificity for each signature (S4 Table).

Results

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 2382 OTUs, the other deep-sea areas dataset 48,559,807 reads corresponding to 4148 OTUs and the shallow water dataset comprised 26,349,529 reads and 3745 OTUs. Details of the number of reads retained at each step and for each basin are detailed in S1 Table.

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 groups of foraminifera: globothalamids, tubothalamids, and monothalamids. However, less than 50% of OTUs were assigned. VSEARCH assigned the greatest fractions of sequences (46.2%), followed by BLAST (24.1%) and IDTAXA (10.2%). The monothalamids, including environmental sequences (ENFOR) and Monothalamids X, were the most abundant groups of foraminifera (S1 Fig, more details in S3 Table). Globothalamids and tubothalamids were the minority in the three assignments. According to the VSEARCH assignment, globothalamids and tubothalamids made up roughly 4.9% (561,586) of reads, monothalamids, including ENFOR and Monothalamids X, represented 41.28% (5,554,157) of reads, while unassigned OTUs accounted for 53.73% (21,466,294 reads).

From sequence alignment of 693 unassigned OTUs, a total of 61 DNA signatures were identified corresponding to 30 lineages and 31 sub-lineages (S4 Table). The length of signatures varied between 12 and 53 nucleotides. Most of the signatures (51) were located at the beginning of the 37F variable region, comprising the six conservative nucleotides "GACAGG"

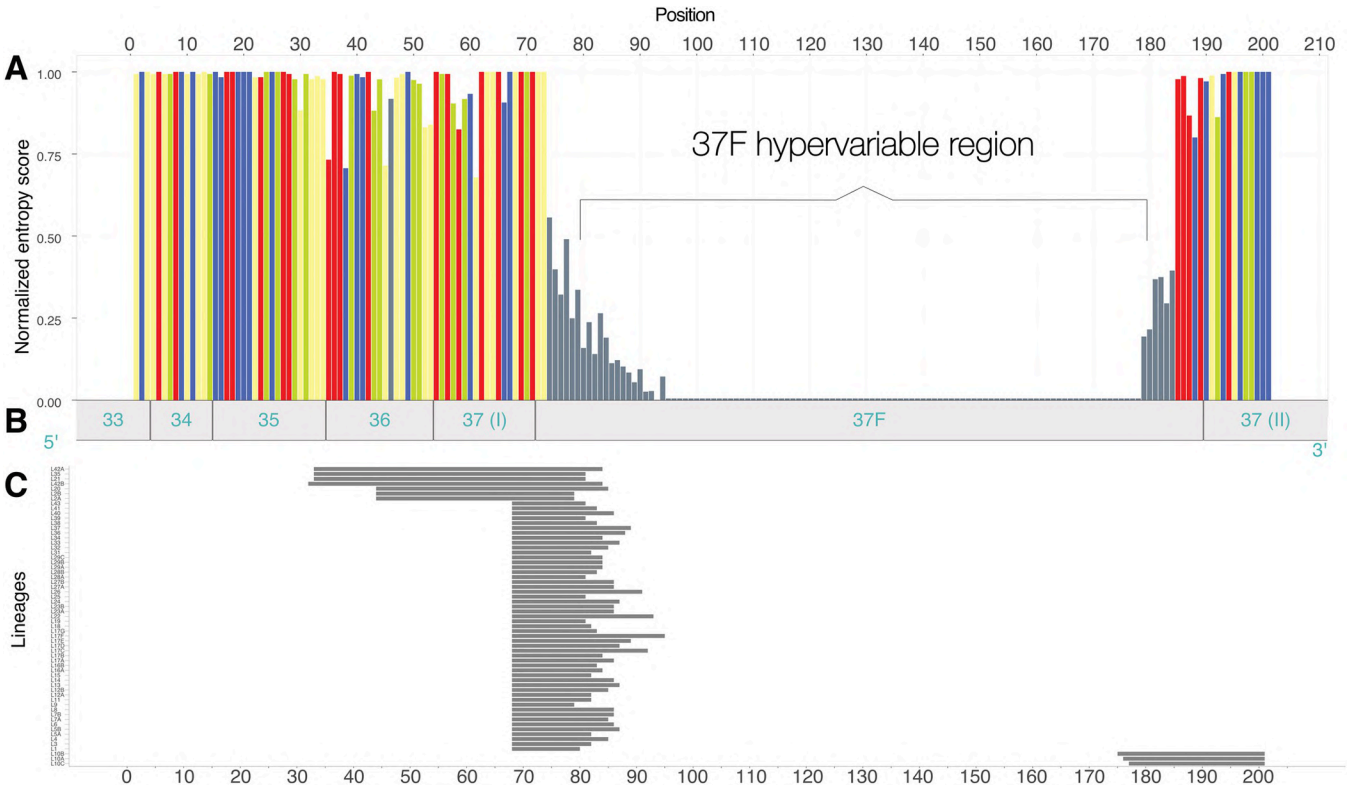


Fig 1. Positions of signatures in the foraminiferal 18S rRNA gene. (A) entropy plot and (B) foraminiferal regions from 33 to 37 after [27], (C) position and length of signatures.

https://doi.org/10.1371/journal.pone.0298440.g001

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 sub-lineages (Fig 1).

By searching for the signatures without a mismatch (i.e., with 100% similarity), we could identify 109 additional OTUs in the CCZ dataset (see S5 Table). In total, 802 OTUs (corresponding to 34% of the total number of OTUs and 62% of the total number of reads) were assigned to novel lineages. The signature approach allowed to reduce the number of unassigned OTUs 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 a large proportion of OTUs assigned to novel lineages among the monothalamids (34%) and the undetermined monothalamids (Monothalamids X, 54%). One of the novel lineages (L21) was assigned to both monothalamids and tubothalamids, but this requires confirmation by single-cell sequencing. No signature was found among globothalamid sequences.

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 S2 Fig. 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).

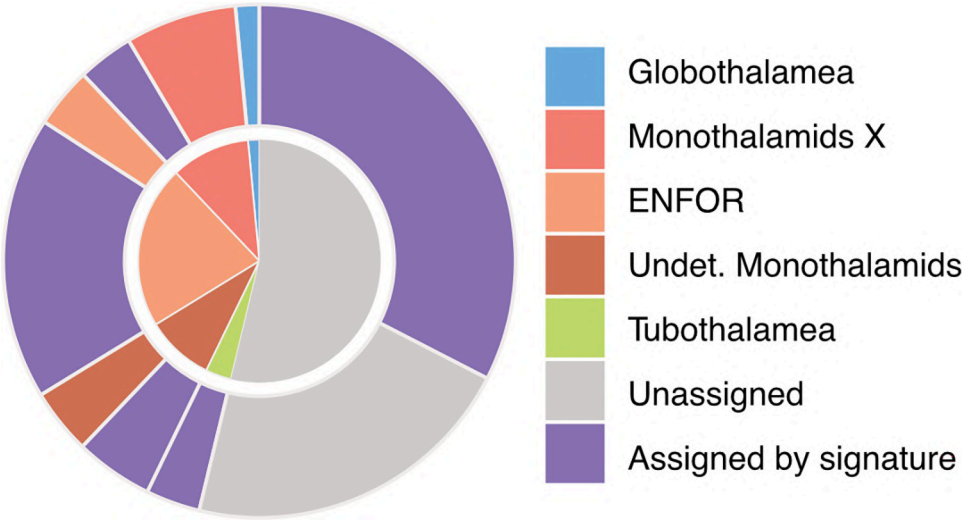


Fig 2. The pie chart shows the proportion of foraminifera groups before and after being assigned by signatures. The inner pie chart represents the result of VSEARCH assignments, and the outer ring represents a combined assignment including VSEARCH and signature-based approach (in purple). The foraminiferal groups are assigned by signatures including the new lineages in unassigned, monothalamids, ENFOR, and other (undetermined) monothalamids.

https://doi.org/10.1371/journal.pone.0298440.g002

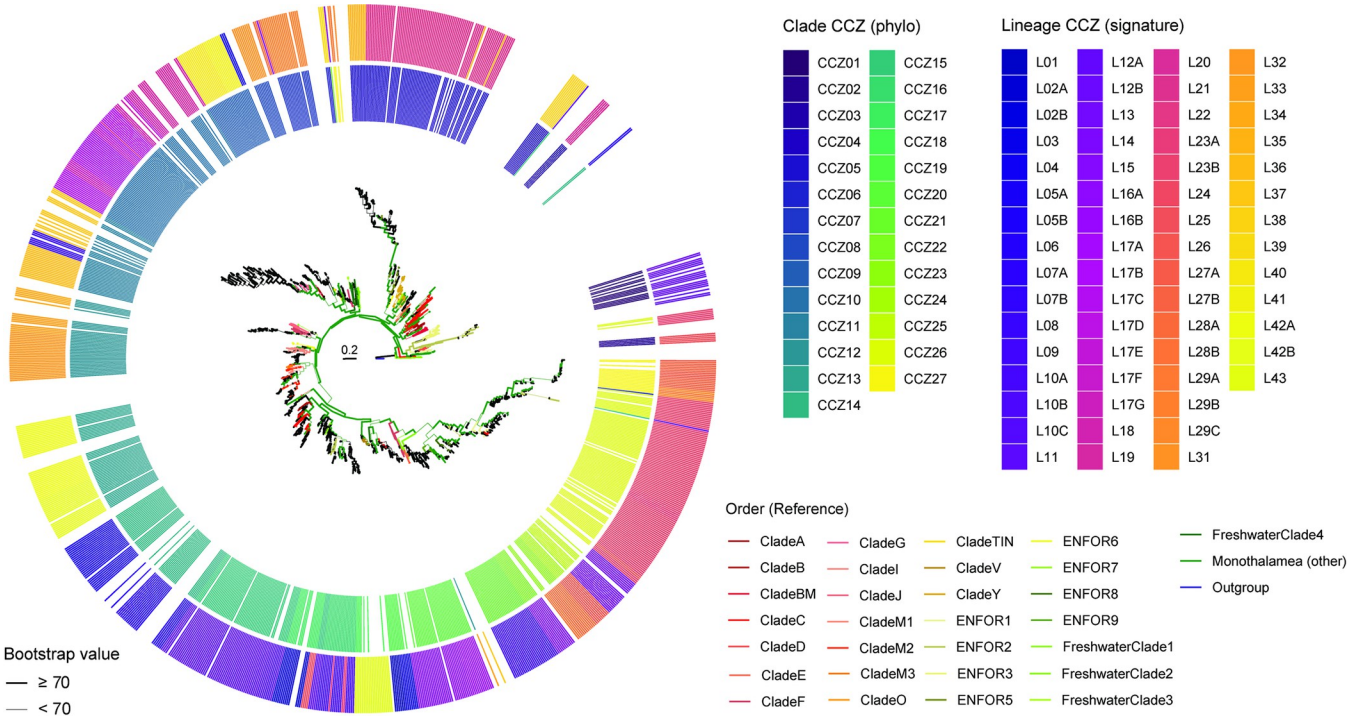


Fig 3. Phylogenetic diversity and novelty of foraminiferal OTUs identified by signatures. Phylogenetic analysis of selected OTUs representing new lineages and reference sequences of monothalamids from Clade A to Clade Y and some freshwater clades. Tree branches are colored at the Order level. All sequences 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 $\geq$ 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).

https://doi.org/10.1371/journal.pone.0298440.g003

Phylogenetic analysis indicated that the signatures of the assigned lineages were more similar to each other than to those of distant ones (Figs 3 and S2). 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.

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 UK-1 areas shared all the lineages whereas in BGR he lineage

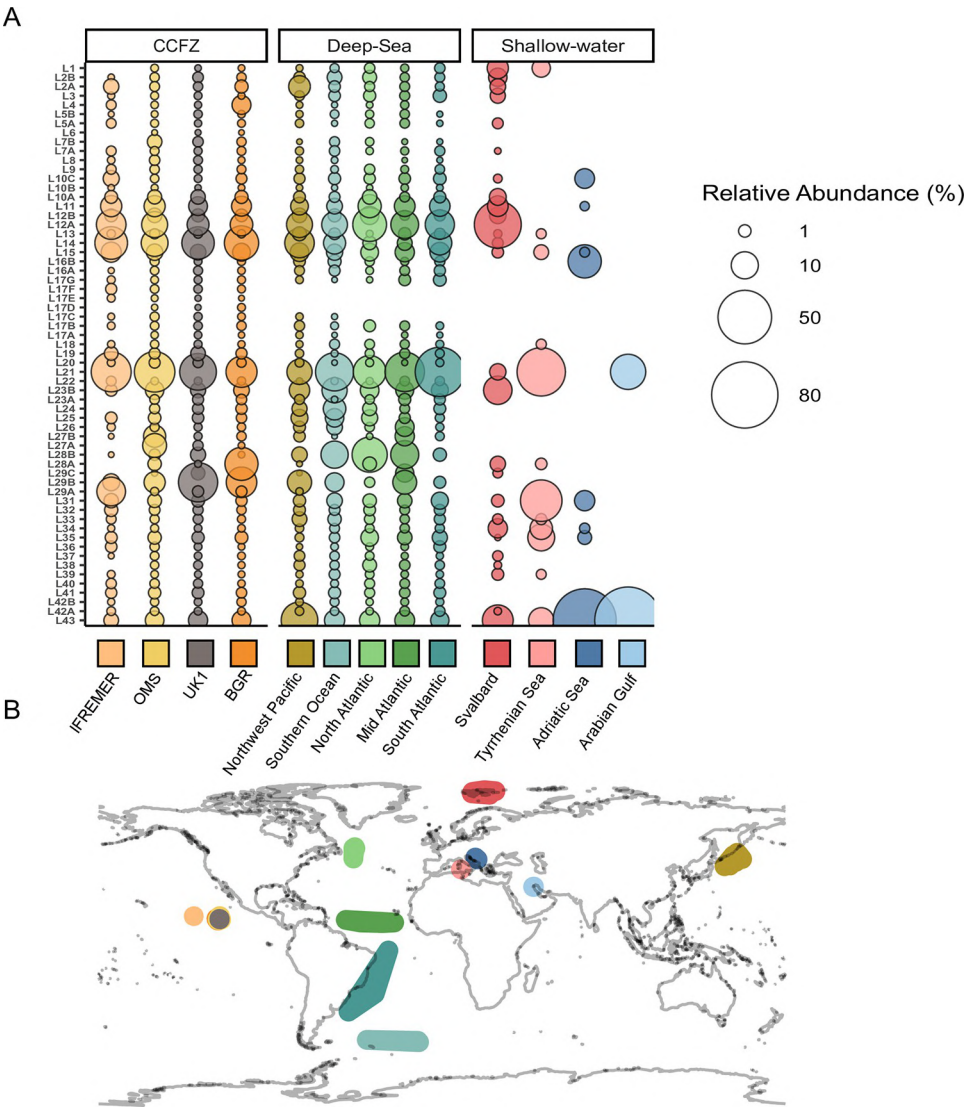


Fig 4. Bubble chart showing the proportions of 61 lineages present and their distribution in the studied regions. (A) The bubble sizes show the relative abundance of lineages per area. The bigger the bubble, the more abundant the lineage is in each area. (B) A map showing CCFZ, other deep-sea areas, and shallow water sites (<200 m depth).

https://doi.org/10.1371/journal.pone.0298440.g004

L29Cwas absent. The IFREMER area, located in the westernmost part of CFFZ, has the lowest number of lineages (49) shared with the eastern part of CCFZ sites. Comparing CCFZ to deep-sea sites, 85% of lineages were the most deep-sea regions. Only five lineages were endemic to CCFZ (absent in all other areas): L6, L17D, E, F, and 27A. 56 lineages occurred in the North-west Pacific, 53 in the Southern Ocean, and 50 in the three regions of the Atlantic Ocean. Lineages 28A and 29C only appeared in the North Atlantic and the mid-Atlantic, respectively. L4 was present in the North and mid-Atlantic and L8 and L17A were found in the mid and south of Atlantic.

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 the most abundant and had in common with the other three cosmopolitan lineages a very short signature.

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 the OTUs 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 in terms of OTUs' number (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.

Discussion

Despite the advances introduced by metabarcoding, taxonomically unassigned sequences remain an issue for researchers interested in biological diversity assessment and ecology. As shown by our study, about half of the deep-sea metabarcodes could not be assigned. This proportion is even higher if we also consider as unassigned the metabarcodes that were classified only at higher levels (phylum or class). Indeed, the assignment at such a high taxonomic 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 to 54% when using bioinformatics tools (VSEARCH, IDTAXA, and BLAST) to 80% using the signature approach (Fig 2). In total, 61 new foraminiferal lineages have been defined based on DNA signatures. As expected, most of these lineages belong to monothalamids, a paraphyletic assemblage of early-evolved single-chambered foraminifera [52], which are generally overlooked in conventional foraminiferal surveys [61]. Our study confirms the importance of this group in the deep-sea environment [21] and provides a general scaffold for its classification.

Besides this taxonomic aspect, our approach can also contribute to a better understanding of the ecology and geographic distribution of deep-sea foraminifera. This information could be lost if the unassigned foraminiferal sequences are lumped together. Some authors analyzed metabarcoding data at the level of ASV or OTU, for example, in the study of patchiness of deep-sea foraminifera [62] or their distribution along the depth gradient [36] or even in coastal

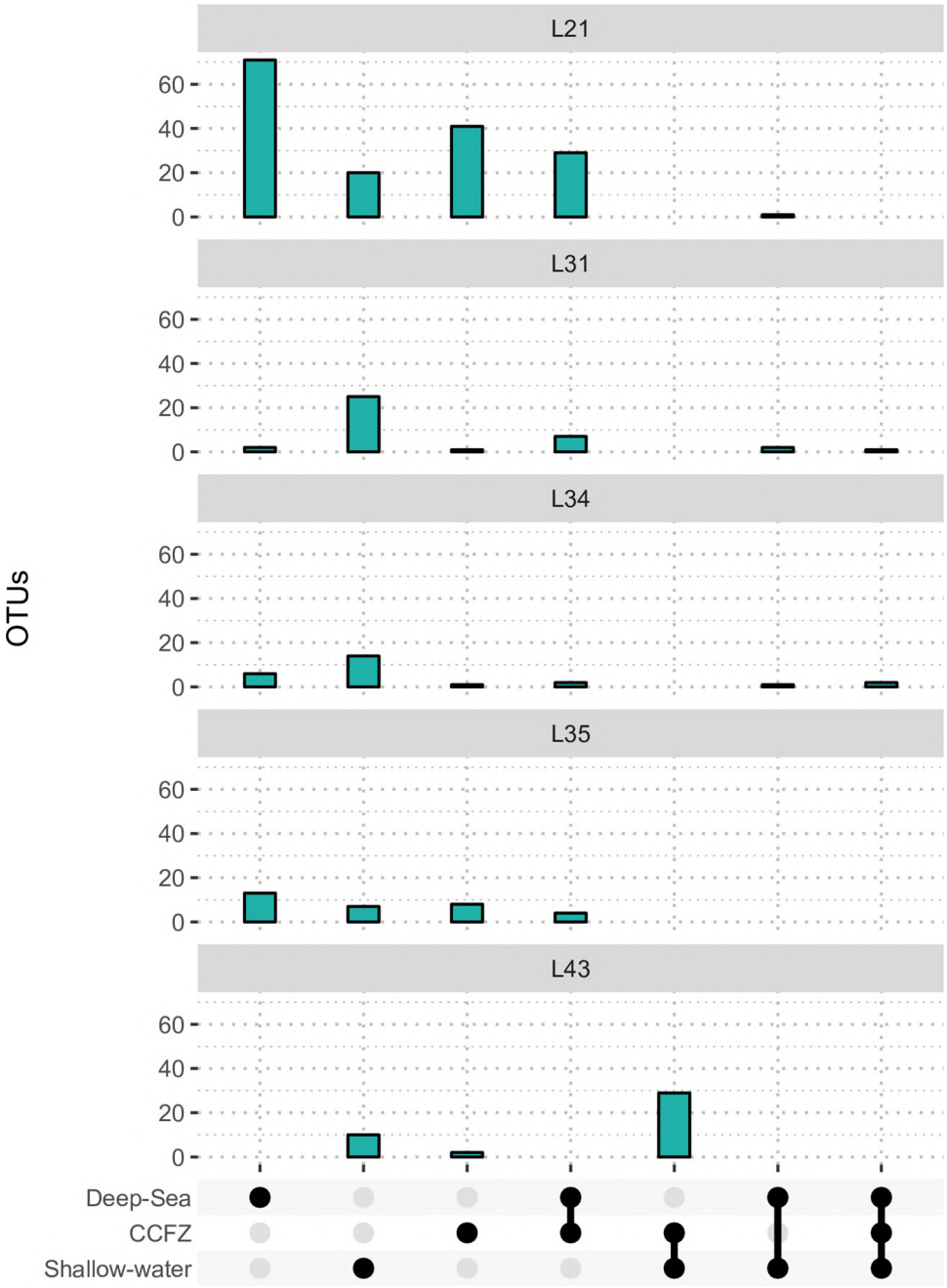


Fig 5. UpSet chart showing the five most abundant lineages. It comprises OTUs shared between CCFZ, deep-sea, and shallow-water samples. All duplicate OTUs were removed and the number of OTUs is a conservative estimate per habitat.

https://doi.org/10.1371/journal.pone.0298440.g005

biomonitoring [63]. 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 [64]. By classifying ASV/OTUs at higher taxonomic levels our approach facilitates their correlation with environmental variables.

The advantages 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 [28,65]. 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 demonstrates that the numerous foraminiferal lineages are specifically deep-sea. It is well documented that the giant monothalamous foraminifera belonging to Xenophyophorea occur exclusively on abyssal plains [66]. Nevertheless, according to our study, the number of foraminiferal lineages adapted to the deep sea might be much higher 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 the beginning of 37F variable region [35,67]. Thus, the longer signatures might better define the lower taxonomic level and can reduce the risk of misidentification as in the case of L21, a short signature whose assignment and placement were within monothalamids and tubothalamids species. 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 [68].

Furthermore, not all foraminiferal species can be distinguished in this region, 37f, as shown by [69] where it was not possible to discriminate Cibicidoides species. This can be solved by increasing the number of metabarcodes obtained through single-cell analysis. Once a comprehensive database of foraminiferal metabarcodes is established, one would have to develop a further signature-based approach to make it useful for taxonomical and ecological studies.

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 bp) 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 single-cell polymorphism analysis.

To conclude, we view our approach as an inclusive tool that allows expanding the information inferred from metabarcoding data to the currently 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 and/or poorly explored habitats, such as the deep sea. It can help in the case of DNA-based environmental monitoring that targets particular groups of bioindicators or in paleo-metabarcoding reconstructions of past biodiversity. Its efficiency will certainly increase if the metabarcoding data are combined with single-cell high-throughput barcoding, but this taxonomy-free approach can be viewed as a practical way to uncover hidden information present in hitherto unassigned metabarcoding data.

Supporting information

S1 Fig. 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 Monothalamids (Undet. Monothalamids) are coloured in shades of orange. More details in S3 Table.
(TIFF)

S2 Fig. Extended phylogenetic tree of CCFZ monothalamid sequences (ASVXXX) and monothalamids reference sequences (PAWXXX). The tree was constructed using the maximum likelihood method. The size of circles at nodes represents bootstrap support. The first column of CCFZ sequences is the name, the second the clade, and the third the lineage.
(PDF)

S1 Table. List of datasets. Features datasets from CCFZ, Deep-Sea and Shallow-water basins and their accession number.
(XLSX)

S2 Table. Filtering reads. Number of reads at each step of filtering per dataset and area.
(XLSX)

S3 Table. Number of reads per taxonomic method. Distribution of main foraminifera classes depending on the method of taxonomic assignment, only sequences having more than 100 reads were taken into account.
(XLSX)

S4 Table. List of lineages and clades. Details Lineages and clades including the signature (sequence), abundance and number of OTUs.
(XLSX)

S5 Table. Taxonomic assignment of sequences that could be assigned by the signatures. Other assignment methods such as VSEARCH, IDTAXA and BLAST are also displayed for comparison.
(XLSX)

Acknowledgments

The authors are thankful to Adrian Glover and Swee Cheng Lim for helping in collecting samples, to Oceanne Himmighofen and Maria Holzmann for providing DNA barcoding data from CCFZ foraminifera, and to Andrew Gooday for helpful discussion.

Author Contributions

Conceptualization: Inès Barrenechea Angeles, Ngoc-Loi Nguyen, Jan Pawlowski.

Data curation: Inès Barrenechea Angeles, Ngoc-Loi Nguyen, Mattia Greco, Jan Pawlowski.

Formal analysis: Inès Barrenechea Angeles, Ngoc-Loi Nguyen, Mattia Greco, Jan Pawlowski.

Funding acquisition: Koh Siang Tan, Jan Pawlowski.

Investigation: Inès Barrenechea Angeles, Ngoc-Loi Nguyen, Mattia Greco, Koh Siang Tan, Jan Pawlowski.

Methodology: Inès Barrenechea Angeles, Ngoc-Loi Nguyen, Jan Pawlowski.

Resources: Koh Siang Tan.

Software: Mattia Greco.

Supervision: Jan Pawlowski.

Validation: Inès Barrenechea Angeles.

Visualization: Inès Barrenechea Angeles, Ngoc-Loi Nguyen, Mattia Greco.

Writing – original draft: Inès Barrenechea Angeles, Ngoc-Loi Nguyen, Mattia Greco, Koh Siang Tan, Jan Pawlowski.

Writing – review & editing: Inès Barrenechea Angeles, Ngoc-Loi Nguyen, Mattia Greco, Koh Siang Tan, Jan Pawlowski.

References

1. Gold Z, Curd EE, Goodwin KD, Choi ES, Frable BW, Thompson AR, et al. Improving metabarcoding taxonomic assignment: A case study of fishes in a large marine ecosystem. Mol Ecol Resour. 2021; 21 (7):2546–64. https://doi.org/10.1111/1755-0998.13450 PMID: 34235858
2. Hestetun JT, Bye-Ingebrigtsen E, Nilsson RH, Glover AG, Johansen P-O, Dahlgren TG. Significant taxon sampling gaps in DNA databases limit the operational use of marine macrofauna metabarcoding. Mar Biodivers. 2020; 50(5):70. https://doi.org/10.1007/s12526-020-01093-5
3. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, et al. Insights into the phylogeny and coding potential of microbial dark matter. Nature. 2013; 499(7459):431–7. https://doi.org/10.1038/nature12352 PMID: 23851394
4. Schultz J, Modolon F, Peixoto RS, Rosado AS. Shedding light on the composition of extreme microbial dark matter: alternative approaches for culturing extremophiles. Front Microbiol. 2023; 14. https://doi.org/10.3389/fmicb.2023.1167718 PMID: 37333658
5. Zamkovaya T, Foster JS, de Crécy-Lagard V, Conesa A. A network approach to elucidate and prioritize microbial dark matter in microbial communities. ISME J. 2021; 15(1):228–44. https://doi.org/10.1038/s41396-020-00777-x PMID: 32963345
6. Bowman JS. Identification of microbial dark matter in Antarctic environments. Front Microbiol. 2018; 9. https://doi.org/10.3389/fmicb.2018.03165 PMID: 30619224
7. de Vargas C, Norris R, Zaninetti L, Gibb SW, Pawlowski J. Molecular evidence of cryptic speciation in planktonic foraminifers and their relation to oceanic provinces. Proc Natl Acad Sci U S A. 1999; 96 (6):2864–8. https://doi.org/10.1073/pnas.96.6.2864 PMID: 10077602
8. Forster D, Dunthorn M, Stoeck T, Mahé F. Comparison of three clustering approaches for detecting novel environmental microbial diversity. PeerJ. 2016; 4:e1692. https://doi.org/10.7717/peerj.1692 PMID: 26966652
9. Logares R, Audic S, Bass D, Bittner L, Boutte C, Christen R, et al. Patterns of rare and abundant marine microbial eukaryotes. Curr Biol. 2014; 24(8):813–21. https://doi.org/10.1016/j.cub.2014.02.050 PMID: 24704080
10. Hleap JS, Littlefair JE, Steinke D, Hebert PDN, Cristescu ME. Assessment of current taxonomic assignment strategies for metabarcoding eukaryotes. Mol Ecol Resour. 2021; 21(7):2190–203. https://doi.org/10.1111/1755-0998.13407 PMID: 33905615
11. Latz MAC, Grujcic V, Brugel S, Lycken J, John U, Karlson B, et al. Short- and long-read metabarcoding of the eukaryotic rRNA operon: Evaluation of primers and comparison to shotgun metagenomics sequencing. Mol Ecol Resour. 2022; 22(6):2304–18. https://doi.org/10.1111/1755-0998.13623 PMID: 35437888
12. Taberlet P, Coissac E, Pompanon F, Gielly L, Miquel C, Valentini A, et al. Power and limitations of the chloroplast trn L (UAA) intron for plant DNA barcoding. Nucleic Acids Res. 2007; 35(3):e14–e. https://doi.org/10.1093/nar/gkl938 PMID: 17169982
13. Albuquerque P, Mendes MV, Santos CL, Moradas-Ferreira P, Tavares F. DNA signature-based approaches for bacterial detection and identification. Sci Total Environ. 2009; 407(12):3641–51. https://doi.org/10.1016/j.scitotenv.2008.10.054 PMID: 19062077
14. Phillippy AM, Mason JA, Ayanbule K, Sommer DD, Taviani E, Huq A, et al. Comprehensive DNA signature discovery and validation. PLoS Comput Biol. 2007; 3(5):e98. https://doi.org/10.1371/journal.pcbi.0030098 PMID: 17511514

15. Hugenholtz P, Chuvochina M, Oren A, Parks DH, Soo RM. Prokaryotic taxonomy and nomenclature in the age of big sequence data. ISME J. 2021; 15(7):1879–92. https://doi.org/10.1038/s41396-021-00941-x PMID: 33824426

16. West PT, Probst AJ, Grigoriev IV, Thomas BC, Banfield JF. Genome-reconstruction for eukaryotes from complex natural microbial communities. Genome Res. 2018; 28(4):569–80. https://doi.org/10.1101/gr.228429.117 PMID: 29496730

17. Zielske S, Haase M. Molecular phylogeny and a modified approach of character-based barcoding refining the taxonomy of New Caledonian freshwater gastropods (Caenogastropoda, Truncatelloidea, Tateidae). Mol Phylogenet Evol. 2015; 89:171–81. https://doi.org/10.1016/j.ympev.2015.04.020 PMID: 25929789

18. Ganser MH, Santoferrara LF, Agatha S. Molecular signature characters complement taxonomic diagnoses: A bioinformatic approach exemplified by ciliated protists (Ciliophora, Oligotrichea). Mol Phylogenet Evol. 2022; 170:107433. https://doi.org/10.1016/j.ympev.2022.107433 PMID: 35131424

19. Cordier T, Angeles IB, Henry N, Lejzerowicz F, Berney C, Morard R, et al. Patterns of eukaryotic diversity from the surface to the deep-ocean sediment. Sci Adv. 2022; 8(5):eabj9309. https://doi.org/10.1126/sciadv.abj9309 PMID: 35119936

20. Sinniger F, Pawlowski J, Harii S, Gooday AJ, Yamamoto H, Chevaldonné P, et al. Worldwide analysis of sedimentary DNA reveals major gaps in taxonomic knowledge of deep-sea benthos. Front Mar Sci. 2016; 3. https://doi.org/10.3389/fmars.2016.00092

21. Gooday AJ. Deep-Sea Benthic Foraminifera. In: Cochran JK, Bokuniewicz HJ, Yager PL, editors. Encyclopedia of Ocean Sciences. Third Edition ed. Oxford: Academic Press; 2019. p. 684–705.

22. Gooday AJ. Benthic Foraminifera. In: Steele JH, editor. Encyclopedia of Ocean Sciences. Oxford: Academic Press; 2001. p. 274–86.

23. Gooday AJ, Schoenle A, Dolan JR, Arndt H. Protist diversity and function in the dark ocean–Challenging the paradigms of deep-sea ecology with special emphasis on foraminiferans and naked protists. Eur J Protistol. 2020; 75:125721. https://doi.org/10.1016/j.ejop.2020.125721 PMID: 32575029

24. Cornelius N, Gooday AJ. 'Live' (stained) deep-sea benthic foraminiferans in the western Weddell Sea: trends in abundance, diversity and taxonomic composition along a depth transect. Deep Sea Res Part II Top Stud Oceanogr. 2004; 51(14):1571–602. https://doi.org/10.1016/j.dsr2.2004.06.024

25. Levin LA, Amon DJ, Lily H. Challenges to the sustainability of deep-seabed mining. Nat Sustain. 2020; 3(10):784–94. https://doi.org/10.1038/s41893-020-0558-x

26. Pawlowski J, Lejzerowicz F, Esling P. Next-generation environmental diversity surveys of Foraminifera: Preparing the future. Biol Bull. 2014; 227(2):93–106. https://doi.org/10.1086/BBLv227n2p93 PMID: 25411369.

27. Pawlowski J, Lecroq B. Short rDNA barcodes for species identification in foraminifera. J Eukaryot Microbiol. 2010; 57(2):197–205. https://doi.org/10.1111/j.1550-7408.2009.00468.x PMID: 20113377

28. Lecroq B, Lejzerowicz F, Bachar D, Christen R, Esling P, Baerlocher L, et al. Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. Proc Natl Acad Sci U S A. 2011; 108(32):13177–82. https://doi.org/10.1073/pnas.1018426108 PMID: 21788523; PubMed Central PMCID: PMC3156150.

29. Lejzerowicz F, Gooday AJ, Barrenechea Angeles I, Cordier T, Morard R, Apothéloz-Perret-Gentil L, et al. Eukaryotic biodiversity and spatial patterns in the Clarion-Clipperton Zone and other abyssal regions: Insights from sediment DNA and RNA metabarcoding. Front Mar Sci. 2021; 8. https://doi.org/10.3389/fmars.2021.671033

30. Jones DOB, Simon-Lledó E, Amon DJ, Bett BJ, Caulle C, Clément L, et al. Environment, ecology, and potential effectiveness of an area protected from deep-sea mining (Clarion Clipperton Zone, abyssal Pacific). Prog Oceanogr. 2021; 197:102653. https://doi.org/10.1016/j.pocean.2021.102653

31. Rabone M, Wiethase JH, Simon-Lledó E, Emery AM, Jones DOB, Dahlgren TG, et al. How many metazoan species live in the world's largest mineral exploration region? Curr Biol. 2023; 33(12):2383–96.e5. https://doi.org/10.1016/j.cub.2023.04.052 PMID: 37236182

32. Washburn TW, Jones DOB, Wei C-L, Smith CR. Environmental Heterogeneity Throughout the Clarion-Clipperton Zone and the Potential Representativity of the APEI Network. Front Mar Sci. 2021; 8. https://doi.org/10.3389/fmars.2021.661685

33. Goineau A, Gooday AJ. Novel benthic foraminifera are abundant and diverse in an area of the abyssal equatorial Pacific licensed for polymetallic nodule exploration. Sci Rep. 2017; 7(1):45288. https://doi.org/10.1038/srep45288 PMID: 28382941

34. Gooday AJ, Lejzerowicz F, Goineau A, Holzmann M, Kamenskaya O, Kitazato H, et al. The biodiversity and distribution of abyssal benthic foraminifera and their possible ecological roles: A synthesis across the Clarion-Clipperton Zone. Front Mar Sci. 2021; 8. https://doi.org/10.3389/fmars.2021.634726

35. Esling P, Lejzerowicz F, Pawlowski J. Accurate multiplexing and filtering for high-throughput amplicon-sequencing. Nucleic Acids Res. 2015; 43(5):2513–24. https://doi.org/10.1093/nar/gkv107 PMID: 25690897

36. Cordier T, Barrenechea I, Lejzerowicz F, Reo E, Pawlowski J. Benthic foraminiferal DNA metabarcodes significantly vary along a gradient from abyssal to hadal depths and between each side of the Kuril-Kamchatka trench. Prog Oceanogr. 2019; 178:102175. https://doi.org/10.1016/j.pocean.2019.102175

37. Cavaliere M, Barrenechea Angeles I, Montresor M, Bucci C, Brocani L, Balassi E, et al. Assessing the ecological quality status of the highly polluted Bagnoli area (Tyrrhenian Sea, Italy) using foraminiferal eDNA metabarcoding. Sci Total Environ. 2021; 790:147871. https://doi.org/10.1016/j.scitotenv.2021.147871 PMID: 34098278

38. Al-Enezi E, Francescangeli F, Balassi E, Borderie S, Al-Hazeem S, Al-Salameen F, et al. Benthic foraminifera as proxies for the environmental quality assessment of the Kuwait Bay (Kuwait, Arabian Gulf): Morphological and metabarcoding approaches. Sci Total Environ. 2022; 833:155093. https://doi.org/10.1016/j.scitotenv.2022.155093 PMID: 35421459

39. Cordier T, Frontalini F, Cermakova K, Apothéloz-Perret-Gentil L, Treglia M, Scantamburlo E, et al. Multi-marker eDNA metabarcoding survey to assess the environmental impact of three offshore gas platforms in the North Adriatic Sea (Italy). Mar Environ Res. 2019; 146:24–34. https://doi.org/10.1016/j.marenvres.2018.12.009 PMID: 30890270

40. Frontalini F, Greco M, Di Bella L, Lejzerowicz F, Reo E, Caruso A, et al. Assessing the effect of mercury pollution on cultured benthic foraminifera community using morphological and eDNA metabarcoding approaches. Mar Pollut Bull. 2018; 129(2):512–24. https://doi.org/10.1016/j.marpolbul.2017.10.022 PMID: 29033170

41. Greco M, Lejzerowicz F, Reo E, Caruso A, Maccotta A, Coccioni R, et al. Environmental RNA outperforms eDNA metabarcoding in assessing impact of marine pollution: A chromium-spiked mesocosm test. Chemosphere. 2022; 298:134239. https://doi.org/10.1016/j.chemosphere.2022.134239 PMID: 35292278

42. Nguyen N-L, Pawłowska J, Angeles IB, Zajaczkowski M, Pawłowski J. Metabarcoding reveals high diversity of benthic foraminifera linked to water masses circulation at coastal Svalbard. Geobiology. 2023; 21(1):133–50. https://doi.org/10.1111/gbi.12530 PMID: 36259453

43. Pawlowski J. Introduction to the Molecular Systematics of Foraminifera. Micropaleontology. 2000; 46:1–12.

44. Dufresne Y, Lejzerowicz F, Perret-Gentil LA, Pawlowski J, Cordier T. SLIM: a flexible web application for the reproducible processing of environmental DNA metabarcoding data. BMC Bioinform. 2019; 20(1):88. https://doi.org/10.1186/s12859-019-2663-2 PMID: 30782112; PubMed Central PMCID: PMC6381720.

45. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016; 13(7):581–3. https://doi.org/10.1038/nmeth.3869 PMID: 27214047; PubMed Central PMCID: PMC4927377.

46. Frøslev TG, Kjøller R, Bruun HH, Ejrnæs R, Brunbjerg AK, Pietroni C, et al. Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates. Nat Commun. 2017; 8(1):1188. https://doi.org/10.1038/s41467-017-01312-x PMID: 29084957

47. Brandt MI, Trouche B, Quintric L, Günther B, Wincker P, Poulain J, et al. Bioinformatic pipelines combining denoising and clustering tools allow for more comprehensive prokaryotic and eukaryotic metabarcoding. Mol Ecol Resour. 2021; 21(6):1904–21. https://doi.org/10.1111/1755-0998.13398 PMID: 33835712

48. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016; 4:e2584. https://doi.org/10.7717/peerj.2584 PMID: 27781170

49. Murali A, Bhargava A, Wright ES. IDTAXA: a novel approach for accurate taxonomic classification of microbiome sequences. Microbiome. 2018; 6(1):140. https://doi.org/10.1186/s40168-018-0521-5 PMID: 30092815

50. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinform. 2009; 10(1):421. https://doi.org/10.1186/1471-2105-10-421 PMID: 20003500

51. Morard R, Darling KF, Mahé F, Audic S, Ujiié Y, Weiner AKM, et al. PFR2: a curated database of planktonic foraminifera 18S ribosomal DNA as a resource for studies of plankton ecology, biogeography and evolution. Mol Ecol Resour. 2015; 15(6):1472–85. https://doi.org/10.1111/1755-0998.12410 PMID: 25828689

52. Pawlowski J, Holzmann M, Tyszka J. New supraordinal classification of Foraminifera: Molecules meet morphology. Mar Micropaleontol. 2013; 100:1–10. https://doi.org/10.1016/j.marmicro.2013.04.002

53. Pawlowski J, Fontaine D, da Silva AA, Guiard J. Novel lineages of Southern Ocean deep-sea foraminifera revealed by environmental DNA sequencing. Deep Sea Res Part II Top Stud Oceanogr. 2011; 58 (19):1996–2003. https://doi.org/10.1016/j.dsr2.2011.01.009

54. Gooday AJ, Rothe N, Pearce RB. New and poorly known benthic foraminifera (Protista, Rhizaria) inhabiting the shells of planktonic foraminifera on the bathyal Mid-Atlantic Ridge. Mar Biol Res. 2013; 9(5–6):447–61. https://doi.org/10.1080/17451000.2012.750365

55. Moodley L. "Squatter" behaviour in soft-shelled foraminifera. Mar Micropaleontol. 1990; 16(1):149–53. https://doi.org/10.1016/0377-8398(90)90033-I

56. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2017; 20(4):1160–6. https://doi.org/10.1093/bib/bbx108 PMID: 28968734

57. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2014; 32(1):268–74. https://doi.org/10.1093/molbev/msu300 PMID: 25371430

58. Trifinopoulos J, Nguyen L-T, von Haeseler A, Minh BQ. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res. 2016; 44(W1):W232–W5. https://doi.org/10.1093/nar/gkw256 PMID: 27084950

59. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol. 2017; 35(2):518–22. https://doi.org/10.1093/molbev/msx281 PMID: 29077904

60. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017; 8(1):28–36. https://doi.org/10.1111/2041-210X.12628

61. Pawlowski J, Holzmann M, Berney C, Fahrni J, Gooday AJ, Cedhagen T, et al. The evolution of early Foraminifera. Proc Natl Acad Sci U S A. 2003; 100(20):11494–8. https://doi.org/10.1073/pnas.2035132100 PMID: 14504394

62. Lejzerowicz F, Esling P, Pawlowski J. Patchiness of deep-sea benthic Foraminifera across the Southern Ocean: Insights from high-throughput DNA sequencing. Deep Sea Res Part II Top Stud Oceanogr. 2014; 108:17–26. https://doi.org/10.1016/j.dsr2.2014.07.018

63. Singer D, Fouet MPA, Schweizer M, Mouret A, Quinchard S, Jorissen FJ. Unlocking foraminiferal genetic diversity on estuarine mudflats with eDNA metabarcoding. Sci Total Environ. 2023; 902:165983. https://doi.org/10.1016/j.scitotenv.2023.165983 PMID: 37543334

64. Weber AA-T, Pawlowski J. Wide Occurrence of SSU rDNA Intragenomic Polymorphism in Foraminifera and its Implications for Molecular Species Identification. Protist. 2014; 165(5):645–61. https://doi.org/10.1016/j.protis.2014.07.006 PMID: 25150612

65. Pawlowski J, Fahrni J, Lecroq B, Longet D, Cornelius N, Excoffier L, et al. Bipolar gene flow in deep-sea benthic foraminifera. Mol Ecol. 2007; 16(19):4089–96. https://doi.org/10.1111/j.1365-294X.2007.03465.x PMID: 17725572

66. Gooday AJ, Holzmann M, Caulle C, Goineau A, Kamenskaya O, Weber AAT, et al. Giant protists (xenophyophores, Foraminifera) are exceptionally diverse in parts of the abyssal eastern Pacific licensed for polymetallic nodule exploration. Biol Conserv. 2017; 207:106–16. https://doi.org/10.1016/j.biocon.2017.01.006

67. Pawlowski J, Holzmann M. A plea for DNA barcoding of Foraminifera. J Foraminiferal Res. 2014; 44 (1):62–7. https://doi.org/10.2113/gsjfr.44.1.62

68. Pawlowski J, Bolivar I, Fahrni JF, de Vargas C, Gouy M, Zaninetti L. Extreme differences in rates of molecular evolution of foraminifera revealed by comparison of ribosomal DNA sequences and the fossil record. Mol Biol Evol. 1997; 14(5):498–505. https://doi.org/10.1093/oxfordjournals.molbev.a025786 PMID: 9159927

69. Schweizer M, Fontaine D, Pawlowski J. Phylogenetic position of two Patagonian Cibicididae (Rotaliida, Foraminifera): Cibicidoides dispars (d'Orbigny, 1839) and Cibicidoides variabilis (d'Orbigny, 1826). Rev de Micropaleontol. 2011; 54(3):175–82. https://doi.org/10.1016/j.revmic.2011.03.002

4. Review Article IV

***Sedimentary ancient DNA: a new paleo-genomic tool for reconstructing the history of marine ecosystems*.** Ngoc-Loi Nguyen, Dhanushka Devendra, Natalia Szymańska, Mattia Greco, Inès Barrenechea Angeles, Agnes K. M. Weiner, Jessica Louise Ray, Tristan Cordier, Stijn De Schepper, Jan Pawłowski, and Joanna Pawłowska. Published in *Frontiers in Marine Science* 10:1075 (2023).

frontiers | Frontiers in Marine Science

TYPE Review
PUBLISHED 09 June 2023
DOI 10.3389/fmars.2023.1185435

OPEN ACCESS

EDITED BY
Christian Grenz,
UMR7294 Institut Méditerranéen d'océanographie (MIO), France

REVIEWED BY
Marco J. L. Coolen,
Curtin University, Australia
Mathilde Bourreau,
École Normale Supérieure, France

*CORRESPONDENCE
Ngoc-Loi Nguyen
loinguyen@iopan.pl
Joanna Pawłowska
pawlowska@iopan.pl

RECEIVED 13 March 2023
ACCEPTED 30 May 2023
PUBLISHED 09 June 2023

CITATION
Nguyen N-L, Devendra D, Szymańska N, Greco M, Angeles IB, Weiner AKM, Ray JL, Cordier T, De Schepper S, Pawłowski J and Pawłowska J (2023) Sedimentary ancient DNA: a new paleogenomic tool for reconstructing the history of marine ecosystems.
Front. Mar. Sci. 10:1185435.
doi: 10.3389/fmars.2023.1185435

Sedimentary ancient DNA: a new paleogenomic tool for reconstructing the history of marine ecosystems

Ngoc-Loi Nguyen[1]*, Dhanushka Devendra[1], Natalia Szymańska[1], Mattia Greco[1], Inès Barrenechea Angeles[2], Agnes K. M. Weiner[3], Jessica Louise Ray[3], Tristan Cordier[3], Stijn De Schepper[3], Jan Pawłowski[1] and Joanna Pawłowska[1]*

[1]Department of Paleoceanography, Institute of Oceanology Polish Academy of Sciences, Sopot, Poland, [2]Department of Geosciences, University of Tromsø (UiT) The Arctic University of Norway, Tromsø, Norway, [3]NORCE Climate and Environment, NORCE Norwegian Research Centre AS and Bjerknes Centre for Climate Research, Bergen, Norway

Sedimentary ancient DNA (*seda*DNA) offers a novel retrospective approach to reconstructing the history of marine ecosystems over geological timescales. Until now, the biological proxies used to reconstruct paleoceanographic and paleoecological conditions were limited to organisms whose remains are preserved in the fossil record. The development of ancient DNA analysis techniques substantially expands the range of studied taxa, providing a holistic overview of past biodiversity. Future development of marine *seda*DNA research is expected to dramatically improve our understanding of how the marine biota responded to changing environmental conditions. However, as an emerging approach, marine *seda*DNA holds many challenges, and its ability to recover reliable past biodiversity information needs to be carefully assessed. This review aims to highlight current advances in marine *seda*DNA research and to discuss potential methodological pitfalls and limitations.

KEYWORDS

paleoceanography, *seda*DNA, marine sediment, metabarcoding, metagenomics, marine biodiversity

1 Introduction

Marine sedimentary archives are an important source of information for understanding the environment in which marine organisms lived over geological time scales. The analysis of these archives adopts a multidisciplinary approach, which requires the engagement of experts in geology, organic and inorganic geochemistry, geomorphology, paleoceanography, and micropaleontology, (e.g., Backman et al. (2009); Łącka et al. (2019); Łącka et al. (2020); Pawłowska et al. (2020a); Marino et al. (2022)). Records of past environmental change can be accessed by analyzing indirect sources of

information (so-called proxies) retrieved from marine sediment cores. These proxies can be fossil assemblages, indicator species, or geochemical proxies such as lipid biomarkers, pigments, or sedimentological properties. Such multidisciplinary approaches have greatly expanded our knowledge of marine paleo-communities and their physical and chemical environments, providing invaluable information for predictions on marine ecosystem response to future climate changes and anthropogenic pressures. However, our current view of past environmental changes and evolutionary responses of marine organisms has, until recently, been limited to selected groups of organisms whose remains are preserved in the fossil record. Examples of such fossilizing taxa are mollusks (Machado et al., 2018; Mcgann and Powell, 2022), coccolithophores (Jacques and Luc, 2007; Marino et al., 2022), foraminifera (Kujawa et al., 2021; Pados-Dibattista et al., 2022), diatoms (Oksman et al., 2017; Miettinen, 2018), and dinoflagellate cysts (Ellegaard et al., 2017; Hennissen et al., 2017; Gussone and Friedrich, 2018; Aubry et al., 2020). The compositional variation of these taxa reflects transformations in marine environmental conditions. Our current understanding of the past ocean is largely based on traditional proxies, such as foraminifera, for which oxygen isotope records provided detailed insight into the climate history during the Cenozoic (Zachos et al., 2008) and Mg/Ca ratios provided sea water temperature estimates (Mcclymont et al., 2020). However, these proxies are often used to infer a specific environmental variable (e.g., temperature, salinity) and do not provide an overview of biodiversity changes. Moreover, their taxonomic analysis is often labor-intensive and requires expertise that is not readily available making impractical their applications for long-term or large-scale assessments.

Over the last decade, advances in high-throughput DNA sequencing technologies and laboratory techniques for working with ancient DNA have offered richer, higher-quality analyses that better facilitate the reconstruction and understanding of paleo-communities. An initial study by Coolen and Overmann (1998) successfully used ancient 16S rDNA fragments from lake sediments and the carotenoid okenone to show that bottom-water anoxia and a sulfidic chemocline persisted in Mahoney Lake, Canada, during most of the Holocene. After this study, the number of *seda*DNA-based publications remained relatively low until 2014. Since then, their number has been consistently increasing, yet studies of marine sediments remain less numerous than those of freshwater environments (Figure 1A). Early marine *seda*DNA studies tested the burial of DNA and compared the microbial communities of *seda*DNA and fossil assemblages exploring various marine habitats from coastal polar regions (Coolen et al., 2004; Boere et al., 2009; Pawłowska et al., 2014), to anoxic basins (Coolen et al., 2009; Boere et al., 2011; Coolen, 2011; Coolen et al., 2013) and abyssal plains (Lejzerowicz et al., 2013). An increasing number of studies is now combining *seda*DNA analyses with other proxy data to gain insight into past oceanographic conditions, such as sea-ice cover (Boere et al., 2009; De Schepper et al., 2019; Pawłowska et al., 2020a; Zimmermann et al., 2020; Zimmermann et al., 2021), sea-surface temperatures and subsurface salinities (Zimmermann et al., 2021). As shown by the main keywords associated with *seda*DNA publications, their main focus so far has been: "paleoenvironments", changes in "biodiversity", "climate change" and "long-term dynamics" (Figure 1B).

This review aims to highlight the potential of sedimentary ancient DNA-based research for the characterization of past marine biodiversity and its response to environmental changes. We describe taphonomic or preservation processes and discuss key issues related to procedures for sampling, laboratory processing, and computational techniques involved in marine *seda*DNA investigation. We focus on both planktonic and benthic microbial (prokaryotes and single-cell eukaryotes) and meiofaunal organisms deposited in marine sediments and discuss their potential in studying biodiversity changes across geological timescales.

2 Taphonomy and characteristics of marine *seda*DNA

Marine *seda*DNA research involves the analysis of genetic material preserved in marine sediments from organisms that once lived in the water column or on the ocean floor. Several investigations in different sedimentary basins shed some light on the taphonomy of *seda*DNA in marine settings (Sunday et al., 2014; Kelly et al., 2018; Ellegaard et al., 2020) (Figure 2). The processes that regulate the accumulation and preservation of environmental DNA (*e*DNA) in marine environments are more complex than in the freshwater-terrestrial system. This is because the marine environment with its surface and bottom water currents, sinking water masses, lateral sediment transports, and sometimes very active benthic communities, is more dynamic than freshwater systems (Pedersen et al., 2015; Barrenechea Angeles et al., 2020; Allan et al., 2021).

The planktonic community, dominated by prokaryotes (Kallmeyer et al., 2012), protists (Vargas et al., 2015), and zooplankton (Murrell and Lores, 2004), is responsible for a large portion of the marine *seda*DNA (Barrenechea Angeles et al., 2020). Several studies have described how planktonic DNA is deposited and preserved in sediments (Corinaldesi et al., 2011; Torti et al., 2018). As shown in Figure 2, after the death of planktonic organisms, their DNA has to travel through the water column for tens (shallow shelf) to thousands (open ocean) of meters to reach the sea floor (Smayda, 1971; Bach et al., 2016). During this journey, the planktonic DNA can be associated with organism remains or attached to particulate organic matter, skeletons, bound detritus, or inorganic mineral grains (Iversen and Ploug, 2010; Herndl and Reinthaler, 2013; Wood et al., 2020). Depending on the size of the sinking particles (Cael and White, 2020) and a range of physical and biological mechanisms (aggregation, downwelling and density inversion currents, packaging of cells in fecal pellets) *in situ* (Bach et al., 2016), the plankton sinking can take several days or even longer time to reach the seafloor (Bach et al., 2019; Nooteboom et al., 2019). The degree of planktonic DNA degradation depends on various factors such as organic matter load, temperature, pH, salinity, water depth, light intensity, and organismal activity through the water column (Corinaldesi et al., 2008; Torti et al., 2015; Andruszkiewicz et al., 2017; Collins et al., 2018; Mccartin et al., 2022). Some

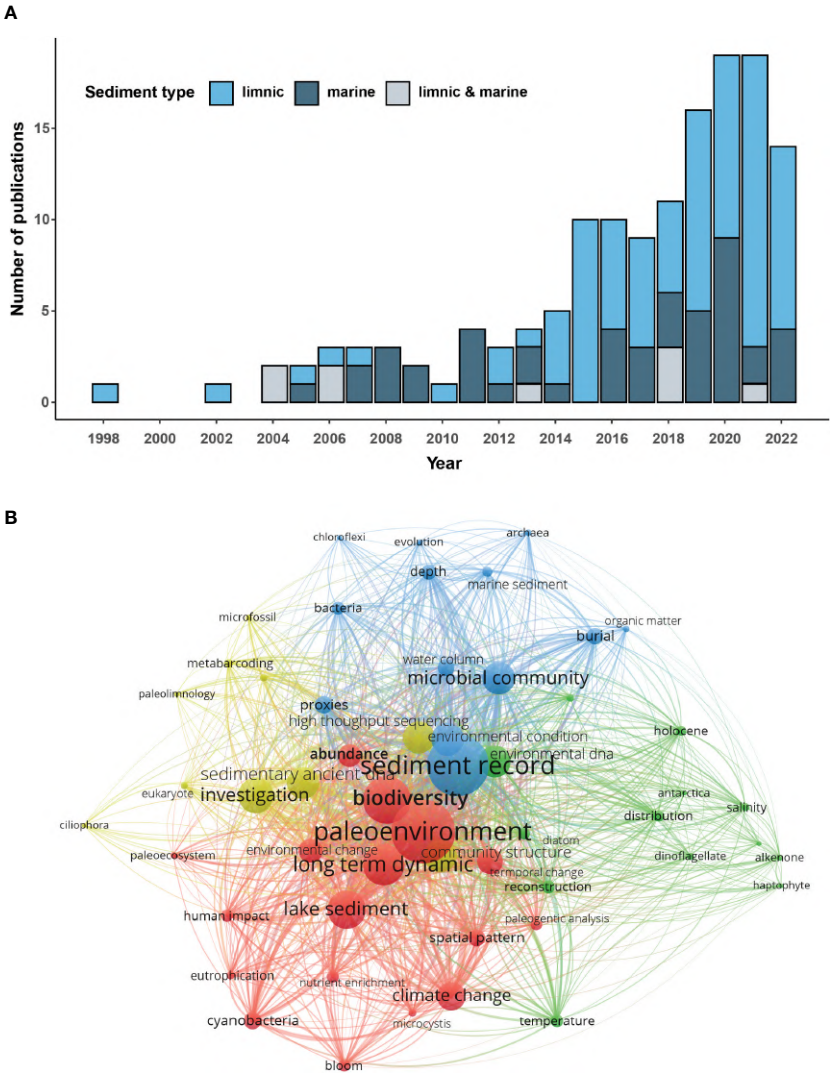


FIGURE 1
(A) Bar chart showing the annual number of published *seda*DNA research articles retrieved from Google Scholar and cross-checked with the literature database of the *seda*DNA scientific society (https://sedadna.github.io/). We queried for papers that (1) presented *seda*DNA data, and (2) targeted either bacteria, archaea, micro-eukaryotes (protists, fungi), meiofauna, or small zooplankton, and verified them manually. **(B)** Co-occurrence network map of keywords found in abstracts of *seda*DNA research papers. The network map was generated using the VOSviewer software (Perianes-Rodriguez et al., 2016). Size of the nodes and font are proportional to the eigenvector centrality; the nodes were colored to show different clusters. A total of 142 original studies on limnic and/or marine *seda*DNA were included in this review (Table S1).

authors suggested that macrobial *e*DNA degradation rates differ between marine, brackish, and freshwater systems (Thomsen et al., 2012; Collins et al., 2018). An experimental and comparative study by Collins et al. (2018) indicated the rate of decay to be 1.6 times faster in inshore waters than in offshore waters. Salinity seemed to be an important factor in steering decay rates, but it could also be related to microbial activities (Collins et al., 2018). In general, degradation rates of *e*DNA decrease in colder seawater (Okabe and Shimazu, 2007; Mccartin et al., 2022) with a higher pH (Collins et al., 2018), high dissolved organic carbon content, and low bacterial activity (Corinaldesi et al., 2008). However, very little is known about how the degradation of planktonic DNA that is going through the water column and deposited on the seafloor can impact the composition of planktonic communities retrieved from *seda*DNA analyses (Morard et al., 2017; Barrenechea Angeles et al., 2020).

The physicochemical characteristics of the sediment in concert with environmental conditions are considered to play important roles in the preservation of *seda*DNA on the seafloor (Corinaldesi et al., 2008; Vuillemin et al., 2019). Early ancient DNA studies indicated that the sediment's characteristics such as high clay, borate, and organic content under cold/frozen and anoxic conditions facilitate optimal DNA preservation (Willerslev et al., 2004; Coolen and Overmann, 2007; Coolen et al., 2013; Furukawa et al., 2013; Torti et al., 2015). Indeed, it has been demonstrated experimentally that organic matter and clay increase the DNA adsorption capacity of sediment (Xue and Feng, 2018). In addition, the adsorption of DNA to sand is also well documented (Lorenz and

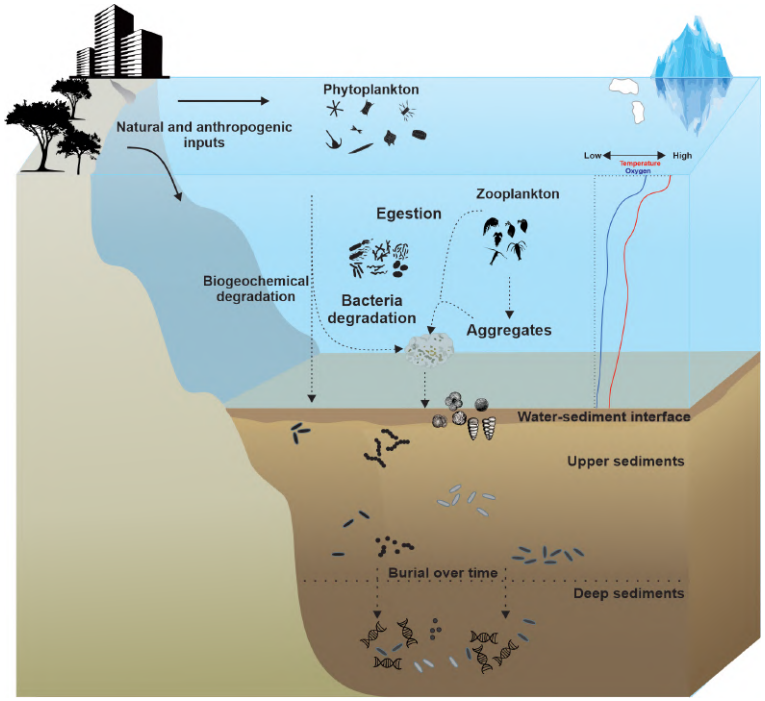


FIGURE 2
Schematic illustration of *e*DNA taphonomic processes in the marine environment. These processes involve *e*DNA distribution, degradation, and/or preferential preservation during the transition from the pelagic to the benthic zones, and from the water-sediment interface into subsurface sediment.

Wackernagel, 1987; Robert et al., 2005). However, overall, the relation between marine sediment characteristics and *seda*DNA preservation over time is poorly understood. Similarly, it remains unclear how bottom water temperature impacts *seda*DNA preservation. Since the bottom water temperature is remarkably stable and colder than 4°C below a depth of 700 m (Locarnini et al., 2018) it is possible that the influence of the bottom water temperature on the preservation of *seda*DNA may be comparable over large parts of the ocean. To date, the majority of *seda*DNA studies have been performed in high-latitude, cold environments, particularly the polar regions (Pawłowska et al., 2014; De Schepper et al., 2019), or deep-sea subsurface sediments (Corinaldesi et al., 2011; More et al., 2018; Armbrecht et al., 2021a). Warmer, shallow, or marginal seas and oceanic regions with strong seasonality were long assumed to be less favorable for the long-term preservation of *e*DNA. However, some studies show that *e*DNA can also be well preserved in temperate and tropical regions, beneath toxic water columns (Coolen et al., 2013; More et al., 2018), in well–oxygenated deep-sea sediments (Lejzerowicz et al., 2013), and tropical reef sediments (Del Carmen Gomez Cabrera et al., 2019).

Although no accurate timeline predicting *seda*DNA preservation potential in various sediment types or geographical regions exists, marine *seda*DNA has been recovered back to the early Quaternary (Figure 3A). Micro-eukaryotic DNA has been recovered from 800-year-old fjord sediments (Pawłowska et al., 2014), 30,000-year-old deep-sea sediments (Lejzerowicz et al., 2013), 43,000-year-old oxygen-minimum zone sediments (More et al., 2018), and 100,000-year-old Greenland Sea sediments (De Schepper et al., 2019). Even longer preservation was documented in sediments dating back to 1–1.4 Ma (Kirkpatrick et al., 2016; Armbrecht et al., 2022). Compared to modern sedimentary DNA, the *seda*DNA molecules are typically degraded into short fragments and may have extensive chemical damage (Armbrecht et al., 2020). However, some studies have shown that ancient DNA fragments longer than 500 base pairs (bp) can be recovered from subsurface sediments, for example, from the eastern Mediterranean Holocene and Pleistocene sapropels with aforementioned conditions for preservation (Coolen and Overmann, 2007; Boere et al., 2011) (see also Table S1; Figure 3A). This could be explained by the ability of some microbes such as bacteria and archaea (Vuillemin et al., 2019; Thomas et al., 2020; Capo et al., 2022), dinoflagellates, and diatoms (Ribeiro et al., 2011; Sanyal et al., 2022), to form dormant or resting stages for extended periods following sediment burial.

3 Marine *seda*DNA processing and analysis

The methods used in *seda*DNA processing and analysis have been reviewed by several authors, e.g. Epp et al. (2019); Fulton and Shapiro (2019); Capo et al. (2021). However, only a few papers concern specifically the marine environment, focusing on the collection, storage, and manipulation of marine sediments for *seda*DNA analysis (Armbrecht et al., 2019; Armbrecht et al., 2020; Selway et al., 2022). Here we will shortly discuss some of these methods with a special focus on challenges related to marine sediments.

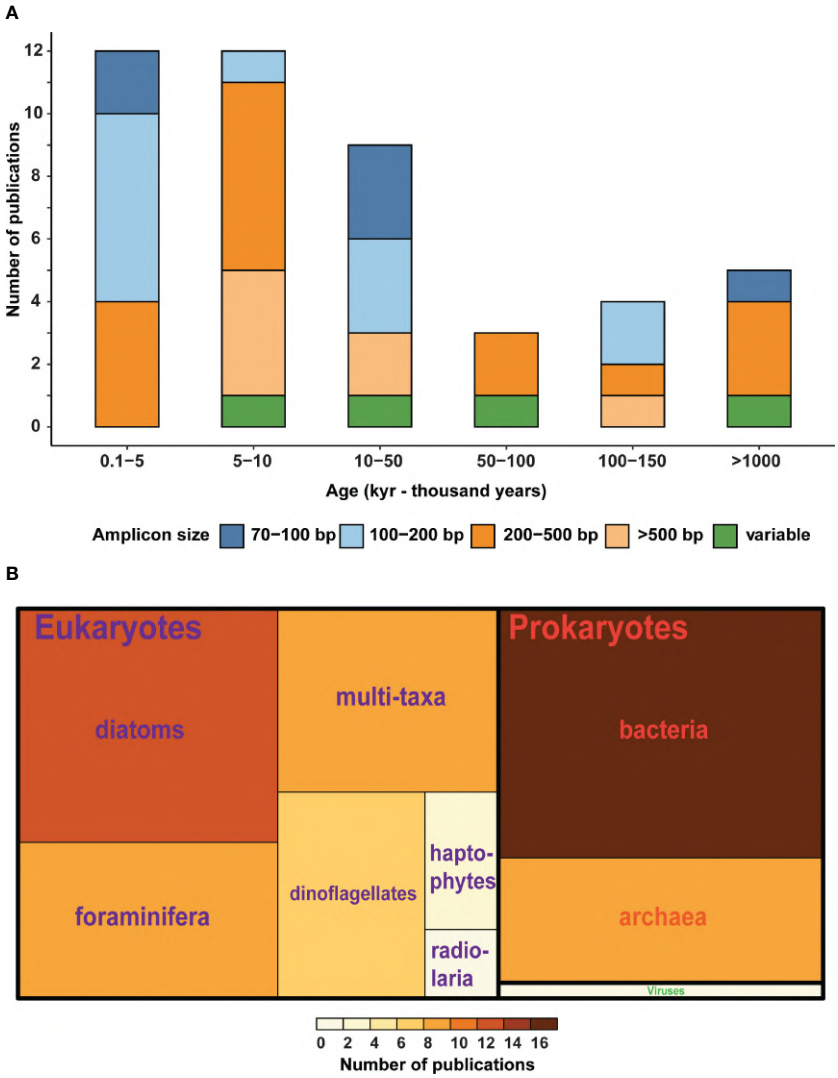


FIGURE 3
(A) Bar chart showing the number of published articles on marine *seda*DNA, indicating the age of the sediment and the fragment length of the analyzed amplicons, variable: a range of fragment sizes is obtained through metagenomics. **(B)** Treemap showing the number of published articles targeting different taxonomic groups of prokaryotes and eukaryotes in marine *seda*DNA. A total of 55 marine *seda*DNA studies were included in this figure (Table S1).

3.1 Sediment coring

The sediment core sampling strategy is the major difference between marine *seda*DNA and other studies involving ancient environmental DNA. In particular, sampling long cores (> 10-meter-long) from depths of hundreds to thousands of meters below the sea surface, while avoiding contamination is the main challenge of marine *seda*DNA. Depending on the functional mechanisms that ensure penetration into the sediment, the following coring methods are commonly used in marine *seda*DNA research: gravity corer (Coolen et al., 2013; Torti et al., 2018; De Schepper et al., 2019; More et al., 2021), piston corer (Zimmermann et al., 2021; Armbrecht et al., 2022), multicorer (Coolen et al., 2013; Lejzerowicz et al., 2013; De Schepper et al., 2019), and boxcorer (Coolen et al., 2006). The gravity and piston corers are used for obtaining sediment records of several meters in length to provide insights into past environmental change over decades, centuries to millennia. Multicore systems, on the other hand, are typically used for short-term records of the seafloor and allow for simultaneous collection of 2, 4 or even 6 short cores (< 1 m).

Since collecting multiple gravity or piston cores in the marine environment can be problematic and costly, splitting the core lengthwise and taking multiple subsamples from the same core is usually the best way to obtain replicate samples for *seda*DNA analysis. Subsequent steps of *seda*DNA analysis should be performed in a sterile environment, ideally in a dedicated ancient DNA laboratory. Subsampling of sediment cores can be performed with sterile knives, spoons, and spatulas (Lejzerowicz et al., 2013; Hou et al., 2014; Szczuciński et al., 2016) or using mini-cores or cut-open syringes (Coolen et al., 2013; Capo et al., 2021; Talas et al., 2021; Armbrecht et al., 2022). The sampling intervals vary according to the objectives of each study, the sedimentation rates

in marine systems, and the weight/volume of samples required for further analysis. Subsampling for *seda*DNA should be conducted rapidly in cold, still-air conditions to avoid exposure to oxygen after the core is opened (Elbaum et al., 2006; Ogata et al., 2021). A recent study shows that although *seda*DNA can be obtained from cores that were split and stored in the dark at 4°C in plastic containers for several years, modern contamination from fungi and other eukaryotes are likely to be observed in such material (Selway et al., 2022). To avoid serious contamination issues, sediment cores should be transported and stored under cold, or even freezing temperatures. Alternatively, sediment cores for *seda*DNA analyses could be subsampled onboard the vessel or immediately upon arrival at the laboratory and stored frozen (Llamas et al., 2017; Selway et al., 2022).

3.2 *seda*DNA extraction

Due to the complexity of marine sediment composition and the wide range of target organisms to be analyzed, choosing a universally optimal *seda*DNA extraction method is challenging (Armbrecht et al., 2020; Murchie et al., 2020; Kang et al., 2021). Extraction of the highly degraded fragments of *seda*DNA has to ensure efficient removal of inhibitors, such as humic acids, while retaining a maximum amount of DNA in the solution. The efficiency of *seda*DNA extraction depends on the complex and variable mineral composition of the sediment (Lekang et al., 2015), and DNA extraction methods (Kang et al., 2021). Due to the interaction between DNA molecules and sediment colloids, different extraction methods can influence the final quantity and quality of total DNA and biodiversity assessments.

Commercially available spin column-based DNA extraction kits (e.g., Qiagen, MP biomedicals, Omega, etc.) are commonly used in marine *seda*DNA studies (Table S1). These products ensure a greater degree of uniformity and consistency than homemade solution-based protocols, especially when extracting DNA from a large number of samples. Yet, the latter, such as the phenol-chloroform method, have other advantages, including being cheaper than commercial kits, usually obtaining good quality and quantity of the extracted DNA, being easily optimized to the sample material, and preferentially recovering longer *seda*DNA fragments (Direito et al., 2012; Armbrecht et al., 2020). However, solution-based DNA extraction protocols can be quite laborious, since toxic chemicals are used and all steps are to be processed manually in a fume hood, which can be inconvenient in a clean laboratory.

It should be noted that using different DNA extraction methods can significantly affect DNA yield and quality and species assemblages (Deiner et al., 2018; Pearman et al., 2020; Kang et al., 2021; Brauer and Bengtsson, 2022). For example, the Qiagen DNeasy PowerMax Soil Kit that allows extraction of up to 10 g of material is recommended by some studies for eukaryote biodiversity surveys from the surface (Pearman et al., 2020; Kang et al., 2021; Pawlowski et al., 2022) and downcore (Epp et al., 2019) sediment samples. The use of DNA-binding spin columns tends to selectively recover large DNA fragments (Armbrecht et al., 2020). Whereas non-column-based methods are more efficient for recovering smaller fragments, particularly when dealing with highly fragmented and degraded *seda*DNA (Armbrecht et al., 2020). Thus, when selecting a DNA extraction protocol, it is important to consider various factors such as the type and quantity of samples, the intended purpose of the study, the availability of equipment, and financial limitations.

3.3 Metabarcoding *vs* metagenomics

Most *seda*DNA studies to date have utilized metabarcoding approaches for the characterization of paleobiodiversity from sediment records. Metabarcoding consists of high-throughput sequencing of PCR-amplified marker genes, thus the success of DNA metabarcoding depends on the selection of an appropriate DNA marker. Ideally, such markers should have sufficiently conserved flanking primer-binding sites to minimize taxonomic biases during PCR amplification (Liu et al., 2020), distinguish targeted taxa, and possess a reference database for assigning taxonomy (Deagle et al., 2014). There are several marker genes commonly used for *seda*DNA metabarcoding studies, including nuclear ribosomal genes for eukaryotic organisms (Pawłowska et al., 2014; De Schepper et al., 2019; Thomas et al., 2019; More et al., 2021), prokaryotes (Torti et al., 2018; Vuillemin et al., 2019), the mitochondrial cytochrome oxidase I (COI) gene for animals (Coolen, 2011; Der Sarkissian et al., 2017), the chloroplast ribulose bisphosphate carboxylase large chain (rcbL) gene for diatoms (Zimmermann et al., 2020; Armbrecht et al., 2021a), and major capsid protein (MCP) for double stranded DNA viruses infecting algae (Coolen, 2011). Based on the research questions, the sequence length of DNA markers must be considered: longer fragments (e.g., 200 bp to 500 bp) usually provide a better ability to discriminate taxa, while shorter fragments (e.g. 100 bp or less) are more likely to be preserved in *seda*DNA (Figure 3A). In some cases, short fragments were also shown to provide effective taxonomic resolution (Meusnier et al., 2008; Pawłowska et al., 2014).

Currently, shotgun metagenomic sequencing is becoming more accessible and popular in marine *seda*DNA research (Armbrecht et al., 2021a; Armbrecht et al., 2021b; Armbrecht et al., 2022; Selway et al., 2022) because it allows detection of all genomic fragments preserved in the sediment independent of length and without the primer bias of metabarcoding (Armbrecht et al., 2021a). In addition, it preserves DNA damage patterns and thus allows for assessing *seda*DNA authenticity. On the other hand, the targeted group of organisms may constitute only a small portion of sequences in shotgun-sequencing samples, making this approach computationally challenging and costly. A recent study showed hybridization capture to be a promising compromise as it increases the yield of target eukaryote *seda*DNA while maintaining the possibility for ancient DNA authentication via damage patterns (Armbrecht et al., 2021b). Since metagenomics approaches allow

sequencing all DNA present in an ancient environmental sample, they have the potential to recover the full paleo-genetic record and provide the most holistic view of past marine biodiversity (Fellows Yates et al., 2021a; Armbrecht et al., 2022; Capo et al., 2022).

3.4 Bioinformatics challenges

The quality assessment and quality control of sequences produced by *sedaDNA* metabarcoding-metagenomics studies are particularly challenging given the potential damage of DNA preserved in the sediments. It is critical to select a bioinformatic pipeline that addresses the challenges specific to the recovery of *sedaDNA* signals. In the context of ancient environmental samples, ASV-based DADA2 (Callahan et al., 2019) may be more accurate than Operational Taxonomical Units (OTUs) for low abundance *sedaDNA* containing damage patterns (substitution) as it allows to detect small variants (Callahan et al., 2019; Porter and Hajibabaei, 2020). It is also important to ensure that the taxonomic classification of *sedaDNA* sequences is carried out using highly curated taxonomic reference databases. Although continuously updated (Guillou et al., 2012; Quast et al., 2012), such databases are largely incomplete, especially for marine meio- and microorganisms. The gaps in reference databases can be a key challenge associated with *e*DNA studies of both paleo- and modern communities since those are often dominated by taxonomically unassigned or unknown sequence fragments (Vargas et al., 2015; Cordier et al., 2022). The situation is also difficult in the case of metagenomic studies targeting marine eukaryotes as the number of reference genomes remains limited (Delmont et al., 2022).

In response to the rapidly growing production of ancient metagenomics data, new bioinformatics techniques have been developed to deal with the particular features of ancient DNA sequences. These techniques encompass a range of approaches, such as the integration of damage detection algorithms, e.g. in mapDamage (Jónsson et al., 2013; Kistler et al., 2017; Weyrich et al., 2017) and metaDMG (Michelsen et al., 2022). They also involve optimizing ancient DNA mapping through the analysis of cytosine residues deamination, utilizing algorithms such as BWA (Schubert et al., 2012; Weyrich et al., 2017), PyDamage (Borry et al., 2021). Furthermore, new tools have been developed to enhance the accuracy of taxonomic assignment of metagenomic reads (PIA, Cribdon et al. (2020), and for processing and analyzing ancient metagenomics shotgun data, specifically targeting ultra-short molecules, (e.g. Collin et al. (2020); Fellows Yates et al. (2021b); Pochon et al. (2022); Neuenschwander et al. (2023)). However, identifying ancient sequences is still a challenge due to the lack of standard bioinformatics pipelines to analyze DNA metabarcoding or shotgun metagenomics data. Using different bioinformatics procedures or genetic databases that lack consistent standards increases ambiguity and bias. Additionally, to ensure in silico reproducibility and facilitate further research, it is crucial to emphasize the need for publishing comprehensive bioinformatics analysis reports. The application and optimization of bioinformatics pipelines should be the focus of future research to accurately identify and authenticate marine microbiota in *sedaDNA*.

4 Current applications of the marine *sedaDNA* approach

The main application of marine *sedaDNA* is the reconstruction of past biodiversity in relation to environmental changes at geological time scales (Figure 4). In this context, the *sedaDNA* approach provides a unique insight into the marine paleo-ecological communities. Its application allows reconstructing a detailed record of community changes from individuals to populations and species that occurred across time, increasing the accuracy of prediction models in anticipating how climate/environmental change affects biodiversity (De Schepper et al., 2019; Epp, 2019). Several studies have demonstrated the potential of *sedaDNA* for screening genetic signals of multiple taxa (as in Figure 3B), including non-fossilizing organisms (Pawłowska et al., 2014; Selway et al., 2022; Barrenechea Angeles et al., 2023). Most studies focused on marine microbial and meiofaunal communities, revealing a huge and largely undescribed diversity of viruses (Coolen, 2011; Pratas and Pinho, 2018; Zheng et al., 2021), bacteria (Hou et al., 2014; More et al., 2019; Nwosu et al., 2021), archaea (Torti et al., 2018; Vuillemin et al., 2019), and protists (Lejzerowicz et al., 2013; Pawłowska et al., 2014; More et al., 2018; Zimmermann et al., 2020; Egge et al., 2021).

The aims of *sedaDNA* studies vary from comparing the diversity of molecular and microfossil assemblages to analyzing the past diversity of non-fossilized taxa. Several studies demonstrated the lack of congruence between molecular and microfossil assemblages (Pawłowska et al., 2014; Barrenechea Angeles et al., 2020). This lack of congruence could be explained by difficulties in extracting DNA from hard-shelled, fossilized taxa, multiple copies of the target gene in the genome, PCR primers biases, or PCR inhibitors, e.g., in the case of several planktonic taxa (Barrenechea Angeles et al., 2020; Pierella Karlusich et al., 2023). Another possible explanation would be that the non-fossilized taxa are much more abundant, and their DNA dominates in sediment samples (Lejzerowicz et al., 2013). This could be the case for non-fossilized monothalamous foraminifera and the acantharian radiolarians that usually are more dominant in *sedaDNA* datasets compared to fossilizing species of the same lineages (Lejzerowicz et al., 2013; Pawłowska et al., 2014). Indeed, they also dominate in modern sediment analyses (Lecroq et al., 2011; Pawłowska et al., 2014; Cordier et al., 2022). Despite this lack of congruence in community composition, the patterns inferred from metabarcoding and microfossil data are often similar. This was the case in one study of planktonic foraminifera microfossil and metabarcoding assemblages that exhibited congruent regional biogeographical patterns (Barrenechea Angeles et al., 2020). Similarly, *sedaDNA* data of non-fossilized soft-walled foraminifera show comparable patterns as microfossil foraminiferal assemblage in 800-year-old fjord sediments (Pawłowska et al., 2014).

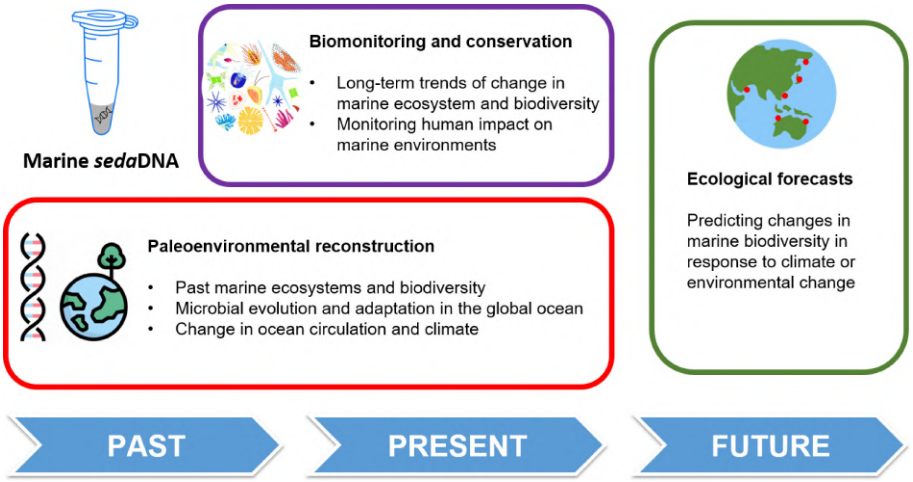


FIGURE 4
Potential and future applications of marine *seda*DNA.

The most promising application of *seda*DNA is to investigate past climates and their impact on oceanographic circulation and species migration. Until now, these studies were mainly based on macro- and microfossils, ignoring a large diversity of non-fossilized species, in particular those belonging to microbial and meiofaunal communities, which are essential for ecosystem functioning. This issue could be overcome by analyzing *seda*DNA from marine sediment cores. For example, changes in ocean circulation patterns were investigated by targeting *seda*DNA of non-fossilized foraminifera in cores collected east of Svalbard (Pawłowska et al., 2020a). Several studies integrate *seda*DNA data with geochemical proxies, including stable and radiogenic isotopes (More et al., 2018; Voldstad et al., 2020), and biomarkers (De Schepper et al., 2019). For example, through analysis of *seda*DNA and hydrogen isotopes of haptophyte-derived alkenones preserved in early Holocene oxygenated lacustrine and mid-to-late-Holocene anoxic coastal Black Sea sediments, planktonic community structures were reconstructed and associated with Holocene climate phases and transitions (Coolen et al., 2013). Paired analysis of *seda*DNA and the sea ice biomarker IP_{25} was used as a proxy for sea-ice reconstructions by targeting diatom *seda*DNA composition in the Fram Strait (Zimmermann et al., 2020), Subarctic North Pacific (Zimmermann et al., 2021), and to trace a sea-ice dinoflagellate east of Greenland (De Schepper et al., 2019). All of these studies provide complementary insights into ecosystem-climate histories, extending our current view of marine communities beyond the diversity found in modern populations and advancing our understanding of their past biogeographic patterns and adaptations to new environmental conditions.

At a shorter timescale, marine *seda*DNA studies also represent a promising complement to traditional biomonitoring surveys. Currently, modern sedimentary DNA metabarcoding is becoming a routine tool for monitoring the human impact on marine systems e.g., through construction, overexploitation, agriculture, habitat loss, and pollution (Bálint et al., 2018; Cordier et al., 2021; Pawlowski et al., 2021). A few studies have shown that the *seda*DNA approach can be used to investigate the impact of anthropogenic activities on the ecosystem over the past century. One of these studies recovered the composition of marine plankton communities from hundred-year-old sediment samples and showed irreversible shifts after the cumulative effect of war and agricultural pollution (Siano et al., 2021). Another study demonstrated dramatic ecosystem changes resulting from a multi-level cascade effect of impacts associated with industrial activities, urbanization, water circulation, and land-use changes in one of the most polluted marine sites in Europe (Barrenechea Angeles et al., 2023). Both studies demonstrate the potential of *seda*DNA to elucidate the effects of human pollution on marine communities, contributing to the reconstruction of reference conditions and helping the conservation and management of marine and coastal ecosystems.

5 Future perspectives

Future developments in the marine *seda*DNA research field will make it easier to integrate paleogenetic signals preserved in the marine seafloor with other proxies (Figure 4). Based on the best understanding of each proxy, we can use the strengths of each approach to their best advantage (i.e., complementary patterns of DNA metabarcoding and morphology). The marine *seda*DNA allows recovering a wide range of organisms that will provide valuable material for searching for new paleo-bioindicators. These could include non-fossilized microbial and meiofaunal taxa that present various kinds of ecological characteristics, or the genotypes of fossilized species, the distribution of which can be used to detect paleo-environmental changes (Pawłowska et al., 2020b). In the future, *seda*DNA studies could also pay more attention to marine

mammals, fish, and macroinvertebrates, traces of which can be detected in sediment samples (Kuwae et al., 2020).

However, several technical challenges inherent to *seda*DNA research need to be considered for future developments in the field. First, incomplete reference databases are the major factor limiting the assignment of sequences to taxonomic names and might lead to divergent results (Balvociute and Huson, 2017). The continuous addition of new reference sequences from modern marine organisms to taxonomic databases is important to further complete the taxonomic picture of marine paleo communities. Second, marine *seda*DNA sampling and laboratory procedures may need further optimization (Armbrecht et al., 2019). In particular, the combination of amplicon sequencing (metabarcoding) with metagenomics through the further development of HTS technologies may improve the authentication of ancient signals in genetic data (Armbrecht et al., 2021a) and allow getting more diverse data from assembled sequences, at least for prokaryotes. Finally, more studies are needed to evaluate the limitations of the *seda*DNA approach, especially the time limits of DNA preservation in marine sediments. There is also a lack of evidence on which physicochemical characteristics of sediments are optimal for long-time DNA preservation.

Given the developments in the field, we foresee that the *seda*DNA approach will be rapidly integrated into routine paleoceanographic research. The approach is technologically mature, thus the costs of including a *seda*DNA module in research projects should continue to decrease, thereby increasing the feasibility of *seda*DNA as a regular inclusion in multi-proxy investigations. Even though there are still some challenging issues to be solved, the information provided by the *seda*DNA data is highly valuable. The DNA-based holistic overview of biodiversity changes through time is unique. Its various applications, from the studies of climate changes and water mass circulation at geological time scales to the monitoring of recent anthropogenic impacts, are of key importance to understanding the past and present state of marine ecosystems. Further development of the *seda*DNA field and its wider integration will not only help to improve our knowledge of past changes affecting the ocean and coastal ecosystems, but it will also help to establish and optimize strategies for their conservation and management.

Author contributions

NL-N, JaP and JoP conceived the review; all authors drafted the manuscript. MG and DD drew the figures. All authors contributed to the article and approved the submitted version.

Funding

The research leading to these results has been funded by Norwegian Financial Mechanism for 2014-2021, project no. 2019/34/H/ST10/00682 in the frame of project "Sedimentary ancient DNA - a new proxy to investigate the impact of environmental change on past and present biodiversity in Nordic Seas". IBA was supported by the Norwegian Petroleum Directorate and the project AKMA, project no. 287869.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2023.1185435/full#supplementary-material

References

Allan, E. A., Dibenedetto, M. H., Lavery, A. C., Govindarajan, A. F., and Zhang, W. G. (2021). Modeling characterization of the vertical and temporal variability of environmental DNA in the mesopelagic ocean. *Sci. Rep.* 11, 21273. doi: 10.1038/s41598-021-00288-5

Andruszkiewicz, E. A., Sassoubre, L. M., and Boehm, A. B. (2017). Persistence of marine fish environmental DNA and the influence of sunlight. *PloS One* 12, e0185043. doi: 10.1371/journal.pone.0185043

Armbrecht, L. H., Coolen, M. J. L., Lejzerowicz, F., George, S. C., Negandhi, K., Suzuki, Y., et al. (2019). Ancient DNA from marine sediments: precautions and considerations for seafloor coring, sample handling and data generation. *Earth-Sci. Rev.* 196, 102887. doi: 10.1016/j.earscirev.2019.102887

Armbrecht, L., Eisenhofer, R., Utge, J., Sibert, E. C., Rocha, F., Ward, R., et al. (2021a). Paleo-diatom composition from Santa Barbara Basin deep-sea sediments: a comparison of 18S-V9 and *diat-rbcL* metabarcoding vs shotgun metagenomics. *ISME Commun.* 1, 66. doi: 10.1038/s43705-021-00070-8

Armbrecht, L., Hallegraeff, G., Bolch, C. J. S., Woodward, C., and Cooper, A. (2021b). Hybridisation capture allows DNA damage analysis of ancient marine eukaryotes. *Sci. Rep.* 11, 3220. doi: 10.1038/s41598-021-82578-6

Armbrecht, L., Herrando-Pérez, S., Eisenhofer, R., Hallegraeff, G. M., Bolch, C. J. S., and Cooper, A. (2020). An optimized method for the extraction of ancient eukaryote DNA from marine sediments. *Mol. Ecol. Resour.* 20, 906–919. doi: 10.1111/1755-0998.13162

Armbrecht, L., Weber, M. E., Raymo, M. E., Peck, V. L., Williams, T., Warnock, J., et al. (2022). Ancient marine sediment DNA reveals diatom transition in Antarctica. *Nat. Commun.* 13, 5787. doi: 10.1038/s41467-022-33494-4

Aubry, A. M. R., De Schepper, S., and De Vernal, A. (2020). Dinocyst and acritarch biostratigraphy of the late Pliocene to early Pleistocene at integrated ocean drilling program site U1307 in the Labrador Sea. *J. Micropalaeontol.* 39, 41–60. doi: 10.5194/jm-39-41-2020

Bach, L. T., Boxhammer, T., Larsen, A., Hildebrandt, N., Schulz, K. G., and Riebesell, U. (2016). Influence of plankton community structure on the sinking velocity of marine aggregates. *Glob. Biogeochem. Cycles* 30, 1145–1165. doi: 10.1002/2016GB005372

Bach, L. T., Stange, P., Taucher, J., Achterberg, E. P., Algueró-Muñiz, M., Horn, H., et al. (2019). The influence of plankton community structure on sinking velocity and remineralization rate of marine aggregates. *Global Biogeochem. Cycles* 33, 971–994. doi: 10.1029/2019GB006256

Backman, J., Fornaciari, E., and Rio, D. (2009). Biochronology and paleoceanography of late Pleistocene and Holocene calcareous nannofossil abundances across the Arctic Basin. *Mar. Micropaleontol.* 72, 86–98. doi: 10.1016/j.marmicro.2009.04.001

Bálint, M., Pfenninger, M., Grossart, H.-P., Taberlet, P., Vellend, M., Leibold, M. A., et al. (2018). Environmental DNA time series in ecology. *Trends Ecol. Evol.* 33, 945–957. doi: 10.1016/j.tree.2018.09.003

Balvociute, M., and Huson, D. H. (2017). SILVA, RDP, Green genes, NCBI and OTT - how do these taxonomies compare? *BMC Genom.* 18, 114. doi: 10.1186/s12864-017-3501-4

Barrenechea Angeles, I., Lejzerowicz, F., Cordier, T., Scheplitz, J., Kucera, M., Ariztegui, D., et al. (2020). Planktonic foraminifera *e*DNA signature deposited on the seafloor remains preserved after burial in marine sediments. *Sci. Rep.* 10, 20351. doi: 10.1038/s41598-020-77179-8

Barrenechea Angeles, I., Romero-Martínez, M. L., Cavaliere, M., Varrella, S., Francescangeli, F., Piredda, R., et al. (2023). Encapsulated in sediments: *e*DNA deciphers the ecosystem history of one of the most polluted European marine sites. *Environ. Int.* 172, 107738. doi: 10.1016/j.envint.2023.107738

Boere, A. C., Abbas, B., Rijpstra, W. I. C., Versteegh, G. J. M., Volkman, J. K., Sinninghe Damsté, J. S., et al. (2009). Late-Holocene succession of dinoflagellates in an Antarctic fjord using a multi-proxy approach: paleoenvironmental genomics, lipid biomarkers and palynomorphs. *Geobiology* 7, 265–281. doi: 10.1111/j.1472-4669.2009.00202.x

Boere, A. C., Rijpstra, W. I. C., De Lange, G. J., Sinninghe Damsté, J. S., and Coolen, M. J. L. (2011). Preservation potential of ancient plankton DNA in Pleistocene marine sediments. *Geobiology* 9, 377–393. doi: 10.1111/j.1472-4669.2011.00290.x

Borry, M., Hübner, A., Rohrlach, A. B., and Warinner, C. (2021). PyDamage: automated ancient damage identification and estimation for contigs in ancient DNA *de novo* assembly. *PeerJ* 9, e11845. doi: 10.7717/peerj.11845

Brauer, A., and Bengtsson, M. M. (2022). DNA Extraction bias is more pronounced for microbial eukaryotes than for prokaryotes. *MicrobiologyOpen* 11, e1323. doi: 10.1002/mbo3.1323

Cael, B. B., and White, A. E. (2020). Sinking versus suspended particle size distributions in the North Pacific subtropical gyre. *Geophys. Res. Lett.* 47, e2020GL087825. doi: 10.1029/2020GL087825

Callahan, B. J., Wong, J., Heiner, C., Oh, S., Theriot, C. M., Gulati, A. S., et al. (2019). High-throughput amplicon sequencing of the full-length 16S rRNA gene with single-nucleotide resolution. *Nucleic Acids Res.* 47, e103–e103. doi: 10.1093/nar/gkz569

Capo, E., Giguet-Covex, C., Rouillard, A., Nota, K., Heintzman, P. D., Vuillemin, A., et al. (2021). Lake sedimentary DNA research on past terrestrial and aquatic biodiversity: overview and recommendations. *Quaternary* 4, 6. doi: 10.3390/quat4010006

Capo, E., Monchamp, M.-E., Coolen, M. J. L., Domaizon, I., Armbrecht, L., and Bertilsson, S. (2022). Environmental paleomicrobiology: using DNA preserved in aquatic sediments to its full potential. *Environ. Microbiol* 24, 2201–2209. doi: 10.1111/1462-2920.15913

Collin, T. C., Drosou, K., O Riordan, J. D., Meshveliani, T., Pinhasi, R., and Feeney, R. N. M. (2020). An open-sourced bioinformatic pipeline for the processing of next-generation sequencing derived nucleotide reads: identification and authentication of ancient metagenomic DNA. *bioRxiv*. doi: 10.1101/2020.04.20.050369

Collins, R. A., Wangensteen, O. S., O'gorman, E. J., Mariani, S., Sims, D. W., and Genner, M. J. (2018). Persistence of environmental DNA in marine systems. *Commun. Biol.* 1, 185. doi: 10.1038/s42003-018-0192-6

Coolen, M. J. (2011). 7000 years of *Emiliania huxleyi* viruses in the Black Sea. *Science* 333, 451–452. doi: 10.1126/science.1200072

Coolen, M. J. L., Boere, A., Abbas, B., Baas, M., Wakeham, S. G., and Sinninghe Damsté, J. S. (2006). Ancient DNA derived from alkenone-biosynthesizing haptophytes and other algae in Holocene sediments from the Black Sea. *Paleoceanography* 21. doi: 10.1029/2005PA001188

Coolen, M. J. L., Muyzer, G., Rijpstra, W. I. C., Schouten, S., Volkman, J. K., and Sinninghe Damsté, J. S. (2004). Combined DNA and lipid analyses of sediments reveal changes in Holocene haptophyte and diatom populations in an Antarctic lake. *Earth Planet. Sci. Lett.* 223, 225–239. doi: 10.1016/j.epsl.2004.04.014

Coolen, M. J., Orsi, W. D., Balkema, C., Quince, C., Harris, K., Sylva, S. P., et al. (2013). Evolution of the plankton paleome in the Black Sea from the deglacial to anthropocene. *Proc. Natl. Acad. Sci. U.S.A.* 110, 8609–8614. doi: 10.1073/pnas.1219283110

Coolen, M. J. L., and Overmann, J. (1998). Analysis of subfossil molecular remains of purple sulfur bacteria in a lake sediment. *Appl. Environ. Microbiol.* 64, 4513–4521. doi: 10.1128/AEM.64.11.4513-4521.1998

Coolen, M. J. L., and Overmann, J. (2007). 217 000-year-old DNA sequences of green sulfur bacteria in Mediterranean sapropels and their implications for the reconstruction of the paleoenvironment. *Environ. Microbiol.* 9, 238–249. doi: 10.1111/j.1462-2920.2006.01134.x

Coolen, M. J. L., Saenz, J. P., Giosan, L., Trowbridge, N. Y., Dimitrov, P., Dimitrov, D., et al. (2009). DNA and lipid molecular stratigraphic records of haptophyte succession in the Black Sea during the Holocene. *Earth Planet. Sci. Lett.* 284, 610–621. doi: 10.1016/j.epsl.2009.05.029

Cordier, T., Alonso-Sáez, L., Apothéloz-Perret-Gentil, L., Aylagas, E., Bohan, D. A., Bouchez, A., et al. (2021). Ecosystems monitoring powered by environmental genomics: a review of current strategies with an implementation roadmap. *Mol. Ecol.* 30, 2937–2958. doi: 10.1111/mec.15472

Cordier, T., Angeles, I. B., Henry, N., Lejzerowicz, F., Berney, C., Morard, R., et al. (2022). Patterns of eukaryotic diversity from the surface to the deep-ocean sediment. *Sci. Adv.* 8, eabj9309. doi: 10.1126/sciadv.abj9309

Corinaldesi, C., Barucca, M., Luna, G. M., and Dell'anno, A. (2011). Preservation, origin and genetic imprint of extracellular DNA in permanently anoxic deep-sea sediments. *Mol. Ecol.* 20, 642–654. doi: 10.1111/j.1365-294X.2010.04958.x

Corinaldesi, C., Beolchini, F., and Dell'anno, A. (2008). Damage and degradation rates of extracellular DNA in marine sediments: implications for the preservation of gene sequences. *Mol. Ecol.* 17, 3939–3951. doi: 10.1111/j.1365-294X.2008.03880.x

Cribdon, B., Ware, R., Smith, O., Gaffney, V., and Allaby, R. G. (2020). PIA: more accurate taxonomic assignment of metagenomic data demonstrated on *sed*aDNA from the North Sea. *Front. Ecol. Evol.* 8. doi: 10.3389/fevo.2020.00084

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

Deiner, K., Lopez, J., Bourne, S., Holman, L., Seymour, M., Grey, E. K., et al. (2018). Optimising the detection of marine taxonomic richness using environmental DNA metabarcoding: the effects of filter material, pore size and extraction method. *Metabarcoding Metagenom.* 2, e28963. doi: 10.3897/mbmg.2.28963

Del Carmen Gomez Cabrera, M., Young, J. M., Roff, G., Staples, T., Ortiz, J. C., Pandolfi, J. M., et al. (2019). Broadening the taxonomic scope of coral reef palaeoecological studies using ancient DNA. *Mol. Ecol.* 28, 2636–2652. doi: 10.1111/mec.15038

Delmont, T. O., Gaia, M., Hinsinger, D. D., Frémont, P., Vanni, C., Fernandez-Guerra, A., et al. (2022). Functional repertoire convergence of distantly related eukaryotic plankton lineages abundant in the sunlit ocean. *Cell Genom.* 2, 100123. doi: 10.1016/j.xgen.2022.100123

Der Sarkissian, C., Pichereau, V., Dupont, C., Ilsoe, P. C., Perrigault, M., Butler, P., et al. (2017). Ancient DNA analysis identifies marine mollusc shells as new metagenomic archives of the past. *Mol. Ecol. Resour.* 17, 835–853. doi: 10.1111/1755-0998.12679

De Schepper, S., Ray, J. L., Skaar, K. S., Sadatzki, H., Ijaz, U. Z., Stein, R., et al. (2019). The potential of sedimentary ancient DNA for reconstructing past sea ice evolution. *ISME J.* 13, 2566–2577. doi: 10.1038/s41396-019-0457-1

Direito, S. O. L., Marees, A., and Röling, W. F. M. (2012). Sensitive life detection strategies for low-biomass environments: optimizing extraction of nucleic acids adsorbing to terrestrial and Mars analogue minerals. *FEMS Microbiol. Ecol.* 81, 111–123. doi: 10.1111/j.1574-6941.2012.01325.x

Egge, E., Elferink, S., Vaulot, D., John, U., Bratbak, G., Larsen, A., et al. (2021). An 18S V4 rRNA metabarcoding dataset of protist diversity in the Atlantic inflow to the Arctic Ocean, through the year and down to 1000 m depth. *Earth Syst. Sci. Data* 13, 4913–4928. doi: 10.5194/essd-13-4913-2021

Elbaum, R., Melamed-Bessudo, C., Boaretto, E., Galili, E., Lev-Yadun, S., Levy, A. A., et al. (2006). Ancient olive DNA in pits: preservation, amplification and sequence analysis. *J. Archaeol. Sci.* 33, 77–88. doi: 10.1016/j.jas.2005.06.011

Ellegaard, M., Clokie, M. R. J., Czypionka, T., Frisch, D., Godhe, A., Kremp, A., et al. (2020). Dead or alive: sediment DNA archives as tools for tracking aquatic evolution and adaptation. *Commun. Biol.* 3, 169. doi: 10.1038/s42003-020-0899-z

Ellegaard, M., Dale, B., Mertens, K. N., Pospelova, V., and Ribeiro, S. (2017). "Dinoflagellate cysts as proxies for Holocene environmental change in estuaries: diversity, abundance and morphology," in *Applications of paleoenvironmental techniques in estuarine studies*. Eds. K. Weckström, K. M. Saunders, P. A. Gell and C. G. Skilbeck (Dordrecht: Springer Netherlands), 295–312.

Epp, L. S. (2019). A global perspective for biodiversity history with ancient environmental DNA. *Mol. Ecol.* 28, 2456–2458. doi: 10.1111/mec.15118

Epp, L. S., Zimmermann, H. H., and Stoof-Leichsenring, K. R. (2019). Sampling and extraction of ancient DNA from sediments. *Methods Mol. Biol.* 1963, 31–44. doi: 10.1007/978-1-4939-9176-1_5

Fellows Yates, J. A., Andrades Valtueña, A., Vågene, Å.J., Cribdon, B., Velsko, I. M., Borry, M., et al. (2021a). Community-curated and standardised metadata of published ancient metagenomic samples with AncientMetagenomeDir. *Sci. Data* 8, 31. doi: 10.1038/s41597-021-00816-y

Fellows Yates, J. A., Lamnidis, T. C., Borry, M., Andrades Valtueña, A., Fagernäs, Z., Clayton, S., et al. (2021b). Reproducible, portable, and efficient ancient genome reconstruction with nf-core/eager. *PeerJ* 9, e10947. doi: 10.7717/peerj.10947

Fulton, T. L., and Shapiro, B. (2019). "Setting up an ancient DNA laboratory," in *Ancient DNA: methods and protocols*. Eds. B. Shapiro, A. Barlow, P. D. Heintzman, M. Hofreiter, J. L. A. Paijmans and A. E. R. Soares (New York, NY: Springer New York), 1–13.

Furukawa, Y., Horiuchi, M., and Kakegawa, T. (2013). Selective stabilization of ribose by borate. *Orig. Life Evol. Biosph.* 43, 353–361. doi: 10.1007/s11084-013-9350-5

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., et al. (2012). The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. *Nucleic Acids Res.* 41, D597–D604. doi: 10.1093/nar/gks1160

Gussone, N., and Friedrich, O. (2018). Cretaceous calcareous dinoflagellate cysts as recorder of δ44/40Caseawater and paleo-temperature using Sr/Ca thermometry. *Chem. Geol.* 488, 138–148. doi: 10.1016/j.chemgeo.2018.04.020

Hennissen, J., Head, M. J., De Schepper, S., and Groeneveld, J. (2017). Dinoflagellate cyst paleoecology during the Pliocene–Pleistocene climatic transition in the North Atlantic. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 470, 81–108. doi: 10.1016/j.palaeo.2016.12.023

Herndl, G. J., and Reinthaler, T. (2013). Microbial control of the dark end of the biological pump. *Nat. Geosci.* 6, 718–724. doi: 10.1038/ngeo1921

Hou, W., Dong, H., Li, G., Yang, J., Coolen, M. J. L., Liu, X., et al. (2014). Identification of photosynthetic plankton communities using sedimentary ancient DNA and their response to late-Holocene climate change on the Tibetan Plateau. *Sci. Rep.* 4, 6648. doi: 10.1038/srep06648

Iversen, M. H., and 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. doi: 10.5194/bg-7-2613-2010

Jacques, G., and Luc, B. (2007). "Chapter ten coccolithophores: from extant populations to fossil assemblages," in *Developments in marine geology*. Eds. C. Hillaire–Marcel and A. Vernal (Elsevier), 409–439.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29, 1682–1684. doi: 10.1093/bioinformatics/btt193

Kallmeyer, J., Pockalny, R., Adhikari, R. R., Smith, D. C., and D'hondt, S. (2012). Global distribution of microbial abundance and biomass in subseafloor sediment. *Proc. Natl. Acad. Sci. U.S.A.* 109, 16213–16216. doi: 10.1073/pnas.1203849109

Kang, W., Anslan, S., Börner, N., Schwarz, A., Schmidt, R., Künzel, S., et al. (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. *Ecol. Indic.* 121, 107070. doi: 10.1016/j.ecolind.2020.107070

Kelly, R. P., Gallego, R., and Jacobs-Palmer, E. (2018). The effect of tides on nearshore environmental DNA. *PeerJ* 6, e4521. doi: 10.7717/peerj.4521

Kirkpatrick, J. B., Walsh, E. A., and 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. doi: 10.1130/G37933.1

Kistler, L., Ware, R., Smith, O., Collins, M., and Allaby, R. G. (2017). A new model for ancient DNA decay based on paleogenomic meta-analysis. *Nucleic Acids Res.* 45, 6310–6320. doi: 10.1093/nar/gkx361

Kujawa, A., Łącka, M., Szymańska, N., Pawłowska, J., Telesiński, M. M., and Zajączkowski, M. (2021). Could Norwegian fjords serve as an analogue for the future of the Svalbard fjords? state and fate of high latitude fjords in the face of progressive "atlantification". *Polar Biol.* 44, 2217–2233. doi: 10.1007/s00300-021-02951-z

Kuwae, M., Tamai, H., Doi, H., Sakata, M. K., Minamoto, T., and Suzuki, Y. (2020). Sedimentary DNA tracks decadal-centennial changes in fish abundance. *Commun. Biol.* 3, 558. doi: 10.1038/s42003-020-01282-9

Łącka, M., Cao, M., Rosell-Melé, A., Pawłowska, J., Kucharska, M., Forwick, M., et al. (2019). Postglacial paleoceanography of the western Barents Sea: implications for alkenone-based sea surface temperatures and primary productivity. *Quat. Sci. Rev.* 224, 105973. doi: 10.1016/j.quascirev.2019.105973

Łącka, M., Michalska, D., Pawłowska, J., Szymańska, N., Szczuciński, W., Forwick, M., et al. (2020). Multiproxy paleoceanographic study from the western Barents Sea reveals dramatic younger dryas onset followed by oscillatory warming trend. *Sci. Rep.* 10, 15667. doi: 10.1038/s41598-020-72747-4

Lecroq, B., Lejzerowicz, F., Bachar, D., Christen, R., Esling, P., Baerlocher, L., et al. (2011). Ultra-deep sequencing of foraminiferal microbarcodes unveils hidden richness of early monothalamous lineages in deep-sea sediments. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13177–13182. doi: 10.1073/pnas.1018426108

Lejzerowicz, F., Esling, P., Majewski, W., Szczucinski, W., Decelle, J., Obadia, C., et al. (2013). Ancient DNA complements microfossil record in deep-sea subsurface sediments. *Biol. Lett.* 9, 20130283. doi: 10.1098/rsbl.2013.0283

Lekang, K., Thompson, E., and Troedsson, C. (2015). A comparison of DNA extraction methods for biodiversity studies of eukaryotes in marine sediments. *Aquat. Microb. Ecol.* 75, 15–25. doi: 10.3354/ame01741

Liu, M., Clarke, L. J., Baker, S. C., Jordan, G. J., and Burridge, C. P. (2020). A practical guide to DNA metabarcoding for entomological ecologists. *Ecol. Entomol.* 45, 373–385. doi: 10.1111/een.12831

Llamas, B., Valverde, G., Fehren-Schmitz, L., Weyrich, L. S., Cooper, A., and Haak, W. (2017). From the field to the laboratory: controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *STAR: Sci. Technol. Archaeol. Res.* 3, 1–14. doi: 10.1080/20548923.2016.1258824

Locarnini, M., Mishonov, A., Baranova, O., Boyer, T., Zweng, M., Garcia, H., et al. (2018). "World ocean atlas 2018," in *Temperature*, vol. 1. (Silver Spring, MD: NOAA Atlas NESDIS), 81.

Lorenz, M. G., and Wackernagel, W. (1987). Adsorption of DNA to sand and variable degradation rates of adsorbed DNA. *Appl. Environ. Microbiol.* 53, 2948–2952. doi: 10.1128/aem.53.12.2948-2952.1987

Machado, G. M. V., Bastos, A. C., Da Silva De Freitas, A., and Baptista Neto, J. A. (2018). Sedimentary, geochemical and micropaleontological responses to sea level variations in the Vitoria Estuary, Espírito Santo. *Radiocarbon* 60, 583–600. doi: 10.1017/RDC.2018.5

Marino, M., Rodrigues, T., Quivelli, O., Girone, A., Maiorano, P., and Bassinot, F. (2022). Paleoproductivity proxies and alkenone precursors in the Western Mediterranean during the early-middle Pleistocene transition. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 601, 111104. doi: 10.1016/j.palaeo.2022.111104

Mccartin, L. J., Vohsen, S. A., Ambrose, S. W., Layden, M., Mcfadden, C. S., Cordes, E. E., et al. (2022). Temperature controls *e*DNA persistence across physicochemical conditions in seawater. *Environ. Sci. Technol.* 56, 8629–8639. doi: 10.1021/acs.est.2c01672

Mcclymont, E. L., Ford, H. L., Ho, S. L., Tindall, J. C., Haywood, A. M., Alonso-Garcia, M., et al. (2020). Lessons from a high-CO2 world: an ocean view from ∼3 million years ago. *Climate Past* 16, 1599–1615. doi: 10.5194/cp-16-1599-2020

Mcgann, M., and Powell, C. L. (2022). Impact of climate change on mollusks and other invertebrate resources at the Dominican University of California archaeological site (CA-MRN-254), Marin County, California. *Quat. Int.* 628, 64–78. doi: 10.1016/j.quaint.2022.02.030

Meusnier, I., Singer, G. A., Landry, J. F., Hickey, D. A., Hebert, P. D., and Hajibabaei, M. (2008). A universal DNA mini-barcode for biodiversity analysis. *BMC Genom.* 9, 214. doi: 10.1186/1471-2164-9-214

Michelsen, C., Pedersen, M. W., Fernandez-Guerra, A., Zhao, L., Petersen, T. C., and Korneliussen, T. S. (2022). metaDMG – a fast and accurate ancient DNA damage toolkit for metagenomic data. *bioRxiv* 2022, 2012. doi: 10.1101/2022.12.06.519264

Miettinen, A. (2018). Diatoms in Arctic regions: potential tools to decipher environmental changes. *Polar Sci.* 18, 220–226. doi: 10.1016/j.polar.2018.04.001

Morard, R., Lejzerowicz, F., Darling, K. F., Lecroq-Bennet, B., Winther Pedersen, M., Orlando, L., et al. (2017). Planktonic foraminifera-derived environmental DNA extracted from abyssal sediments preserves patterns of plankton macroecology. *Biogeosciences* 14, 2741–2754. doi: 10.5194/bg-14-2741-2017

More, K. D., Giosan, L., Grice, K., and Coolen, M. J. L. (2019). Holocene paleodepositional changes reflected in the sedimentary microbiome of the Black Sea. *Geobiology* 17, 436–448. doi: 10.1111/gbi.12338

More, K. D., Orsi, W. D., Galy, V., Giosan, L., He, L., Grice, K., et al. (2018). A 43 kyr record of protist communities and their response to oxygen minimum zone variability in the Northeastern Arabian Sea. *Earth Planet. Sci. Lett.* 496, 248–256. doi: 10.1016/j.epsl.2018.05.045

More, K. D., Wuchter, C., Irigoien, X., Tierney, J. E., Giosan, L., Grice, K., et al. (2021). Subseafloor archaea reflect 139 kyrs of paleodepositional changes in the northern Red Sea. *Geobiology* 19, 162–172. doi: 10.1111/gbi.12421

Murchie, T. J., Kuch, M., Duggan, A. T., Ledger, M. L., Roche, K., Klunk, J., et al. (2020). Optimizing extraction and targeted capture of ancient environmental DNA for reconstructing past environments using the PalaeoChip Arctic-1.0 bait-set. *Quat. Res.* 99, 305–328. doi: 10.1017/qua.2020.59

Murrell, M. C., and Lores, E. M. (2004). Phytoplankton and zooplankton seasonal dynamics in a subtropical estuary: importance of cyanobacteria. *J. Plankton Res.* 26, 371–382. doi: 10.1093/plankt/fbh038

Neuenschwander, S., Cruz Dávalos, D. I., Anchieri, L., Sousa Da Mota, B., Bozzi, D., Rubinacci, S., et al. (2023). Mapache: a flexible pipeline to map ancient DNA. *Bioinformatics* 39, btad028. doi: 10.1093/bioinformatics/btad028

Nooteboom, P. D., Bijl, P. K., Van Sebille, E., Von Der Heydt, A. S., and Dijkstra, H. A. (2019). Transport bias by ocean currents in sedimentary microplankton assemblages: implications for paleoceanographic reconstructions. *Paleoceanogr. Paleoclimatology* 34, 1178–1194. doi: 10.1029/2019PA003606

Nwosu, E. C., Brauer, A., Kaiser, J., Horn, F., Wagner, D., and Liebner, S. (2021). Evaluating sedimentary DNA for tracing changes in cyanobacteria dynamics from sediments spanning the last 350 years of Lake Tiefer See, NE Germany. *J. Paleolimnol.* 66, 279–296. doi: 10.1007/s10933-021-00206-9

Ogata, M., Masuda, R., Harino, H., Sakata, M. K., Hatakeyama, M., Yokoyama, K., et al. (2021). Environmental DNA preserved in marine sediment for detecting jellyfish blooms after a tsunami. *Sci. Rep.* 11, 16830. doi: 10.1038/s41598-021-94286-2

Okabe, S., and Shimazu, Y. (2007). Persistence of host-specific bacteroides–prevotella 16S rRNA genetic markers in environmental waters: effects of temperature and salinity. *Appl. Microbiol. Biotechnol.* 76, 935–944. doi: 10.1007/s00253-007-1048-z

Oksman, M., Weckström, K., Miettinen, A., Juggins, S., Divine, D. V., Jackson, R., et al. (2017). Younger Dryas ice margin retreat triggered by ocean surface warming in central-eastern Baffin Bay. *Nat. Commun.* 8, 1017. doi: 10.1038/s41467-017-01155-6

Pados-Dibattista, T., Pearce, C., Detlef, H., Bendtsen, J., and Seidenkrantz, M. S. (2022). Holocene palaeoceanography of the Northeast Greenland shelf. *Clim. Past.* 18, 103–127. doi: 10.5194/cp-18-103-2022

Pawłowska, J., Łącka, M., Kucharska, M., Pawlowski, J., and Zajączkowski, M. (2020a). Multiproxy evidence of the Neoglacial expansion of Atlantic water to eastern Svalbard. *Clim. Past.* 16, 487–501. doi: 10.5194/cp-16-487-2020

Pawłowska, J., Lejzerowicz, F., Esling, P., Szczucinski, W., Zajaczkowski, M., and Pawlowski, J. (2014). Ancient DNA sheds new light on the Svalbard foraminiferal fossil record of the last millennium. *Geobiology* 12, 277–288. doi: 10.1111/gbi.12087

Pawłowska, J., Wollenburg, J. E., Zajączkowski, M., and Pawlowski, J. (2020b). Planktonic foraminifera genomic variations reflect paleoceanographic changes in the Arctic: evidence from sedimentary ancient DNA. *Sci. Rep.* 10, 15102. doi: 10.1038/s41598-020-72146-9

Pawlowski, J., Bonin, A., Boyer, F., Cordier, T., and Taberlet, P. (2021). Environmental DNA for biomonitoring. *Mol. Ecol.* 30, 2931–2936. doi: 10.1111/mec.16023

Pawlowski, J., Bruce, K., Panksep, K., Aguirre, F. I., Amalfitano, S., Apothéloz-Perret-Gentil, L., et al. (2022). Environmental DNA metabarcoding for benthic monitoring: a review of sediment sampling and DNA extraction methods. *Sci. Total Environ.* 818, 151783. doi: 10.1016/j.scitotenv.2021.151783

Pearman, J. K., Keeley, N. B., Wood, S. A., Laroche, O., Zaiko, A., Thomson-Laing, G., et al. (2020). Comparing sediment DNA extraction methods for assessing organic enrichment associated with marine aquaculture. *PeerJ* 8, e10231. doi: 10.7717/peerj.10231

Pedersen, M. W., Overballe-Petersen, S., Ermini, L., Sarkissian, C. D., Haile, J., Hellstrom, M., et al. (2015). Ancient and modern environmental DNA. *Philos. Trans. R. Soc Lond. B Biol. Sci.* 370, 20130383. doi: 10.1098/rstb.2013.0383

Perianes-Rodriguez, A., Waltman, L., and Van Eck, N. J. (2016). Constructing bibliometric networks: a comparison between full and fractional counting. *J. Informetr.* 10, 1178–1195. doi: 10.1016/j.joi.2016.10.006

Pierella Karlusich, J. J., Pelletier, E., Zinger, L., Lombard, F., Zingone, A., Colin, S., et al. (2023). A robust approach to estimate relative phytoplankton cell abundances from metagenomes. *Mol. Ecol. Resour.* 23, 16–40. doi: 10.1111/1755-0998.13592

Pochon, Z., Bergfeldt, N., Kırdök, E., Vicente, M., Naidoo, T., Valk, T. V. D., et al. (2022). aMeta: an accurate and memory-efficient ancient metagenomic profiling workflow. *bioRxiv* 2022, 2010. doi: 10.1101/2022.10.03.510579

Porter, T. M., and Hajibabaei, M. (2020). Putting COI metabarcoding in context: the utility of exact sequence variants (ESVs) in biodiversity analysis. *Front. Ecol. Evol.* 8. doi: 10.3389/fevo.2020.00248

Pratas, D., and Pinho, A. (2018). "Metagenomic composition analysis of sedimentary ancient DNA from the Isle of Wight," in *26th European Signal Processing Conference, EUSIPCO 2018* (Rome, Italy: IEEE).

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Ribeiro, S., Berge, T., Lundholm, N., Andersen, T. J., Abrantes, F., and Ellegaard, M. (2011). Phytoplankton growth after a century of dormancy illuminates past resilience to catastrophic darkness. *Nat. Commun.* 2, 311. doi: 10.1038/ncomms1314

Robert, K. N., Benjamin, G., John, D. M., David, L. S., David, M., Barbara, R., et al. (2005). Sand DNA - a genetic library of life at the water's edge. *Mar. Ecol. Prog. Ser.* 301, 9–22. doi: 10.3354/meps301009

Sanyal, A., Larsson, J., Van Wirdum, F., Andrén, T., Moros, M., Lönn, M., et al. (2022). Not dead yet: diatom resting spores can survive in nature for several millennia. *Am. J. Bot.* 109, 67–82. doi: 10.1002/ajb2.1780

Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J. F., Al-Rasheid, K., Willerslev, E., et al. (2012). Improving ancient DNA read mapping against modern reference genomes. *BMC Genom.* 13, 178. doi: 10.1186/1471-2164-13-178

Selway, C. A., Armbrecht, L., and Thornalley, D. (2022). An outlook for the acquisition of marine sedimentary ancient DNA (*seda*DNA) from North Atlantic Ocean archive material. *Paleoceanogr. Paleoclimatology* 37, e2021PA004372. doi: 10.1029/2021PA004372

Siano, R., Lassudrie, M., Cuzin, P., Briant, N., Loizeau, V., Schmidt, S., et al. (2021). Sediment archives reveal irreversible shifts in plankton communities after World War II and agricultural pollution. *Curr. Biol.* 31, 2682–2689.e2687. doi: 10.1016/j.cub.2021.03.079

Smayda, T. J. (1971). Normal and accelerated sinking of phytoplankton in the sea. *Mar. Geol.* 11, 105–122. doi: 10.1016/0025-3227(71)90070-3

Sunday, J. M., Popovic, I., Palen, W. J., Foreman, M. G. G., and Hart, M. W. (2014). Ocean circulation model predicts high genetic structure observed in a long-lived pelagic developer. *Mol. Ecol.* 23, 5036–5047. doi: 10.1111/mec.12924

Szczuciński, W., Pawłowska, J., Lejzerowicz, F., Nishimura, Y., Kokociński, M., Majewski, W., et al. (2016). Ancient sedimentary DNA reveals past tsunami deposits. *Mar. Geol.* 381, 29–33. doi: 10.1016/j.margeo.2016.08.006

Talas, L., Stivrins, N., Veski, S., Tedersoo, L., and Kisand, V. (2021). Sedimentary ancient DNA (*seda*DNA) reveals fungal diversity and environmental drivers of community changes throughout the Holocene in the present boreal Lake Lielais Svētiņu (Eastern Latvia). *Microorganisms* 9, 719. doi: 10.3390/microorganisms9040719

Thomas, C., Francke, A., Vogel, H., Wagner, B., and Ariztegui, D. (2020). Weak influence of paleoenvironmental conditions on the subsurface biosphere of Lake Ohrid over the last 515 ka. *Microorganisms* 8, 1736. doi: 10.3390/microorganisms8111736

Thomas, S. P., Shanmuganathan, B., Jaiswal, M. K., Kumaresan, A., and Sadasivam, S. K. (2019). Legacy of a Pleistocene bacterial community: patterns in community dynamics through changing ecosystems. *Microbiol. Res.* 226, 65–73. doi: 10.1016/j.micres.2019.06.001

Thomsen, P. F., Kielgast, J., Iversen, L. L., Møller, P. R., Rasmussen, M., and Willerslev, E. (2012). Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PloS One* 7, e41732. doi: 10.1371/journal.pone.0041732

Torti, A., Jørgensen, B. B., and Lever, M. A. (2018). Preservation of microbial DNA in marine sediments: insights from extracellular DNA pools. *Environ. Microbiol.* 20, 4526–4542. doi: 10.1111/1462-2920.14401

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

Vargas, C. D., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., et al. (2015). Eukaryotic plankton diversity in the sunlit ocean. *Science* 348, 1261605. doi: 10.1126/science.1261605

Voldstad, L. H., Alsos, I. G., Farnsworth, W. R., Heintzman, P. D., Håkansson, L., Kjellman, S. E., et al. (2020). A complete Holocene lake sediment ancient DNA record reveals long-standing high Arctic plant diversity hotspot in Northern Svalbard. *Quat. Sci. Rev.* 234, 106207. doi: 10.1016/j.quascirev.2020.106207

Vuillemin, A., Wankel, S. D., Coskun, Ö.K., Magritsch, T., Vargas, S., Estes, E. R., et al. (2019). Archaea dominate oxic subseafloor communities over multimillion-year time scales. *Sci. Adv.* 5, eaaw4108. doi: 10.1126/sciadv.aaw4108

Weyrich, L. S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., Breen, J., et al. (2017). Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. *Nature* 544, 357–361. doi: 10.1038/nature21674

Willerslev, E., Hansen, A. J., and Poinar, H. N. (2004). Isolation of nucleic acids and cultures from fossil ice and permafrost. *Trends Ecol. Evol.* 19, 141–147. doi: 10.1016/j.tree.2003.11.010

Wood, S. A., Biessy, L., Latchford, J. L., Zaiko, A., Von Ammon, U., Audrezet, F., et al. (2020). Release and degradation of environmental DNA and RNA in a marine system. *Sci. Total Environ.* 704, 135314. doi: 10.1016/j.scitotenv.2019.135314

Xue, J., and Feng, Y. (2018). Determination of adsorption and desorption of DNA molecules on freshwater and marine sediments. *J. Appl. Microbiol.* 124, 1480–1492. doi: 10.1111/jam.13739

Zachos, J. C., Dickens, G. R., and Zeebe, R. E. (2008). An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* 451, 279–283. doi: 10.1038/nature06588

Zheng, X., Liu, W., Dai, X., Zhu, Y., Wang, J., Zhu, Y., et al. (2021). Extraordinary diversity of viruses in deep-sea sediments as revealed by metagenomics without prior virion separation. *Environ. Microbiol.* 23, 728–743. doi: 10.1111/1462-2920.15154

Zimmermann, H. H., Stoof-Leichsenring, K. R., Kruse, S., Müller, J., Stein, R., Tiedemann, R., et al. (2020). Changes in the composition of marine and sea-ice diatoms derived from sedimentary ancient DNA of the eastern Fram Strait over the past 30 000 years. *Ocean Sci.* 16, 1017–1032. doi: 10.5194/os-16-1017-2020

Zimmermann, H. H., Stoof-Leichsenring, K. R., Kruse, S., Nürnberg, D., Tiedemann, R., and 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. *Paleoceanogr. Paleoclimatology* 36, e2020PA004091. doi: 10.1029/2020PA004091

VI. FUNDING

This PhD thesis was carried out from 2021 to 2024 in the Laboratory of Environmental DNA, Department of Paleoceanography, IOPAN, within the Tricity Doctoral School, Polish Academy of Sciences. The study was funded by the Norwegian Financial Mechanism for 2014–2021 within the GRIEG Programme, grant no. 2019/34/H/ST10/00682, and the Polish National Science Centre (NCN) within the PRELUDIUM Programme, grant no. 2023/49/N/ST10/01626.

VII. ADDITIONAL INFORMATION FROM THE AUTHOR

Other collaborations / in preparing manuscripts

1. Uncovering the foraminiferal diversity of the Nordic Seas from surface to seafloor. In prep. Ngoc-Loi Nguyen, Joanna Pawłowska, Marek Zajaczkowski , Agnes K. M. Weiner, Tristan Cordier, Stijn De Schepper, Jan Pawłowski.
2. Sea ice conditions impact on Arctic marine eukaryotes community across 160,000 years. In prep. Joanna Pawłowska, Ngoc-Loi Nguyen, Mattia Greco, Tristan Cordier, Jutta Wollenburg, Jens Matthiessen, Jan Pawłowski.
3. Hidden Foraminiferal diversity near the grounding line: sedaDNA from deglacial sediments in the western Ross Sea, Antarctica. In prep. Ewa Demianiuk, Wojciech Majewski, Mateusz Baca, Inès Barrenechea Angeles, Ngoc-Loi Nguyen, John Anderson, Jan Pawłowski.
4. Post-industrial biotic changes near the former whaling station, King Edward Cove, South Georgia. In prep. Wojciech Majewski, Witold Szczuciński, Ngoc-Loi Nguyen, Joanna Pawłowska, Małgorzata Szymczak-Żyła.

Co-author statement

Bergen, 16th January 2024

Dr Agnes Weiner

NORCE–Norwegian Research Centre
Jahnebakken 5
5007 Bergen
Norway

To Whom It May Concern:

Declaration of co-authorship

I hereby declare that I am a co-author of the publications lead by Mr Ngoc-Loi Nguyen, entitled:

Sedimentary ancient DNA: a new paleogenomic tool for reconstructing the history of marine ecosystems. Ngoc-Loi Nguyen, Dhanushka Devendra, Natalia Szymańska, Mattia Greco, Inès Barrenechea Angeles, Agnes Weiner, Jessica Ray, Tristan Cordier, Stijn De Schepper, Jan Pawłowski and Joanna Pawłowska. Front. Mar. Sci., vol. 10, 2023, p. 1185435, DOI: 10.3389/fmars.2023.1185435.

Not all plankton taxa are equally archived on the seafloor: evidence from water column and sediment environmental DNA. Ngoc-Loi Nguyen, Joanna Pawłowska, Marek Zajaczkowski, Agnes Weiner, Tristan Cordier, Danielle Grant, Stijn De Schepper, Jan Pawłowski. Submitted to Comm. Biol., 2024.

My participation in the preparation of these publications consisted of collecting sample material, helping with data interpretation, and commenting on the written text.

With kind regards,

A. Weiner

Agnes Weiner
Senior Researcher
NORCE

UiT The Arctic University of Norway

Inès Barrenechea Angeles, Ph.D
The Arctic University of Norway
Department of Geosciences
9010-Tromsø

DECLARATION OF CO-AUTHORSHIP

I hereby declare that I am a co-author of the publication entitled:

1. *Metabarcoding reveals high diversity of benthic foraminifera linked to water masses circulation at coastal Svalbard*. Ngoc-Loi Nguyen, Joanna Pawłowska, Inès Barrenechea Angeles, Marek Zajaczkowski, Jan Pawłowski. Geobiology, vol. 21, no. 1, 2023, pp. 133-150, DOI: 10.1111/gbi.12530.
2. *Sedimentary ancient DNA: a new paleogenomic tool for reconstructing the history of marine ecosystems*. Ngoc-Loi Nguyen, Dhanushka Devendra, Natalia Szymańska, Mattia Greco, Ines Barrenechea Angeles, Agnes Weiner, Jessica Ray, Tristan Cordier, Stijn De Schepper, Jan Pawłowski and Joanna Pawłowska. Front. Mar. Sci., vol. 10, 2023, p. 1185435, DOI: 10.3389/fmars.2023.1185435.
3. *Assigning the unassigned: a signature-based classification of rDNA metabarcodes reveals new deep-sea diversity*. Ines Barrenechea Angeles, Ngoc-Loi Nguyen, Mattia Greco, Koh Siang Tan, Jan Pawlowski. Under review in PLOS One, 2024.

My participation in the preparation of these publications consisted of writing-original draft (3), writing-reviewing and editing (1,2,3), methodology (3), formal analysis (3), conceptualization (3), data curation (1,3), investigation (3), visualization (3)

Co-author's name and signature Date: 16 January 2024

Ines Barrenechea Angeles

Prof. Jan Pawłowski
Instytut Oceanologii Polskiej Akademii Nauk
ul. Powstańców Warszawy 55
81-712 Sopot

DECLARATION OF CO-AUTHORSHIP

I hereby declare that I am a co-author of the publication entitled:

1. *Metabarcoding reveals high diversity of benthic foraminifera linked to water masses circulation at coastal Svalbard*. Ngoc-Loi Nguyen, Joanna Pawłowska, Inès Barrenechea Angeles, Marek Zajaczkowski and Jan Pawłowski. Geobiology, vol. 21, no. 1, 2023, pp. 133-150, DOI: 10.1111/gbi.12530.
2. *Sedimentary ancient DNA: a new paleogenomic tool for reconstructing the history of marine ecosystems*. Ngoc-Loi Nguyen, Dhanushka Devendra, Natalia Szymańska, Mattia Greco Barrenechea Angeles, Agnes Weiner, Jessica Ray, Tristan Cordier, Stijn De Schepper, Jan Pawłowski and Joanna Pawłowska. Front. Mar. Sci., vol. 10, 2023, p. 1185435, DOI: 10.3389/fmars.2023.1185435.
3. *Assigning the unassigned: a signature-based classification of rDNA metabarcodes reveals new deep-sea diversity*. Ines Barrenechea Angeles, Ngoc-Loi Nguyen, Mattia Greco, Koh Siang Tan, and Jan Pawlowski. Under review in PLOS One, 2024.
4. *Not all plankton taxa are equally archived on the seafloor: evidence from water column and sediment environmental DNA*. Ngoc-Loi Nguyen, Joanna Pawłowska, Marek Zajaczkowski, Agnes Weiner, Tristan Cordier, Danielle Grant, Stijn De Schepper, Jan Pawłowski. Submitted to Comm. Biol., 2024.

My participation in the preparation of this publications consisted of funding acquisition, writing, editing and final revision.

Co-author's name and signature: Date: 15 January 2024

Jan Pawlowski

Jessica Louise Ray, Ph.D.
Aqua Kompetanse AS
Havbruksparken Midt-Norge
Storlavika 7
7770 Flatanger
Norway

DECLARATION OF CO-AUTHORSHIP

I hereby declare that I am a co-author of the publication entitled:

1. *Sedimentary ancient DNA: a new paleogenomic tool for reconstructing the history of marine ecosystems*. Ngoc-Loi Nguyen, Dhanushka Devendra, Natalia Szymańska, Mattia Greco, Inès Barrenechea Angeles, Agnes Weiner, Jessica Louise Ray, Tristan Cordier, Stijn De Schepper, Jan Pawłowski and Joanna Pawłowska. Front. Mar. Sci., vol. 10, 2023, p. 1185435, DOI: 10.3389/fmars.2023.1185435.

My participation in the preparation of these publications consisted of scientific discussion, sampling planning, sample analysis planning, manuscript writing.

Co-author's name and signature — Date

Jessica Louise Ray — 2024-01-16

dr hab. Joanna Pawłowska
Instytut Oceanologii Polskiej Akademii Nauk
ul. Powstańców Warszawy 55
81-712 Sopot

DECLARATION OF CO-AUTHORSHIP

I hereby declare that I am a co-author of the publication entitled:

1. *Metabarcoding reveals high diversity of benthic foraminifera linked to water masses circulation at coastal Svalbard*. Ngoc-Loi Nguyen, Joanna Pawłowska, Inès Barrenechea Angeles, Marek Zajaczkowski and Jan Pawłowski. Geobiology, vol. 21, no. 1, 2023, pp. 133-150, DOI: 10.1111/gbi.12530.
2. *Sedimentary ancient DNA: a new paleogenomic tool for reconstructing the history of marine ecosystems*. Ngoc-Loi Nguyen, Dhanushka Devendra, Natalia Szymańska, Mattia Greco, Inès Barrenechea Angeles, Agnes Weiner, Jessica Ray, Tristan Cordier, Stijn De Schepper, Jan Pawłowski and Joanna Pawłowska. Front. Mar. Sci., vol. 10, 2023, p. 1185435, DOI: 10.3389/fmars.2023.1185435.
3. *Not all plankton taxa are equally archived on the seafloor: evidence from water column and sediment environmental DNA*. Ngoc-Loi Nguyen, Joanna Pawłowska, Marek Zajaczkowski, Agnes Weiner, Tristan Cordier, Danielle Grant, Stijn De Schepper, Jan Pawłowski. Submitted to Comm. Biol., 2024.

My participation in the preparation of the publications mentioned above consisted of co-creation of the study concept, participation in collecting study material, coordination of field and laboratory work, consultations of laboratory protocols, and pre-editing the manuscript texts.

Co-author's name and signature: Date: 15 January 2024

Joanna Pawłowska

Prof. Marek Zajaczkowski
Instytut Oceanologii Polskiej Akademii Nauk
ul. Powstańców Warszawy 55
81-712 Sopot

DECLARATION OF CO-AUTHORSHIP

I hereby declare that I am a co-author of the publication entitled:

1. *Metabarcoding reveals high diversity of benthic foraminifera linked to water masses circulation at coastal Svalbard*. Ngoc-Loi Nguyen, Joanna Pawłowska, Inès Barrenechea Angeles, Marek Zajaczkowski and Jan Pawłowski. Geobiology, vol. 21, no. 1, 2023, pp. 133-150, DOI: 10.1111/gbi.12530.
2. *Not all plankton taxa are equally archived on the seafloor: evidence from water column and sediment environmental DNA*. Ngoc-Loi Nguyen, Joanna Pawłowska, Marek Zajaczkowski, Agnes Weiner, Tristan Cordier, Danielle Grant, Stijn De Schepper, Jan Pawłowski. Submitted to Comm. Biol., 2024.

My participation in the preparation of this publications consisted of funding acquisition and final revisions.

Co-author's name and signature: Date: 15 January 2024

Marek Zajaczkowski

Dr. Mattia Greco,
ORCID: https://orcid.org/0000-0003-2416-6235
Institut de Ciències del Mar - CSIC,
Passeig Marítim de la Barceloneta, 37-49
ES08003, Barcelona, Spain

DECLARATION OF CO-AUTHORSHIP

I hereby declare that I am a co-author of the publication entitled:

1. *Sedimentary ancient DNA: a new paleogenomic tool for reconstructing the history of marine ecosystems*. Ngoc-Loi Nguyen, Dhanushka Devendra, Natalia Szymańska, Mattia Greco, Inès Barrenechea Angeles, Agnes Weiner, Jessica Ray, Tristan Cordier, Stijn De Schepper, Jan Pawłowski and Joanna Pawłowska. Front. Mar. Sci., vol. 10, 2023, p. 1185435, DOI: 10.3389/fmars.2023.1185435.

2. *Assigning the unassigned: a signature-based classification of rDNA metabarcodes reveals new deep-sea diversity*. Inés Barrenechea Angeles, Ngoc-Loi Nguyen, Mattia Greco, Koh Siang Tan, Jan Pawlowski. Under review in PLOS One, 2024.

My participation in the preparation of these publications consisted of the following contributions:

For the manuscript *Sedimentary ancient DNA: a new paleogenomic tool for reconstructing the history of marine ecosystems*: Data Curation, Visualization, and Review & Editing of the original draft.

For the manuscript *Assigning the unassigned: a signature-based classification of rDNA metabarcodes reveals new deep-sea diversity*: Data Analysis, Visualization, and Review & Editing of the original draft.

Co-author's name and signature Date

MATTIA GRECO 22/01/2024

Natalia Szymańska, Ph.D.
Instytut Oceanologii Polskiej Akademii Nauk
ul. Powstańców Warszawy 55
81-712 Sopot

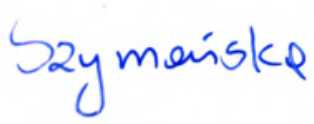
DECLARATION OF CO-AUTHORSHIP

I hereby declare that I am a co-author of the publication entitled:

1. *Sedimentary ancient DNA: a new paleogenomic tool for reconstructing the history of marine ecosystems*. Ngoc-Loi Nguyen, Dhanushka Devendra, Natalia Szymańska, Mattia Greco, Inès Barrenechea Angeles, Agnes Weiner, Jessica Ray, Tristan Cordier, Stijn De Schepper, Jan Pawłowski and Joanna Pawłowska. Front. Mar. Sci., vol. 10, 2023, p. 1185435, DOI: 10.3389/fmars.2023.1185435.

My participation in the preparation of this publications consisted of revisions of the manuscript.

Co-author's name and signature: Date: 15.01.2024



Natalia Szymańska

Bergen, 15 January 2024

Dr Stijn De Schepper

Norwegian Research Centre
Jahnebakken 5
5008 Bergen
Norway

To Whom It May Concern:

Declaration of co-authorship

I hereby declare that I am a co-author of the publications lead by Mr Ngoc-Loi Nguyen, entitled:

Sedimentary ancient DNA: a new paleogenomic tool for reconstructing the history of marine ecosystems. Ngoc-Loi Nguyen, Dhanushka Devendra, Natalia Szymańska, Mattia Greco, Inès Barrenechea Angeles, Agnes Weiner, Jessica Ray, Tristan Cordier, Stijn De Schepper, Jan Pawłowski and Joanna Pawłowska. Front. Mar. Sci., vol. 10, 2023, p. 1185435, DOI: 10.3389/fmars.2023.1185435.

Not all plankton taxa are equally archived on the seafloor: evidence from water column and sediment environmental DNA. Ngoc-Loi Nguyen, Joanna Pawłowska, Marek Zajaczkowski, Agnes Weiner, Tristan Cordier, Danielle Grant, Stijn De Schepper, Jan Pawłowski. Submitted to Comm. Biol., 2024.

My participation in the preparation of these publications consisted of collecting sample material, helping with data interpretation and commenting on the written text.

With kind regards,

Stijn De Schepper
Research Professor
NORCE

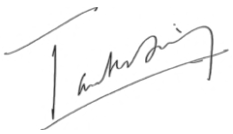
Koh Siang Tan, Ph.D
Tropical Marine Science Institute
National University of Singapore
18 Kent Ridge Road, Singapore 119227.

DECLARATION OF CO-AUTHORSHIP

I hereby declare that I am a co-author of the publication entitled:

1. *Assigning the unassigned: a signature-based classification of rDNA metabarcodes reveals new deep-sea diversity*. Ines Barrenechea Angeles, Ngoc-Loi Nguyen, Mattia Greco, Koh Siang Tan, Jan Pawlowski. Under review in PLOS One, 2024.

My participation in the preparation of this publication consisted of acquisition of data and review of draft.



Co-author's name and signature: Tan Koh Siang Date: 15 January 2024

Tristan Cordier
Senior researcher

+47 56 10 75 13
trco@norceresearch.no

DECLARATION OF CO-AUTHORSHIP

I hereby declare that I am a co-author of the publication entitled:

1. *Sedimentary ancient DNA: a new paleogenomic tool for reconstructing the history of marine ecosystems*. Ngoc-Loi Nguyen, Dhanushka Devendra, Natalia Szymańska, Mattia Greco, Inès Barrenechea, Agnes Weiner, Jessica Ray, Tristan Cordier, Stijn De Schepper, Jan Pawłowski and Joanna Pawłowska. Front. Mar. Sci., vol. 10, 2023, p. 1185435, DOI: 10.3389/fmars.2023.1185435

2. *Not all plankton taxa are equally archived on the seafloor: evidence from water column and sediment environmental DNA*. Ngoc-Loi Nguyen, Joanna Pawłowska, Marek Zajaczkowski, Agnes Weiner, Tristan Cordier, Danielle Grant, Stijn De Schepper and Jan Pawłowski. Submitted to Commun. Biol., 2024.

My participation in the preparation of these publications consisted of advising with the data analysis (statistical, bioinformatics), and by editing and giving feedback on the above-mentioned manuscripts.

Tristan Cordier Bergen, 15th of January 2024