ASSESSMENT OF EUKARYOTIC DIVERSITY AROUND ROBERT ISLAND IN ANTARCTICA BY METABARCODING

by

Yağmur Tarhana

B.Sc. in Molecular Biology and Genetics, Boğaziçi University, 2019

Submitted to the Institute of Environmental Sciences
in partial fulfillment of the requirements for the degree of
Master of Sciences
in
Environmental Sciences

Boğaziçi University
2022

ACKNOWLEDGEMENTS

Foremost, I would like to thank my advisor Prof. Raşit Bilgin for his endless support. I am forever grateful for the encouragement and invaluable insight during my study. He is an excellent mentor who has been supportive throughout the whole thesis processes.

I would like to thank the committee members Assist. Prof. Berat Haznedaroğlu and Assoc. Prof. İsmail K. Sağlam for their time and valuable feedback.

My deepest thanks to Kübra Karaman and Elif İçağasıoğlu for their contributions to the project, also I would like to thank to Ceylan Yücel, and Aygün Karaçay for their help in the laboratory processes.

I am thankful to my beloved friend Umut Çakır for providing information about data analysing. I have learned a lot from our discussions.

Most notably, I would like to thank my family and my cat, Minik, for their emotional support throughout all these years.

The numerical calculations reported in this paper were partially performed at TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources). I would like to thank to TRUBA for offering the opportunity to analyse a huge amount of data.

This study was carried under the auspices of Turkish Republic Presidency, supported by the Ministry of Science, Industry and Technology, and coordinated by Istanbul Technical University (ITU) Polar Research Center (PolReC).

ABSTRACT

ASSESSMENT OF EUKARYOTIC DIVERSITY AROUND ROBERT ISLAND IN ANTARCTICA BY METABARCODING

The rapid growth of human population and overuse of natural resources lead to the deterioration of the balance of ecosystems. Disruption of the environment adversely affects biodiversity and can even cause species extinctions. Since the traditional species detection methods can be expensive and time-consuming, a new, cost-effective, and environmental-friendly method has emerged as an alternative. The new environmental-DNA (eDNA) metabarcoding method is based on DNA isolation, amplification of barcode region, and sequencing of samples taken from environmental systems followed by data analysis for taxonomic assignment. In this study, water samples were collected around Robert Island in Antarctica as a part of the second Turkish Antarctic Expedition (TAE-II) in 2018 to monitor species diversity. Then, mitochondrial cytochrome oxidase I (COI) barcode regions were amplified and sequenced on Illumina platform. Sequences were formatted via OBITools, aligned with BLAST+, and compared with three reference databases (MIDORI, WoRMS, and GenBank) for taxonomic assignment. A total of 39 taxa were found, of which 18 were identified as animal species, but most of the species were algae. Identifying only one taxon in common in the three databases revealed the inconsistency between the reference databases. Additionally, 166 OTUs in MIDORI and 179 OTUs in WoRMS and GenBank databases remained unidentified, indicating that they are deficient in terms of the reference sequence information of taxa found in Antarctica. We conclude that various methods and databases should be used in tandem for biodiversity characterization via metabarcoding in order to increase accuracy and capture species diversity, at least in Antarctica.

ÖZET

ANTARKTİKA'DAKİ ROBERT ADASI ÇEVRESİNDEKİ ÖKARYOT ÇEŞİTLİLİĞİNİN METABARKODLAMA İLE BELİRLENMESİ

Modern dünyada insan nüfusunun hızla artması ve doğal kaynakların aşırı kullanımı ekosistemlerin dengesinin bozulmasına yol açmaktadır. Çevrenin bozulması biyoçeşitliliği olumsuz etkilemekte ve hatta türlerin yok olmasına neden olmaktadır. Tür çeşitliliğinin belirlenmesinde kullanılan geleneksel yöntemler pahalı ve zaman alıcı olabildiğinden, alternatif olarak yeni, uygun maliyetli ve çevre dostu bir yöntem ortaya çıkmıştır. Yeni çevresel DNA (eDNA) metabarkodlama yöntemi, çevresel sistemlerden alınan örneklerin DNA izolasyonu, PCR ile barkod bölgesinin amplifikasyonu, dizilenmesi ve ardından taksonomik atama yapılması adımlarına dayanmaktadır. Bu çalışmada, tür çeşitliliğini tespit etmek amacıyla 2018 yılında 2. Türkiye Ulusal Antarktika Seferi (TAE II) kapsamında Antarktika'daki Robert Adası çevresinden su örnekleri toplanmıştır. Daha sonra mitokondriyal sitokrom oksidaz I (CO1) barkod bölgeleri çoğaltılmış ve Illumina platformu ile dizilenmiştir. Diziler, OBITools aracılığıyla biçimlendirilip, BLAST+ ile hizalanıp, taksonomik atama için üç referans veri tabanı (MIDORI, WoRMS ve GenBank) ile karşılaştırılmıştır. Toplamda 39 takson bulunmuştur, bunların 20 tanesi alg, 18'i hayvan ve 1 tanesi bakteri türü olarak tanımlanmıştır. Üç veri tabanında da ortak olan sadece bir taksonun tanımlanması, farklı referans veri tabanları arasındaki uyuşmazlığı ortaya çıkarmıştır. Ek olarak, MIDORI'de 166 OTU, WoRMS ve GenBank veritabanlarında 179 OTU tanımlanamamıştır; bu da veri tabanlarının Antarktika'da bulunan taksonlara ait dizi bilgilerinde eksiklikler olduğunu göstermektedir. Antarktika'daki biyoçeşitliliğin tespit edilmesi için metabarkodlama ile yapılan çalışmalarda hata payını azaltmak için çeşitli yöntemler ve çeşitli veri tabanlarının birlikte kullanılması gerektiği sonucuna vardık.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....iii
ABSTRACT.....iv
ÖZET.....v
TABLE OF CONTENTS.....vi
LIST OF FIGURES.....viii
LIST OF TABLES.....xii
LIST OF SYMBOLS/ABBREVIATIONS.....xiv
1. INTRODUCTION.....1
2. LITERATURE REVIEW.....5
3. MATERIALS AND METHODS.....10
3.1. Sample Collection.....10
3.2. DNA Extraction.....15
3.2.1. Protocol 1: Isolation from the sealed filter.....15
3.2.2. Protocol 2: Isolation from open filter.....16
3.3. PCR Amplification and Gel Electrophoresis.....17
3.4. PCR Purification, Pooling and Sequencing.....18
3.5. Data Analyses.....19
4. RESULTS AND DISCUSSION.....20
4.1. Amplification with PCR.....20
4.2. Analyses of Sequences.....21
4.3. Species Identification by MIDORI Database.....23
4.4. Species Identification by WoRMS Database.....30
4.5. Species Identification by GenBank Database.....41
4.6. Comparison of Species Identification in MIDORI, WoRMS and GenBank.....51
5. LIMITATIONS OF THE STUDY.....59
6. CONCLUSIONS AND RECOMMENDATIONS.....60
REFERENCES.....62

LIST OF FIGURES

Figure 2.1. The distribution of the databases used in the 30 articles (at least two of these databases were used in 13 articles.) 6

Figure 2.2. The distribution of the marine environments from which the samples were collected by the number of articles. The samples in five articles were collected from more than one sea.....7

Figure 2.3. The simple bar graph for the distribution of the most observed phyla by the number of articles.....8

Figure 3.1. The map shows the location of Robert Island (red square) in Antarctica.....12

Figure 3.2. The locations of Robert Island, Deception Island, and Nansen Island in Antarctic Peninsula.....13

Figure 3.3. The locations where the samples were collected around Robert Island.....13

Figure 4.1. The gel image shows the results of PCR amplification of the water samples filtered from coastal regions of Robert Island. AR 6.2 is the negative control.....20

Figure 4.2. The gel image of PCR amplifications of the water samples filtered from Deception Island, Nansen Island, and offshore of Robert Island.....21

Figure 4.3. The pie chart shows the number of reads of sequences corresponding to each species.....25

Figure 4.4. The numbers of, genera, families, orders, classes, phyla, and kingdoms identified in the three locations.....27

Figure 4.5. The numbers of species identified in the three locations as a bar chart.....27

Figure 4.6. The dendrogram shows the similarity of the species diversity found in the filters according to the results obtained from MIDORI database.....28

Figure 4.7. The dendrogram displays the similarity of the species diversity found in the PCR replicates according to the results obtained from MIDORI database.....29

Figure 4.8. The pie chart shows the number of reads of sequences corresponding to each species.....32

Figure 4.9. The numbers of genera, families, orders, classes, phyla, and kingdoms identified in the five locations.....37

Figure 4.10. The numbers of species identified in the five locations.....37

Figure 4.11. The number of locations where each species found.....38

Figure 4.12. The dendrogram shows the similarity of the species diversity found in the filters according to the results obtained from WoRMS database.....39

Figure 4.13. The dendrogram displays the similarity of the species diversity found in the PCR replicates according to the results obtained from WoRMS database.....40

Figure 4.14. The pie chart shows the number of reads of sequences corresponding to each species identified in GenBank.....42

Figure 4.15. The pie chart shows the number of reads of sequences corresponding to each genus identified in GenBank.....43

Figure 4.16. The pie chart shows the number of reads of sequences corresponding to each family identified in GenBank.....44

Figure 4.17. The pie chart shows the number of reads of sequences corresponding to each order detected in GenBank.....45

Figure 4.18. The pie chart shows the number of reads of sequences corresponding to each class found in GenBank.....46

Figure 4.19. The pie chart shows the number of reads of sequences corresponding to each phylum identified in GenBank.....47

Figure 4.20. The pie chart shows the number of reads of sequences corresponding to each kingdom identified in GenBank.....48

Figure 4.21. The column graph represents the number of unique taxa identified in each location based on the GenBank database.....48

Figure 4.22. The dendrogram indicates the similarity of the species diversity found in the different filters according to the results obtained from GenBank.....49

Figure 4.23. The dendrogram indicates the similarity of the species diversity found in the different PCR replicates according to the results obtained from GenBank.....50

Figure 4.24. The Venn diagram shows the distribution of the taxa identified in the total dataset according to the three databases.....51

Figure 4.25. The stacked column graph demonstrates the read counts of species found in MIDORI, WoRMS, and GenBank databases.....52

Figure 4.26. The stacked column graph displays the presence of taxa found in MIDORI, WoRMS, and GenBank databases.....53

Figure 4.27. The column graph shows the number of different taxa contained in the phylum detected, according to the data obtained from the three databases: MIDORI, WoRMS, and GenBank.....54

Figure 4.28. The pie chart displaying the number of sequence reads of each phylum found in the three databases; MIDORI, WoRMS, and GenBank.....54

LIST OF TABLES

Table 3.1. Sampling dates and locations (longitudes and latitudes) 11

Table 3.2. Amount of filtered water, filtration place and equipment information.....14

Table 3.3. Forward and reverse primers that were used for the first PCR amplification.....17

Table 4.1. The name of the islands and locations where the samples were collected with their filter and PCR replicate's IDs.....22

Table 4.2. Taxonomic lineages of unique taxa found in all locations were identified using MIDORI database.....24

Table 4.3. Taxonomic lineage of unique taxon found in Telephone Cove on Deception Island.....25

Table 4.4. Taxonomic lineage of unique taxa found in Carlota Cove on Robert Island.....26

Table 4.5. Taxonomic lineage of unique taxa found in Coppermine Cove on Robert Island.....26

Table 4.6. Taxonomic lineage of unique taxa obtained from all water samples collected from five regions (Coppermine Cove, Carlota Cove, Telephone Cove, Turkish Point and Offshore) on three islands (Robert Island, Deception Island and Nansen Island) were identified using WoRMS database.....31

Table 4.7. Taxonomic lineage of unique taxa found in Telephone Cove on Deception Island.....33

Table 4.8. Taxonomic lineage of unique taxa found in Turkish Point on Nansen Island.....33

Table 4.9. Taxonomic lineage of unique taxa found in Carlota Cove on Robert Island.....34

Table 4.10. Taxonomic lineage of unique taxa found in Coppermine Cove on Robert Island.....35

Table 4.11. Taxonomic lineage of unique taxa found in the offshore around Robert Island.....36

Table 4.12. The species were identified in only one region.....38

Table 4.13. The table shows whether the 39 species identified in this study were previously recorded in Antarctica.....57

LIST OF SYMBOLS/ABBREVIATIONS

Symbol	Explanation
µl	Microliter
µM	Micromolar
#	Number
$MgCl_2$	Magnesium Chloride
°C	Celsius

Abbreviation	Explanation
A	Adenine
BOLD	Barcode of Life Data System
bp	Base pair
C	Cytosine
CO1	Cytochrome Oxidase I
DNA	Deoxyribonucleic Acid
eDNA	Environmental Deoxyribonucleic Acid
EMBL	European Molecular Biology Laboratory
eRNA	Environmental Ribonucleic Acid
G	Guanine
Gen.	Genus
HPC	High Performance Computing
ml	Milliliter
N/A	Not available
NC	Negative Control
NCBI	National Center for Biotechnology Information
ng	Nanogram
NGS	Next Generation Sequencing
OTU	Operational Taxonomic Unit
PCR	Polymerase Chain Reaction
rRNA	Ribosomal Ribonucleic Acid
SD	Standard Deviation
sp.	Species
T	Thymine

TAE II	Turkish Antarctic Expedition II

1. INTRODUCTION

Biodiversity can be expressed as the heterogeneity of genes, species, and communities in ecosystems, and this diversity is shaped by various geographical, environmental, and biological factors (Cardinale et al., 2012). It is essential for the ability of species to cope with stress and for the sustainability of ecosystems, and also it is important economically and socially as it provides services to humans such as food, shelter, medicine, and water purification (Rafferty 2019). Hence, the conservation of biological diversity is crucial for preserving ecosystem productivity and functioning. Climate change is one of the current environmental problems that destroy natural habitats, with negative impacts on biological diversity now and in the future (Bellard et al., 2012). Scientists predict that species adaptability to the environment will decrease, and species distributions will alter dramatically, with some species having had to migrate and some even going extinct in the future due to climate change (Bellard et al., 2012).

A critical step in understanding the current and future impact of climate change on ecosystems, and hence contributing to prevent and alleviate biodiversity loss due to global warming, is the identification of the biodiversity of a region to form baseline data (Bellard et al., 2012; Blowes et al., 2019). A relatively recently developed alternative method to monitor and characterize biodiversity in a region is the environmental DNA (eDNA) metabarcoding approach, rooted in DNA-based taxonomy analysis (Ruppert et al., 2019). eDNA refers to all DNA molecules released from different organisms to the environment through their skin particles, secretions, germ cells, urine, blood, *etc.*, and these samples can be obtained from various sources such as water, soil, air, honey, or feces (Banchi et al., 2020; Bohmann et al., 2014; Bovo et al., 2018).

eDNA samples contain information about a wide array of species, and characterization of each species is possible with amplifying and sequencing specific gene regions, known as “DNA barcodes” (Taberlet et al., 2012). In DNA barcoding and metabarcoding studies, the mitochondrial genes are often used as barcodes in animals instead of the nuclear genes because mitochondrial genes do not contain introns, evolve rapidly, and are not prone to recombination as they are only inherited maternally (Hebert et al., 2003; Levine et al., 2019). The mitochondrial cytochrome c oxidase I (COI) gene is a standard barcode region used in DNA metabarcoding (Elbrecht & Leese, 2017) due to the presence of universal primers to amplify COI genes and species-associated barcode data that are stored and accessible in comprehensive public databases (e.g. Barcode of Life Database-BOLD).

When analyzing biodiversity in marine ecosystems, it is evident that traditional methods (e.g., visual techniques) are challenging to apply as DNA samples are rapidly degraded or flow away with water (Goldberg et al., 2016; Liu et al., 2019). For all these reasons, the use of the eDNA metabarcoding method in studies of species diversity in marine environments gained popularity as an appropriate method in terms of giving more sensitive results, the requirement of less effort, and cost-effectiveness (Djurhuus et al., 2018; Liu et al., 2019). Besides, considering the marine metabarcoding studies carried out to date, CO1 barcodes have been used in diversity analyzes of very large groups from benthic organisms to higher eukaryotic organisms in various aquatic environments. For example, Leray and Knowlton (2015) analyzed benthic communities around Atlantic Coast, and Collins et al. (2019) surveyed the diversity of freshwater fishes in the North Sea. Considering all the eDNA metabarcoding studies performed, it can be seen that the most studied water is the North Atlantic. Moreover, Arthropoda, Annelida, Mollusca, and Chordata are the most identified phyla by scientists when CO1 was used as a barcode.

After amplification of CO1 barcodes with PCR, the amplicons are sequenced with next-generation sequencing (NGS) technology, which generates a substantial amount of data that are difficult to store and analyze without using bioinformatics tools (Yoccoz, 2012). There is no single data analysis method or a specific computational tool in metabarcoding studies, and scientists have developed different approaches to interpret NGS results. OBITools, a freely available software package consisting of python scripts, has emerged as an effective tool used in eDNA metabarcoding studies for data management and manipulation (Boyer et al., 2016; Yoccoz, 2012). This tool can also be used to perform *in silico* PCR, a method to find the best primers for a gene to amplify with PCR. OBITools can read and process different file formats (e.g., FASTA, FASTQ) and was developed to perform these processes, particularly by taking taxonomic annotation into consideration (Boyer et al. 2016; Coissac et al. 2012). Although OBITools is very useful for cleaning and sorting data, it has some limitations for further analysis because the reference database and taxonomy need to be downloaded and require a large amount of disk space (Boyer et al. 2016; Coissac et al. 2012; Valsecchi et al. 2020).

For an extended period, the classical view on the biodiversity of Antarctica was that the continent exhibited low levels of species diversity due to its harsh conditions. This view might indeed be the case in the Southern Ocean for some groups (such as teleost fishes, gastropods, bivalves, and certain crustaceans); the original work on these groups was considered to represent the entire continent in a simplistic projection (Chown et al., 2015). However, more recent research suggests that the pattern of decreasing diversity as one goes to the poles from Ecuador applies more to the North Pole rather

than the South Pole (Clarke, 2008). For instance, Brandt et al. (Brandt et al., 2007) reported 675 isopod species from the Weddell Sea, of which more than 80% were newly discovered. DeBroyer et al (De Broyer et al., 2014) showed that certain groups such as hexacorals except certain scleractinians, pycnogonids, and bryozoans show diversity levels compared to those in temperate and tropical habitats, only secondary to that of coral reefs. The same study also showed high levels of endemicity, up to 50-97% in certain groups (e.g., sponges, tube worms, amphipods, isopods, sea spiders, and notothenioid fishes). These high levels of endemism were attributed to complicated geographical processes such as glaciation and isolation that interacted together to shape the Southern Ocean biota based on climate projections, the Antarctic Peninsula its surroundings, including the South Shetland Islands, is one of the regions on the earth expected to be most drastically affected by climate change in the 2nd half of the 21st century (Siegert et al., 2019). The Antarctic ecosystem has a quite fragile environmental structure, and especially the Antarctic Peninsula is highly affected by increased human activities and global warming compared to the other parts of the planet (Vaughan et al., 2003). Many species living in the region cannot adapt to these rapidly changing climatic conditions due to their long life cycles. The fact that the Southern Ocean is largely isolated from other marine environments by the Antarctic Circumpolar Current makes it difficult for living creatures to migrate to the environments with more favorable conditions (Peck & Welch, 2004). The current status of the marine and terrestrial biodiversity of this region, whose climate is changing day by day, needs to be better known, and this information should be presented clearly especially to decision makers who shape policies to protect the continent (Grant & Linse, 2009).

In the last decade, the approach of metabarcoding of environmental DNA (eDNA) gained traction and popularity in describing and monitoring biodiversity. The approach has also been used in exploring Antarctic bacterial and eukaryotic biodiversity. Considering the eukaryotes, the groups explored included microscopic species such as protists (Zoccarato et al., 2016), algae (Câmara et al., 2021; Davey et al., 2019; Soto et al., 2020), fungi (Canini et al., 2020; Coleine et al., 2018, 2020; Lacerda et al., 2020; Ogaki et al., 2020; Rosa et al., 2020), and invertebrates (Czechowski et al., 2017). In addition, broad biodiversity surveys for eukaryotes, as well as prokaryotes, were carried out in certain studies (Fraser et al., 2018; Rippin et al., 2018). In the studies above, eDNA samples were collected from various environments and surfaces, including soil crusts from geothermal sites and coastal sites, algal assemblages, plastic surfaces, sandstone rocks, deep-sea sediments, air, and ice cores.

In the present study, we examined eDNA samples collected from Robert Island in Antarctica during the Turkish Antarctic Expedition II (TAE II) in 2018. The diversity assessment of vertebrate,

invertebrate, and algal organisms in this marine environment was done by the CO1 metabarcoding approach. The main goal of this study is to introduce a new approach for data analysis that we think is more useful than the other bioinformatical approaches. We use the BLAST+ method and OBITools combined with an in-house script. Secondly, we aimed to compare three databases, MIDORI, WoRMS, and GenBank, to demonstrate the deficiencies in CO1 reference databases in Antarctica. Our other aim is to help describe biodiversity in this under-explored area to add to the baseline biodiversity information in the Antarctic peninsula.

2. LITERATURE REVIEW

Within the last decade, eDNA metabarcoding studies gained momentum in biodiversity analysis. There are many different methodological approaches to metabarcoding to assess marine eDNA samples. This literature research was undertaken using the keywords of cytochrome oxidase, marine, and metabarcoding on the Web of Science. As a result, there were 48 article hits, 18 of which were excluded because of not being research papers or not being about eDNA. Based on the remaining 30 papers, I checked the most commonly used methods to assign sequences to a reference database and to process raw metabarcoding data, the geographical location of the sampling, the levels of taxonomic diversity detected, and the locus of choice for DNA metabarcoding, the details of which are discussed below.

The most used method to assign sequences to databases included the GenBank BLAST service from the National Center for Biotechnology Information (NCBI) nucleotide database. This approach was adopted in 20 papers out of 30, corresponding to about 67 percent of the studies reviewed. To illustrate, Von Ammon et al. (2019) have used Blast for taxonomy assignment to investigate the relationship between eDNA and eRNA samples, and Jeunen et al. (2019) have also used BLAST to see the habitat-specificity and dispersal of eDNA. Similarly, Borrell et al. (2017), Cheang et al. (2020), and Holman et al. (2019) have utilized the BLAST method for the detection of invasive, indicator, and indigenous species, respectively.

The other databases used in 30 research papers were Barcode of Life Database (BOLD) (15 papers), World Register of Marine Species (WoRMS) (four papers), and European Molecular Biology (EMBL) (four papers). Two of the databases have been combined in 13 papers. In addition, Von Ammon et al. (2019), Holman et al. (2019), and Wood et al. (2019) did their taxonomic assignments by using the metazoan mitochondrial gene sequence reference database MIDORI (which is a more tightly curated version of GenBank), and Günther et al. (2018) used Senckenberg Barcode reference library within EMBL. The distribution of the databases used in the 30 articles is visualized with a pie chart (Figure 2.1).

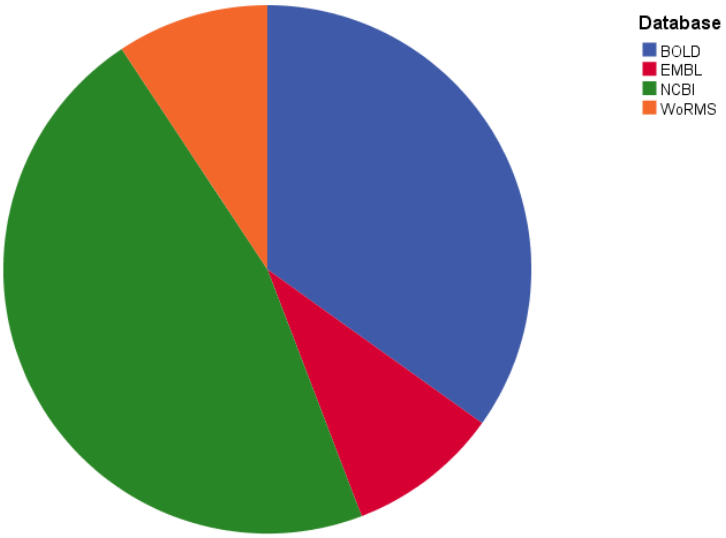


Figure 2.1. The distribution of the databases used in the 30 articles (at least two of these databases were used in 13 articles.)

Another method for taxonomy assignment is *in silico* PCR (or *eco*PCR). *In silico* PCR is a computational procedure to simulate PCR with a given set of primers, checks the primer specificity, and provides the theoretical PCR results, and this method was used in about 20 percent (six papers) of the marine eDNA metabarcoding studies. Antich et al. (2020), Wangensteen et al. (2018), and Günther et al. (2018) used *in silico* PCR to determine the efficiency of metabarcoding in species detection. In two studies, conducted by Collins et al. (2019) and Aglieri et al. (2020), both methods (BLAST and *in silico* PCR) were used to compare traditional surveys with eDNA surveys. The remaining six papers include different Linux-based methods such as Usearch (Edgar, 2010).

As discussed in the introduction, OBITools is one of the software packages used for cleaning and sorting raw metabarcoding data, and besides, OBITools is a valuable tool to perform *in silico* PCR. OBITools uses the *eco*PCR command to check the quality of barcode and *ecotag* command to assign sequences to taxa (Boyer et al., 2016). In five of 30 studies, OBITools was used to perform *eco*PCR on CO1 barcodes. Antich et al. (2020), Turon et al. (2020), and Wangensteen et al. (2018) used *eco*PCR as integrated in OBITools to analyze the taxonomy of various marine benthic communities. Additionally, Aglieri et al. (2020) performed *eco*PCR within OBITools to find taxonomic identities of Mediterranean fishes, and Collins et al. (2019) applied *in silico* PCR using a tool rather than OBITools, namely MFEprimer v2.0 (Qu et al., 2012) for their research on freshwater fishes in the British Isles.

Although environmental DNA studies date back to the late 1980s (Ogram et al. 1987), the approach began to develop and attract attention in the early 2000s (Rondon et al., 2000), and since then millions of different species have been identified from many different marine environments. When the seas from which the samples were taken are compared in the 30 articles reviewed (Figure 2.2), it is seen that the most studied location is the North Atlantic Ocean, where samples were collected from in 16 studies. The North Atlantic Ocean included Florida coasts, Bays of Biscay, and the North Sea. To illustrate, surface water samples from three coral reef sites within South Florida were collected by Sawaya et al. (2019) in 2015 to demonstrate eukaryotic diversity, and by Djurhuus et al. (2018) in 2016 to find zooplankton diversity. eDNA samples in eight studies were collected from the Mediterranean Sea, making it the second most studied marine environment. The Pacific Ocean is also a highly researched water (six papers in Southern Pacific and one paper in Northern Pacific), and the Baltic Sea was surveyed in three articles. One of the surveys was conducted by Atherton and Jondelius (2020) where they sampled marine sediments on Swedish Coasts to identify meiobenthic species diversity. Cahill et al. (2018) and Pearman et al. (2020) used samples collected from three different seas (Baltic Sea, Mediterranean Sea, and North Atlantic Ocean); on the other hand, Dias et al. (2017) collected the samples from all over the world. None of the 30 articles studied the seas around Antarctica, the primary geographic region of interest for this study.

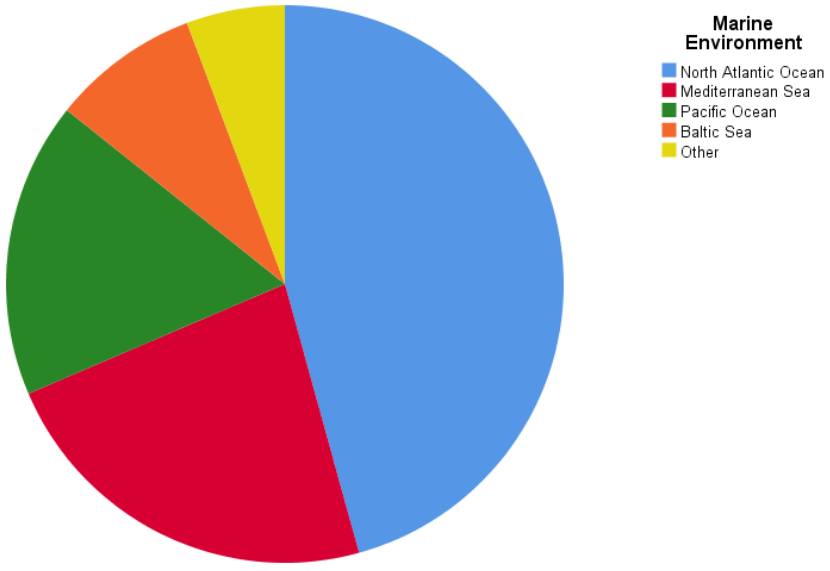


Figure 2.2. The distribution of the marine environments from which the samples were collected by the number of articles. The samples in five articles were collected from more than one sea.

The idea of using the CO1 gene as a DNA barcode in taxonomy assignments was first put forward in 2003 by Hebert et al. In this study, the research group achieved a hundred percent success with CO1 profiles on the taxonomic identification of Lepidoptera species, which is one of the most

diverse animal groups. Since then, many species have been identified using CO1 barcodes in eDNA metabarcoding studies, but most of the studies showed taxonomic assignments at the phylum level. Arthropoda, Mollusca, Annelida, and Chordata are the most found phyla in the 30 studies (Figure 2.3). The most diverse phyla results were represented by Antich et al. (2020), Ransome et al. (2017), and Borrell et al. (2017). Antich et al. (2020) studied benthic species diversity, and they identified 22 metazoan phyla; though some are not the phyla which benthos belong to. Using more specific primers instead of universal CO1 primers can help to overcome this problem (Antich et al., 2020). Ransome et al. (2017) found 28 phyla, but most of them did not show high threshold matches. This result supports the need for a more comprehensive CO1 reference database for more accurate species identification.

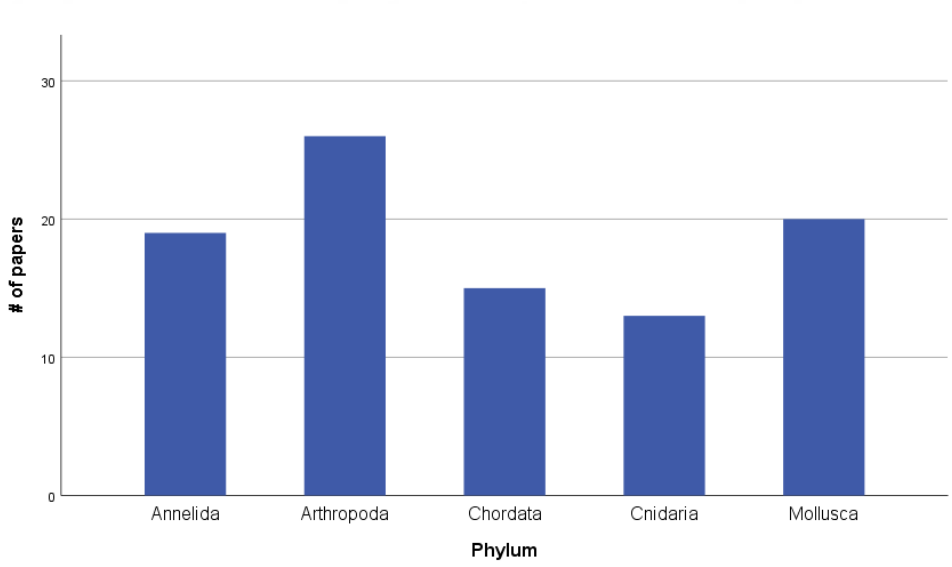


Figure 2.3. The simple bar graph for the distribution of the most observed phyla by the number of articles.

Finally, in 11 studies, details of biodiversity were determined using only the CO1 gene, but there are several barcode genes used other than the CO1 gene, and researchers usually prefer to compare the efficiencies of one or more barcodes in diversity analysis. Nuclear 18S small ribosomal subunit (18S rDNA) gene is the most common DNA barcode used for comparison with CO1. 10 of the 30 papers reviewed used the 18S gene in metabarcoding besides the CO1 gene. The results of the study carried out by Atherton and Jondelius (2020) displayed that the CO1 gene has a limited effect on identifying meiofauna communities against 18S. However, this finding also proves that CO1 is a good barcode for animal species, but not suitable for the non-metazoan species (Hebert et al., 2003). 12S and 16S rDNA genes were also used as DNA barcodes in different studies.

Another approach to determine the efficacy of the eDNA metabarcoding method is to compare the findings with morphological assessments. A comparison between the metabarcoding and morphology-based species identification was made in nine articles. The studies indicate that although both methods are very informative in diversity analysis, metabarcoding method offers more comprehensive taxonomic information, thus, the choice of method depends on the purpose of the study (Steyaert et al. 2020; Schroeder et al. 2020; Djurhuus et al. 2018; Cahill et al. 2018; Pearman et al. 2020; Borrell et al. 2017; Aglieri et al. 2020; Deagle et al. 2018; Jeunen et al. 2019).

3. MATERIALS AND METHODS

3.1. Sample Collection

During the second Turkish Antarctic Expedition (TAE II) in 2018, environmental DNA samples were collected from two coves on Robert Island, which can be reached on foot from the mainland; these were Carlota and Coppermine Coves. In addition, water samples were collected from offshore around Robert Island. eDNA samples were also collected from the islands of Nansen and Deception, which were visited as part of the expedition. In total, water samples were collected from five different regions on three islands. Detailed information about the sample collection is given in Table 3.1, and the exact locations are shown on the map in Figure 3.1, Figure 3.2, and Figure 3.3.

Table 3.1. Sampling dates and locations (longitudes and latitudes).

Sample ID	**Sampling Date**	**Location 1**	**Location 2**	**Longitude**	**Latitude**
AD 1.1	13/04/2018	Deception Island	Telephone Cove	60° 38' 38.64" W	62° 55' 17.56" S
AD 1.2	13/04/2018	Deception Island	Telephone Cove	60° 38' 38.64" W	62° 55' 17.56" S
AN 1	11/04/2018	Nansen Island	Turkish Point	68°07'64.67'' W	64°54' 37.74'' S
AN 2	12/04/2018	Nansen Island	Turkish Point	68°07'64.67'' W	64°54' 37.74'' S
AR 1.1	29/03/2018	Robert Island	Coppermine Cove	59° 42' 11.96" W	62° 22' 46.24" S
AR 1.2	29/03/2018	Robert Island	Coppermine Cove	59° 42' 11.96" W	62° 22' 46.24" S
AR 1.3	29/03/2018	Robert Island	Coppermine Cove	59° 42' 11.96" W	62° 22' 46.24" S
AR 2.1	29/03/2018	Robert Island	Carlota Cove	59° 42' 08.38" W	62° 22' 35.83" S
AR 2.2	29/03/2018	Robert Island	Carlota Cove	59° 42' 08.38" W	62° 22' 35.83" S
AR 3	30/03/2018	Robert Island	Coppermine Cove	59° 41' 41.28" W	62° 22' 45.44" S
AR 4.1	31/03/2018	Robert Island	Carlota Cove (ASPA)	59° 42' 08.38" W	62° 22' 35.83" S
AR 4.2	31/03/2018	Robert Island	Carlota Cove (ASPA)	59° 42' 08.38" W	62° 22' 35.83" S
AR 5.1	03/04/2018	Robert Island	Carlota Cove	59° 42' 08.38" W	62° 22' 35.83" S
AR 5.2	03/04/2018	Robert Island	Carlota Cove	59° 42' 08.38" W	62° 22' 35.83" S
AR 6.1	06/04/2018	Robert Island	Carlota Cove	59° 42' 08.38" W	62° 22' 35.83" S
AR 6.2 (NC)	06/04/2018	Robert Island	Carlota Cove	59° 42' 08.38" W	62° 22' 35.83" S
AR 7	07/04/2018	Robert Island	Coppermine Cove	59° 42' 11.96" W	62° 22' 46.24" S
AR 8	07/04/2018	Robert Island	Carlota Cove	59° 42' 08.38" W	62° 22' 35.83" S
AR 9	08/04/2018	Robert Island	Offshore	59° 42' 25.46" W	62° 23' 25.44" S



Figure 3.1. The map shows the location of Robert Island (red square) in Antarctica.



Figure 3.2. The locations of Robert Island, Deception Island, and Nansen Island in Antarctic Peninsula.

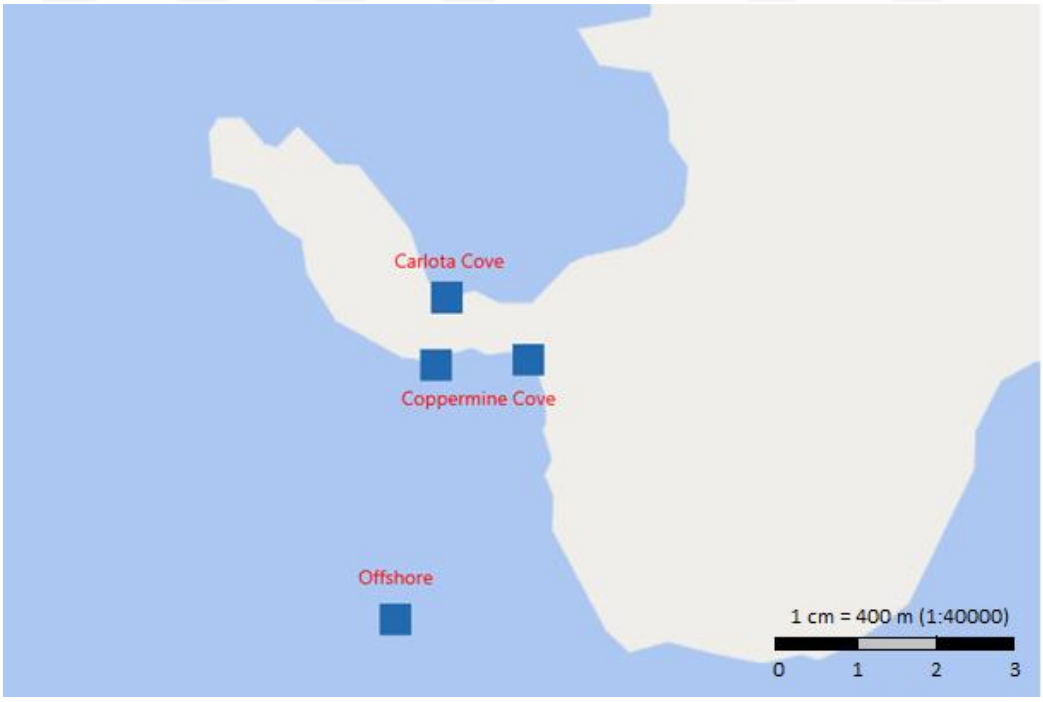


Figure 3.3. The locations where the samples were collected around Robert Island.

eDNA samples were filtered using syringes and sterivex filters of 0.22 μm pore size (Merck, Germany). The water in the syringe was filtered by attaching a sterile sterivex filter to the syringe tip. This process was repeated until the filter was clogged. For this reason, as seen in Table 3.2, different amounts of water could be filtered with each filter. A maximum of 3 liters of water was filtered through the filters. After the filtration process, 2 ml of PBS chemical was passed through the filters. Then the outlet of the filter was closed with parafilm, and the filters were filled with 2 ml of DNA Shield (AMBRD, Turkey). Afterward, the inlet of the filter was closed with parafilm.

Table 3.2. Amount of filtered water, filtration place and equipment information.

Sample ID	**Filtration equipment**	**Water quantity (ml)**	**Filtration site**
AD 1.1	Syringe	2000	On site
AD 1.2	Syringe	2000	On site
AN 1	Syringe	3000	On site
AN 2	Syringe	3000	On ship
AR 1.1	Syringe	1000	On site
AR 1.2	Syringe	1000	On site
AR 1.3	Syringe	1000	On site
AR 2.1	Syringe	800	On site
AR 2.2	Syringe	1300	On site
AR 3	Syringe	1600	On site
AR 4.1	Syringe	700	On site
AR 4.2	Syringe	750	On site
AR 5.1	Syringe	750	On site
AR 5.2	Syringe	600	On site
AR 6.1	Syringe	1200	On camp
AR 6.2 (NC)	Syringe	1000	On camp
AR 7	Syringe	1300	On camp
AR 8	Syringe	850	On camp
AR 9	Syringe	3000	On ship

3.2 DNA Extraction

Qiagen DNeasy Blood & Tissue Kit was used for total DNA extraction. Two different methods were applied.

3.2.1. Protocol 1: Isolation from the sealed filter

1. With the help of a 2 ml syringe, the liquid in the Sterivex filter was transferred to an empty 2 ml Eppendorf tube.
2. The solutions were centrifuged at 6000g (8000 rpm) for 30-45 minutes, and the supernatant was discarded.
3. 180 µL of ATL buffer and 20 µL of proteinase K mixture were added to each sample.
4. After vortexing for 15 seconds, the samples were incubated for 24 hours in a water bath at 56 °C.
5. The Sterivex filter capsule was left to dry in the fume hood with the inlet side down.
6. A mixture of 180 µL of ATL buffer, 20 µL of proteinase K, and 600 µL of distilled water was prepared for each sample.
7. After closing the outlet side cover, 800 µL of the prepared lysis mixture was poured so that it flowed between the outer part of the filter and the side surfaces of the capsule. The hatch on the entrance side was also closed and shaken for a few seconds.
8. Samples were incubated in a water bath at 56 °C for 24 hours. During the incubation, the samples were shaken from time to time.
9. After the end of the incubation, the filters were vigorously shaken again.
10. With the help of a 2 ml syringe, all the liquid in the capsule was withdrawn from the inlet and divided into several 2 ml Eppendorf tubes by knowing their volume. (The volume of liquid added to the filter is 800 ml, but in some samples, a decrease or increase in this liquid was observed after incubation.)
11. The same volume of 99% cold ethanol (molecular grade) and AL buffer was added to the sample.
12. The samples were shaken vigorously.
13. A maximum of 700 µL was taken once and transferred to 2 ml DNeasy Mini Spin Columns.
14. Centrifuged for 1 minute at 4°C and 6000g (8000 rpm).
15. The supernatant was discarded.
16. The 10^{th}, $11^{th,}$ and 12^{th} steps were repeated until the entire volume was filtered.

17. The DNeasy spin column was transferred to new 2 ml tubes, 500 µL of AW1 buffer was added, and centrifuged at 6000g (8000 rpm) for 1 minute.
18. The DNeasy spin column was transferred to new 2 ml tubes, 500 µL of AW2 buffer was added, and the membrane was dried by centrifuging at 20000g (14000 rpm) for 3 minutes. Then the spin columns were transferred to new tubes and centrifuged at 17000g (13000 rpm) for 1 minute.
19. Spin columns were transferred to new 1.5- or 2-ml tubes (with open caps).
20. 50 µL of AE buffer was added to the tubes.
21. Spin columns were incubated at room temperature for 10 minutes.
22. The samples were centrifuged at 6000g (8000 rpm) for 1 minute.
23. Spin columns were discarded.
24. DNA was transferred to clean tubes and stored.

3.2.2. Protocol 2: Isolation from open filter

1. With the help of a 2 ml syringe, the liquid in the Sterivex filter was transferred to an empty 2 ml Eppendorf tube.
2. The solutions were centrifuged at 6000g (8000 rpm) for 30-45 minutes and the supernatant was discarded.
3. 180 µL of ATL buffer and 20 µL of proteinase K mixture were added to each sample.
4. After vortexing for 15 seconds, the samples were incubated for 24 hours at 56 °C 400 rpm in a shaking heat block.
5. The filter paper was cut into small pieces and placed in 2 ml Eppendorf tubes.
6. A mixture of 360 µL of ATL buffer, 40 µL of proteinase K and 600 µL of distilled water was added to each tube.
7. The samples were incubated for 24 hours at 56 °C 400 rpm in a shaking heat block.
8. After incubation, the samples were divided into 2 ml Eppendorf tubes in known volumes.
9. The same volume of 99% cold ethanol (molecular grade) and AL buffer was added to the samples.
10. The samples were shaken vigorously.
11. A maximum of 700 µL was taken at once and transferred to 2 ml DNeasy Mini Spin Columns.
12. Centrifuged for 1 minute at 4°C and 6000g (8000 rpm).
13. The supernatant was discarded.
14. The 10^{th}, 11^{th} and 12^{th} steps were repeated until the entire volume was filtered.

15. The DNeasy spin column was transferred to new 2 ml tubes, 500 µL of AW1 buffer was added, and centrifuged at 6000g (8000 rpm) for 1 minute.
16. The DNeasy spin column was transferred to new 2 ml tubes, 500 µL of AW2 buffer was added, and the membrane was dried by centrifuging at 20000g (14000 rpm) for 3 minutes. Then the spin columns were transferred to new tubes and centrifuged at 17000g (13000 rpm) for 1 minute.
17. Spin columns were transferred to new 1.5- or 2-ml tubes (with open caps).
18. 50 µL of AE buffer were added to the tubes.
19. Spin columns were incubated at room temperature for 10 minutes.
20. The samples were centrifuged at 6000g (8000 rpm) for 1 minute.
21. Spin columns were discarded.
22. DNA was transferred to clean tubes and stored.

3.3. PCR Amplification and Gel Electrophoresis

CO1 barcode regions on the isolated DNA samples were amplified with amplicon-specific primers. The names and sequences of forward and reverse primers are shown on Table 3.

Table 3.3. Forward and reverse primers that were used for the first PCR

Primer Name	**Sequences**	**Primers with Illumina overhangs**
mlCOIintF (Forward)	GGWACWGGWTGAACW GTWTAYCCYCC	TCGTCGGCAGCGTCAGATGTGTATAAGA GACAGGG**W**AC**W**GG**W**TGAAC**W**GT**W**TA **Y**CC**Y**CC
dgHCO-2198 (Reverse)	TAAACTTCAGGGTGAC CAAARAAYCA	TAAACTTCAGGGTGACCAAA**R**AA**Y**CAG TCTCGTGGGCTCGGAGATGTGTATAAGA GACAG

Each PCR tube had a 25 µl reaction volume which contained 12.5 µl 2X PCR ready-mix (AMBRD, Turkey), 0.6 µl forward primer (10 µM), 0.6 µl reverse primer (10 µM), 7.8 µl distilled water, 0.5 µl $MgCl_2$ and 3 µl DNA sample. PCR cycles were 30 seconds at 98 °C, 35 cycles of 10 seconds at 98 °C, 30 seconds at 48 °C, 45 seconds at 72 °C followed by a final extension of 5 minutes at 72 °C. Then, gel electrophoresis procedure was applied to observe the presence of DNA in PCR product, using negative and positive controls. Remaining PCR products were used for a second PCR amplification. In this case, different index primers were used, therefore each PCR product had a unique combination of forward and reverse primers. Similar to the first PCR, each PCR-2 tube had a

25 µl reaction volume in the second PCR. Each tube comprised of 12.5 µl 2X PCR ready-mix (AMBRD, Turkey), 0.1 µl i7 primer (10 µM), 0.1 µl i5 primer (10 µM), 8 µl distilled water, and 2.5 µl of PCR-1 product. Again, gel electrophoresis procedure was applied to check the presence of DNA using a 1% agarose gel.

3.4. PCR Purification, Pooling and Sequencing

The second PCR products were purified with magnetic bead purification method. There were 19 µl PCR products in each tube. The procedure is listed below.

1. Each PCR product was placed into a clean tube.
2. The Agencourt AMPure XP bottle was shaken and 34.2 µl of the Agencourt AMPure XP was added to each sample.
   The amount of Agencourt AMPure XP was calculated as the following equation:
   Volume of Agencourt AMPure XP per reaction = 1.8 x (Reaction Volume)
3. The solutions were mixed by pipette mixing 10 times and incubated for five minutes at room temperature.
4. Tubes were placed on a magnetic rack for two minutes.
5. The supernatants were aspirated from the tubes without touching the beads.
6. 200 µl of 70% ethanol was added to each tube and the tubes were incubated for 30 seconds at room temperature. Then, ethanol was aspirated and discarded. This step was repeated for two times.
7. The tubes were removed from the magnetic rack.
8. 40 µl of elution buffer was added to each tube and incubated for two minutes at room temperature.
9. The tubes were placed on the magnetic rack and incubated for one minute at room temperature.
10. The supernatants were taken and stored into a clean tube.

The concentration of DNA in each sample was measured with Qubit fluorometer. Firstly, a working solution was prepared mixing 199 µl of Qubit buffer and 1 µl of Qubit reagent. Then, 199 µl of working solution was mixed with 1 µl of sample in a Qubit tube and measured. The samples were diluted according to their DNA concentration and pooled in a sequencing tube. The pooled DNA products were sent to Gen-Era for commercial sequencing on the Illumina Miseq Platform, generating 20 million of 300 bp paired-end reads.

3.5. Data Analyses

All computational analyses were conducted using High Performance Computing System (HPC) provided by the Turkish National e-Science e-Infrastructure (TRUBA).

DNA sequences were trimmed with Trimmomatic 0.39 (Bolger et al., 2014). The first and last 26 nucleotides of each sequence were cropped to remove Illumina adaptors. The quality of sequences was checked with FastQC (Andrews, 2010).

OBITools v1. (Boyer et al., 2016) program was used to filter, clean, and sort the data. A for-loop was created to apply OBITools commands (illuminapairedend, obigrep, obiuniq, obiannotate, and obiclean) to each sequence. The sequences were concatenated to generate a single fasta file. Taxonomy annotations were done by using three different approaches and databases. Firstly, MIDORI database was used for alignment and taxonomic lineages. Secondly, taxonomic lineages of the unique taxa were identified from WoRMS, using an in-house Python script. Thirdly, the same BLAST+ output was used, but the taxonomic lineages were taken from GenBank database. All dendrograms were built using PRIMER-e v7. software to show the hierarchical clustering between filters and PCR replicates based on the databases. The number of reads for each sequence was calculated using obiclean command through OBITools.

4. RESULTS AND DISCUSSION

4.1. Amplification with PCR

25 ng of DNA extracted from each filter were amplified with PCR to check for amplification. It can be seen from Figure 6 and Figure 7, 14 filters out of 19 had proper DNA bands on the gel; AR 2.2, AR 4.2, AR 6.1, AR7, and AR8 did not have any PCR products.

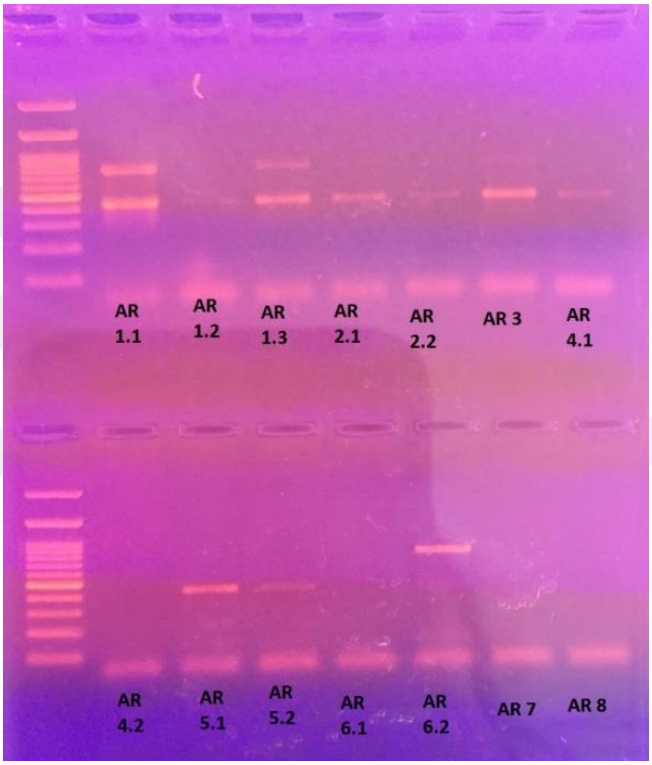


Figure 4.1. The gel image shows the results of PCR amplification of the water samples filtered from coastal regions of Robert Island. AR 6.2 is the negative control.

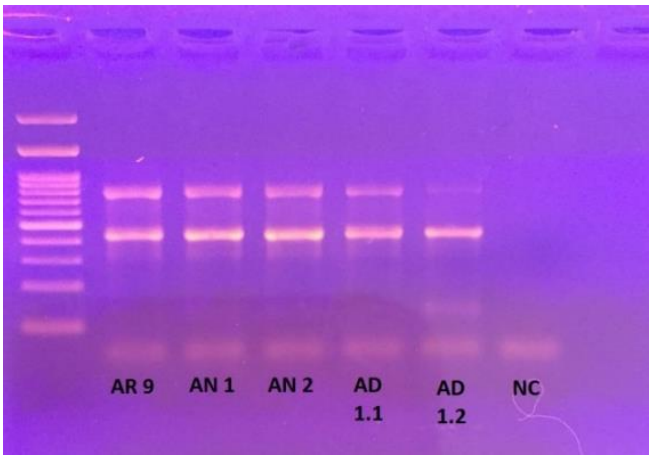


Figure 4.2. The gel image of PCR amplifications of the water samples filtered from Deception Island, Nansen Island, and offshore of Robert Island.

4.2. Analyses of sequences

After conducting the OBITools workflow which includes filtering and editing of files, six PCR replicates did not have any sequences and they were discarded from the data. The filter IDs and the number of replicates is shown in Table 4.1. There were 1269 sequences in the combined fasta file and 317 sequences out of these were unique sequences. The length of the sequences varied between 304 – 352 bp, and the mean of the sequence length was 313 bp (*SD* = 5). The sequence lengths are reliable since the primer set used in this project targets CO1 fragments that are313 bp in length. Additionally, the filter AR 6.2 was the negative control (PCR replicates are 3-4, 3-5, and 3-6) and was produced in Carlota Cove on Robert Island.

Table 4.1. The name of the islands and locations where the samples were collected with their filter and PCR replicate's IDs.

Island	**Location**	**Filter**	**PCR Replicate**
Deception Island	Telephone Cove	AD 1.1	4-7
			4-8
			4-10
		AD 1.2	5-1
			5-2
Nansen Island	Turkish Point	AN 1	4-1
			4-3
		AN 2	4-4
			4-5
			4-6
Robert Island	Carlota Cove	AR 2.1	2-4
			2-5
		AR 4.1	2-7
			2-8
		AR 5.1	2-10
			3-1
		AR 5.2	3-2
			3-3
		AR 6.2 (Negative Control)	3-4
			3-5
			3-6
	Coppermine Cove	AR 1.1	1-1
			1-3
		AR 1.2	1-4
			1-6
		AR 1.3	1-8
			1-10
		AR 3	2-3
	Offshore	AR 9	3-7
			3-8
			3-10

4.3. Species Identification by MIDORI Database

For a second approach, MIDORI database was used for the alignment and taxonomic classifications of the queries obtained from OBITools. RDP Classifier program in MIDORI server (Machida et al., 2017) was employed for the alignment. The queries were aligned with the longest CO1 database (Confidence cut-off = 0.8). Nine taxa were ascertained from MIDORI reference database, but one taxon was discarded from data since it was found in a negative control sample. Species with percent identity that was 97% or higher were taken into consideration for taxonomic assignment. Based on MIDORI database, those nine taxa were identified in three out of five locations (Telephone Cove, Carlota Cove, and Coppermine Cove).

The taxonomic classification of eight taxa is displayed in Table 4.2. There was only one kingdom which is Animalia. Four phyla were found, and these phyla were divided into seven classes, seven orders, nine families, nine genera, and nine species. In addition to this information, the number of reads for the sequences corresponding to each species were calculated and are displayed in Figure 4.3. The highest number of sequence reads belongs to *Halozetes marinus*, a mite, which has 55 (55% of the total reads). *Harpagifer georgianus*, a fish, comprises 17% of the data with 17 sequence reads in MIDORI.

Table 4.2. Taxonomic lineages of unique taxa found in all locations were identified using MIDORI database.

Kingdom	**Phylum**	**Class**	**Order**	**Family**	**Genus**	**Species**	**# of reads**
Animalia	Chordata	Actinopteri	Perciformes	Harpagiferidae	*Harpagifer*	*Harpagifer georgianus*	17
Animalia	Chordata	Actinopteri	Perciformes	Nototheniidae	*Notothenia*	*Notothenia coriiceps*	3
Animalia	Chordata	Chondrichthyes	Myliobatiformes	Myliobatidae	*Myliobatis*	*Myliobatis aquila*	3
Animalia	Arthropoda	Malacostraca	Amphipoda	Pontogeneiidae	*Gondogeneia*	*Gondogeneia antarctica*	3
Animalia	Arthropoda	Maxillopoda	Calanoida	Calanidae	*Ctenocalanus*	*Ctenocalanus citer*	8
Animalia	Arthropoda	Arachnida	Oribatida	Ameronothridae	*Ctenocalanus*	*Halozetes marinus*	55
Animalia	Porifera	Demospongiae	Halichondrida	Halichondriidae	*Halichondria*	*Halichondria panicea*	8
Animalia	Tardigrada	Eutardigrada	Parachela	Hypsibiidae	*Acutuncus*	*Acutuncus antarcticus*	3

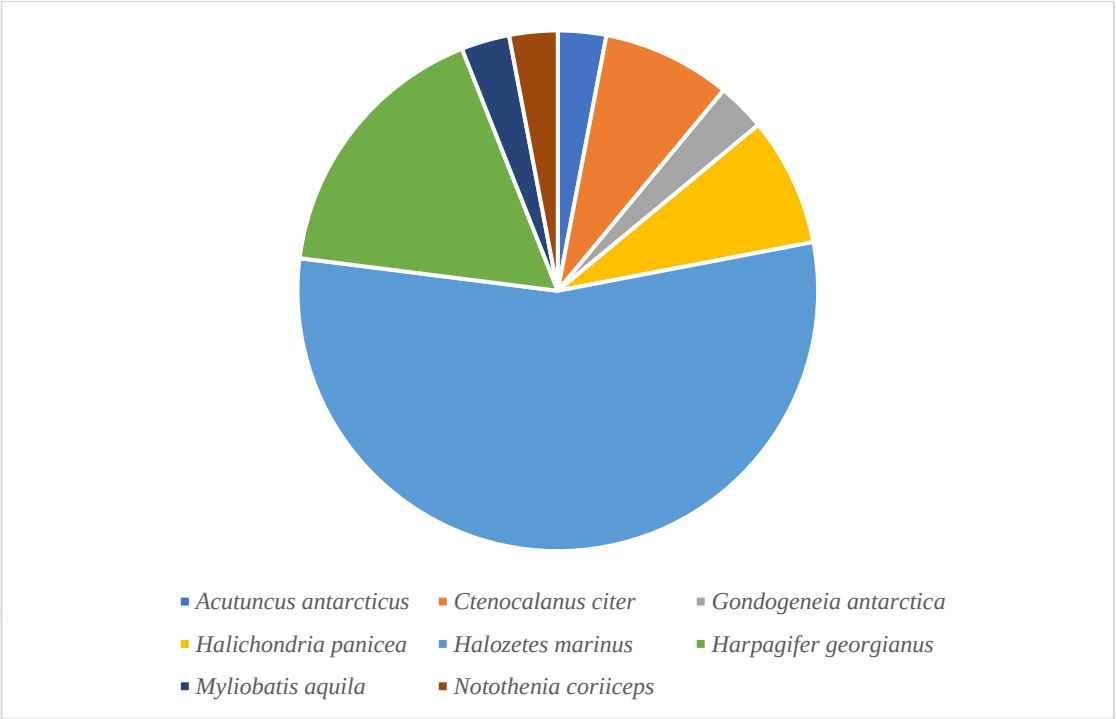


Figure 4.3. The pie chart shows the number of reads of sequences corresponding to each species.

The only species identified in Telephone Cove on Deception Island is *Ctenocalanus citer* as shown in Table 4.3.

Table 4.3. Taxonomic lineage of unique taxon found in Telephone Cove on Deception Island.

Kingdom	**Phylum**	**Class**	**Order**	**Family**	**Genus**	**Species**
Animalia	Arthropoda	Maxillopoda	Calanoida	Calanidae	Ctenocalanus	*Ctenocalanus citer*

Carlota Cove on Robert Island contains information of two taxa: *Acutuncus antarcticus*, and *Myliobatis aquila*. These belong to one kingdom, two phyla, two classes, two orders, two families, and two genera (Table 4.4).

Table 4.4. Taxonomic lineage of unique taxa found in Carlota Cove on Robert Island.

Kingdom	**Phylum**	**Class**	**Order**	**Family**	**Genus**	**Species**
Animalia	Tardigrada	Eutardigrada	Parachela	Hypsibiidae	Acutuncus	*Acutuncus antarcticus*
Animalia	Chordata	Chondrichthyes	Myliobatiformes	Myliobatidae	Myliobatis	*Myliobatis aquila*

The taxonomic classifications of eDNA samples gathered from Coppermine Cove are shown in Table 4.5. The samples include one kingdom, three phyla, four classes, four orders, five families, and five genera.

Table 4.5. Taxonomic lineage of unique taxa found in Coppermine Cove on Robert Island.

Kingdom	**Phylum**	**Class**	**Order**	**Family**	**Genus**	**Species**
Animalia	Chordata	Actinopteri	Perciformes	Harpagiferidae	Harpagifer	*Harpagifer georgianus*
Animalia	Chordata	Actinopteri	Perciformes	Nototheniidae	Notothenia	*Notothenia coriiceps*
Animalia	Arthropoda	Malacostraca	Amphipoda	Pontogeneiidae	Gondogeneia	*Gondogeneia antarctica*
Animalia	Arthropoda	Arachnida	Oribatida	Ameronothridae	Halozetes	*Halozetes marinus*
Animalia	Porifera	Demospongiae	Halichondrida	Halichondriidae	Halichondria	*Halichondria panicea*

Figure 4.4 is a clustered column chart that shows the distribution of the numbers of taxonomic levels by region. According to the chart, Coppermine Cove has the most number of species which is five.

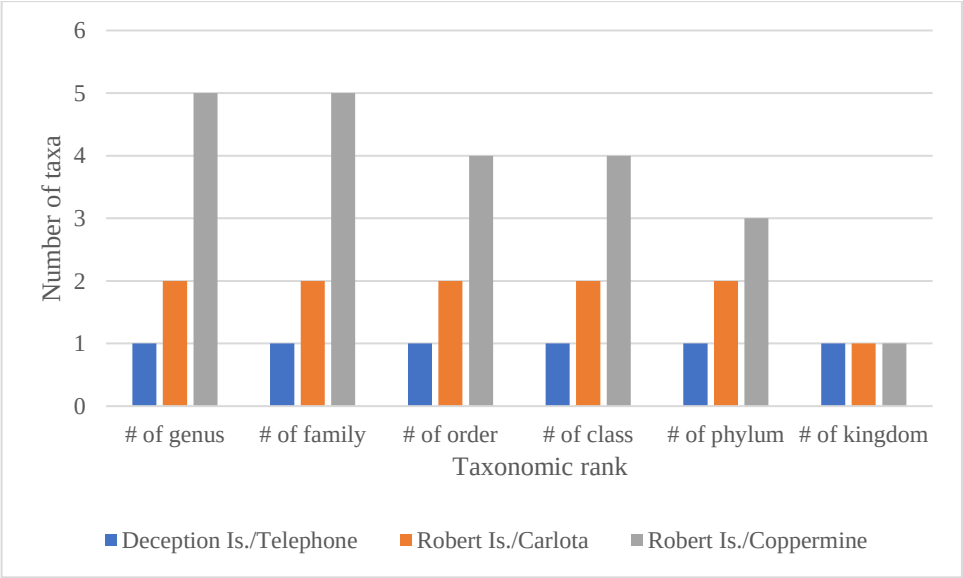


Figure 4.4. The numbers of genera, families, orders, classes, phyla, and kingdoms identified in the three locations.

The number of species found in each location was demonstrated in a column chart in Figure 4.5. The regions with the most species diversity is Coppermine Cove on Robert Island.

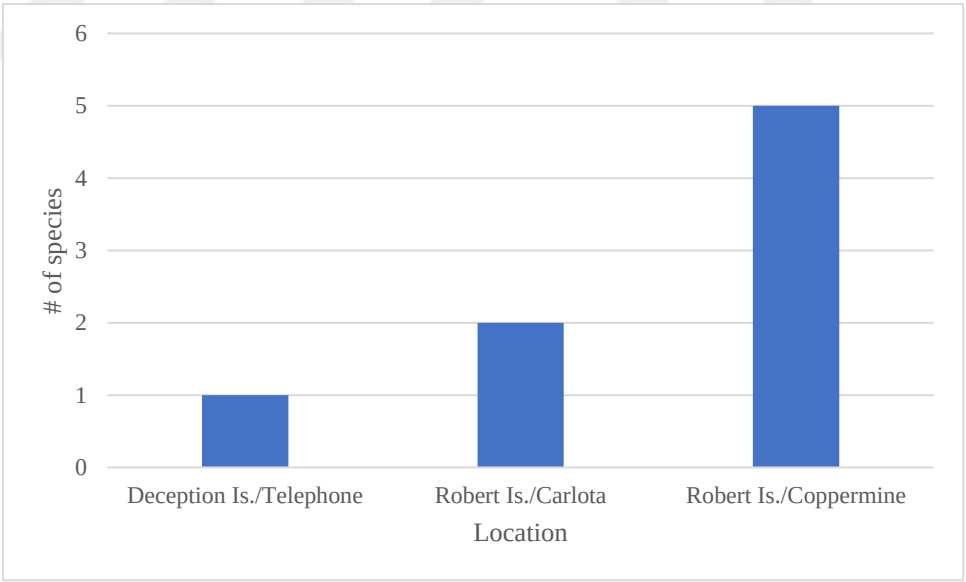


Figure 4.5. The numbers of species identified in the three locations as a bar chart.

The dendrogram that shows the clustering of the species diversity in the filters is given below (Figure 4.6). The filters AR 1.1, AR 1.2, and AR 1.3 are clustered together, and they were collected from Coppermine Cove on Robert Island. The filters AR 2.1, AR4.1, and AD 1.2 do not cluster with any other filters.

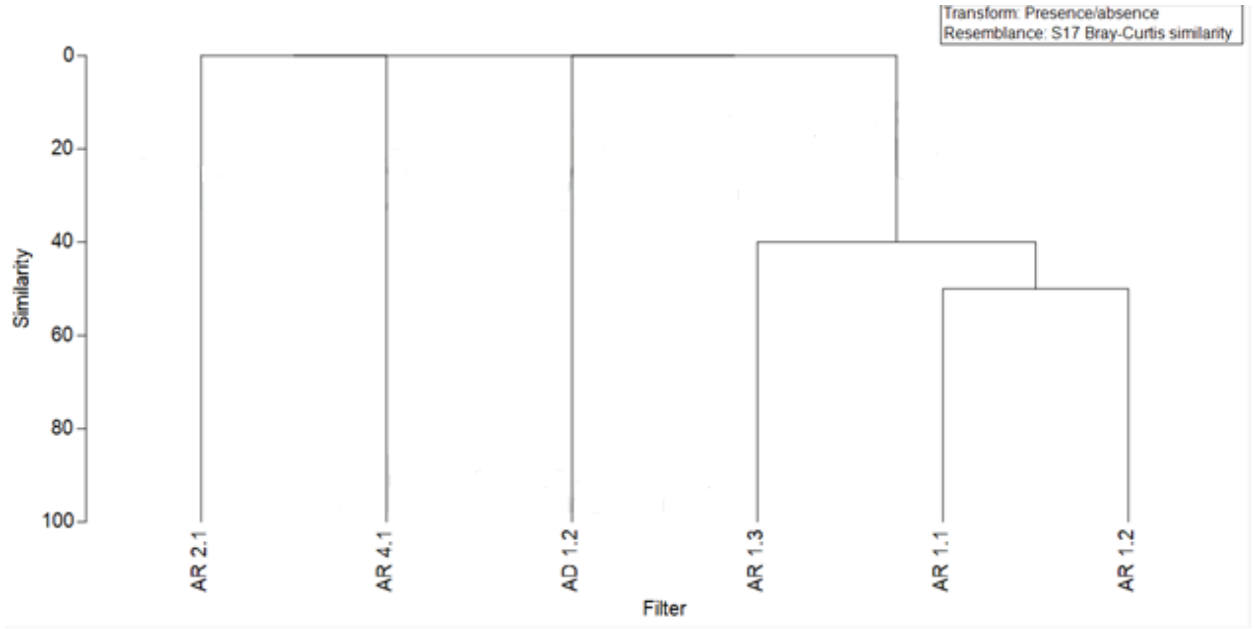


Figure 4.6. The dendrogram shows the similarity of the species diversity found in the filters according to the results obtained from MIDORI database.

Figure 4.7 indicates the hierarchical clustering of the species diversity found in PCR replicates using MIDORI database. According to the dendrogram, the PCR replicates acquired from the filter AR 1.2 filter (1-4 and 1-6) include the same information about the species detected. Similarly, PCR replicates 1-1 and 1-8 contain the same species even though they were taken from different filters (1-1 is from AR 1.1 and 1-8 is from AR 1.3). The samples collected from Deception Island (5-1 and 5-2) are also share the same types of species. However, PCR replicates 2-7 and 2-5 involve completely different species than the others, based on the characterization through the MIDORI database.

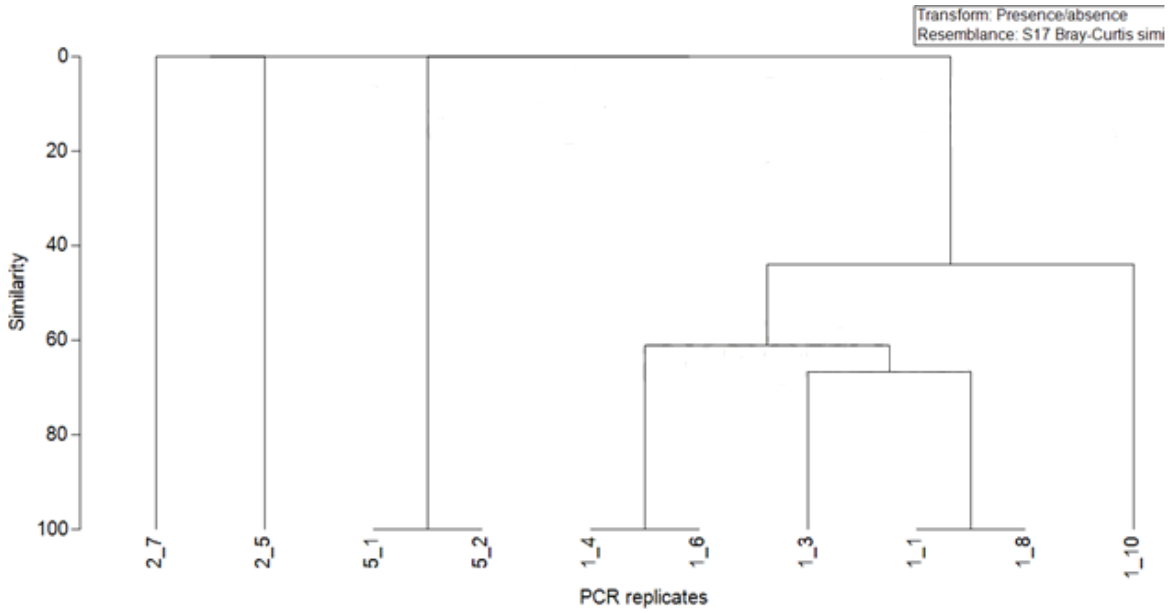


Figure 4.7. The dendrogram displays the similarity of the species diversity found in the PCR replicates according to the results obtained from MIDORI database.

The RDP Classifier algorithm was employed in the MIDORI Server to align sequences for taxonomic assignment instead of BLAST+. The RDP classifier is a naive Bayesian classifier initially designed to rapidly assign taxonomies based on rRNA sequence data (Wang et al., 2007). But the classifier can be adapted to classify mitochondrial DNA genes. The RDP Classifier has been trained for each reference database and made ready to use in MIDORI software. Porter and Hajibabaei (2018) used RDP Classifier with a CO1 training set for taxonomic assignment of insect species by metabarcoding. They also compared the trained classifier with the other popular method, BLAST. They found that the RDP classifier taxonomically allocates more queries per minute than the top BLAST hits method. Still, the accuracy of taxonomic assignment is higher for the BLAST hit method than for the RDP classifier method. These findings support the results obtained from MIDORI software since the BLAST+ required about 16x more time to give results than RDP Classifier in this study. The number of reads per sequence is relatively higher in the BLAST+ outcome than in the RDP classifier. In this data, only animal species were identified in MIDORI, which is reliable because the CO1 database used by the RDP classifier program in MIDORI includes only metazoan sequences. Thus, the analysis resulted in three fish, a shellfish, a water bear, a copepod, a mite, and a sponge.

166 operational taxonomic units (OTUs) remained uncharacterized in the MIDORI data, which corresponds to 95.4% of the total unique sequences. A higher percentage of OTUs indicates deficiencies in the MIDORI CO1 reference database. Another explanation for the high number of

OTUs and the low number of identified taxa may be that the RDP Classifier was not adequately trained for CO1 sequences.

Sample collection type dramatically influences the number of unidentified OTUs and the taxonomic composition of metazoan communities (Gielings et al., 2021). As it is mentioned in the materials and methods part, the quantity of the filtered water in Coppermine Cove on Robert Island (3000 ml) is higher than the quantity of the filtered water in Carlota Cove on Robert Island (1500 ml) and Telephone Cove on Deception Island (2000 ml). Therefore, we expect more sequence reads for identified taxa in Coppermine Cove. In the MIDORI database, as expected, the highest number of sequence reads was calculated in Coppermine Cove (84 reads in total).

The number of reads of sequences can give us information about the species abundance. In other words, as the number of individuals of the same species increases, the amount of DNA they will release into the environment will also increase. According to the MIDORI database, we can assume that *Halozetes marinus* was the most abundant organism on Robert Island at that time.

4.4. Species Identification by WoRMS Database

In this approach, unique sequences with BLAST percent identities are 97% or higher were retained. The taxonomic classification of each unique hits was identified using an in-house script. The scientific names of the taxa were used to gather taxonomic information from WoRMS database. 20 taxa were identified in total. Two taxa out of 20 are found in negative control, thus they were eliminated from the entire data set. Thus, the final number of taxa detected in WoRMS database was equal to 18. Of these, 16 taxa were identified to the species level, one taxon to the genus level, and one taxon to the family level.

Table 4.6 shows the lineages of all samples collected from five locations. There include three kingdoms, nine phyla, 11 classes, 15 orders, 17 families, 17 genera and 16 species. The pie chart in Figure 4.8 displays the read numbers of sequences of species identified via WoRMS database. Accordingly, *Petalonia fascia* (seaweed) has the highest number of reads (922, 27% of the total reads). *Bathycoccus prasinos* (algae) and *Sarcopeltis skottsbergii* (seaweed) have the second and the third highest number of reads that are 502 and 462 (15% and 14% of the total reads), respectively.

Table 4.6. Taxonomic lineage of unique taxa obtained from all water samples collected from five regions (Coppermine Cove, Carlota Cove, Telephone Cove, Turkish Point and Offshore) on three islands (Robert Island, Deception Island and Nansen Island) were identified using WoRMS database.

Kingdom	Phylum	Class	Order	Family	Genus	Species	# of Reads
Animalia	Cnidaria	Staurozoa	Stauromedusae	Haliclystidae	Haliclystus	*Haliclystus antarcticus*	246
Animalia	Arthropoda	Copepoda	Calanoida	Clausocalanidae	Ctenocalanus	*Ctenocalanus citer*	8
Animalia	Chordata	Actinopteri	Perciformes	Harpagiferidae	Harpagifer	*Harpagifer antarcticus*	17
Animalia	Nematoda	Chromadorea	Rhabditida	Rhabditidae	Litoditis	*Litoditis marina*	11
Plantae	Chlorophyta	Mamiellophyceae	Mamiellales	Bathycoccaceae	Bathycoccus	*Bathycoccus prasinos*	502
Plantae	Rhodophyta	Florideophyceae	Gigartinales	Gigartinaceae	Sarcopeltis	*Sarcopeltis scottsbergii*	462
Plantae	Rhodophyta	Florideophyceae	Gracilariales	Gracilariaceae	Curdiea	*Curdiea racovitzae*	271
Plantae	Rhodophyta	Florideophyceae	Ceramiales	Delesseriaceae	Myriogramme	*Myriogramme manginii*	119
Plantae	Rhodophyta	Florideophyceae	Halymeniales	Halymeniaceae	Cryptonemia	*Cryptonemia sp.*	61
Chromista	Ochrophyta	Dictyochophyceae	Florenciellales	Florenciellales incertae sedis	Pseudochattonella	*Pseudochattonella farcimen*	27
Chromista	Ochrophyta	Chrysophyceae	Parmales	Triparmaceae	Triparma	*Triparma laevis*	5
Chromista	Ochrophyta	Phaeophyceae	Desmarestiales	Desmarestiaceae	Desmarestia	*Desmarestia menziesii*	155
Chromista	Ochrophyta	Phaeophyceae	Ectocarpales	Scytosiphonaceae	Petalonia	*Petalonia fascia*	922
Chromista	Ochrophyta	Phaeophyceae	Ectocarpales	Acinetosporaceae	Pylaiella	*Pylaiella washingtoniensis*	170
Chromista	Ochrophyta	Phaeophyceae	Ascoseirales	Ascoseiraceae	Ascoseira	*Ascoseira mirabilis*	282
Chromista	Haptophyta	Prymnesiophyceae	Phaeocystales	Phaeocystaceae	Phaeocystis	*Phaeocystis antarctica*	43
Chromista	Myzozoa	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Margalefidinium	*Margalefidinium polykrikoides*	69
Chromista	Myzozoa	Dinophyceae	Gymnodiniales	Kareniaceae	Kareniaceae gen.	*Kareniaceae sp.*	8

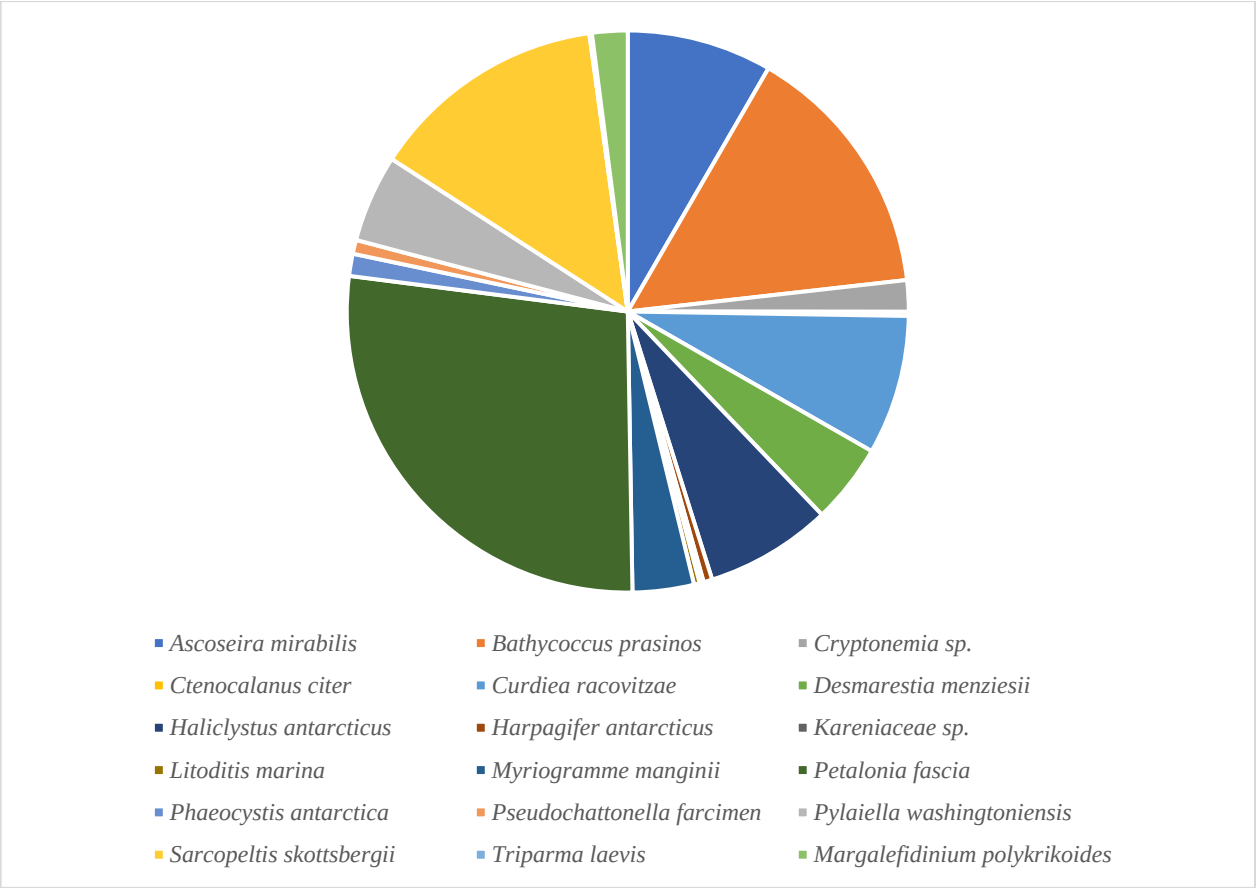


Figure 4.8. The pie chart shows the number of reads of sequences corresponding to each species.

Focusing on individual islands, eDNA samples collected from Telephone Cove on Deception Island contained information on four species. These includes three kingdoms, three phyla, three classes, three orders, four families, and four genera, as shown in Table 4.7.

Table 4.7. Taxonomic lineage of unique taxa found in Telephone Cove on Deception Island.

Kingdom	Phylum	Class	Order	Family	Genus	Species
Animalia	Arthropoda	Copepoda	Calanoida	Clausocalanidae	Ctenocalanus	*Ctenocalanus citer*
Plantae	Chlorophyta	Mamiellophyceae	Mamiellales	Bathycoccaceae	Bathycoccus	*Bathycoccus prasinos*
Chromista	Ochrophyta	Phaeophyceae	Ectocarpales	Scytosiphonaceae	Petalonia	*Petalonia fascia*
Chromista	Ochrophyta	Phaeophyceae	Ectocarpales	Acinetosporaceae	Pylaiella	*Pylaiella washingtoniensis*

Table 4.8 displays the taxonomic classification of species found on Turkish Point on Nansen Island. These belonged to two kingdoms, four phyla, four classes, four orders, four families, and four genera.

Table 4.8. Taxonomic lineage of unique taxa found in Turkish Point on Nansen Island.

Kingdom	Phylum	Class	Order	Family	Genus	Species
Plantae	Chlorophyta	Mamiellophyceae	Mamiellales	Bathycoccaceae	Bathycoccus	*Bathycoccus prasinos*
Plantae	Rhodophyta	Florideophyceae	Gracilariales	Gracilariaceae	Curdiea	*Curdiea racovitzae*
Chromista	Haptophyta	Prymnesiophyceae	Phaeocystales	Phaeocystaceae	Phaeocystis	*Phaeocystis antarctica*
Chromista	Myzozoa	Dinophyceae	Gymnodiniales	Gymnodiniaceae	Margalefidinium	*Margalefidinium polykrikoides*

Nine species were identified in Carlota Cove on Robert Island, and the hierarchical structure of taxonomy is shown in the following table (Table 4.9). The species belonged to three kingdoms, five phyla, five classes, nine orders, nine families, and nine genera.

Table 4.9. Taxonomic lineage of unique taxa found in Carlota Cove on Robert Island.

Kingdom	Phylum	Class	Order	Family	Genus	Species
Animalia	Cnidaria	Staurozoa	Stauromedusae	Haliclystidae	Haliclystus	*Haliclystus antarcticus*
Animalia	Nematoda	Chromadorea	Rhabditida	Rhabditidae	Litoditis	*Litoditis marina*
Plantae	Chlorophyta	Mamiellophyceae	Mamiellales	Bathycoccaceae	Bathycoccus	*Bathycoccus prasinos*
Plantae	Rhodophyta	Florideophyceae	Gigartinales	Gigartinaceae	Sarcopeltis	*Sarcopeltis scottsbergii*
Plantae	Rhodophyta	Florideophyceae	Gracilariales	Gracilariaceae	Curdiea	*Curdiea racovitzae*
Plantae	Rhodophyta	Florideophyceae	Ceramiales	Delesseriaceae	Myriogramme	*Myriogramme manginii*
Plantae	Rhodophyta	Florideophyceae	Halymeniales	Halymeniaceae	Cryptonemia	*Cryptonemia sp.*
Chromista	Ochrophyta	Phaeophyceae	Desmarestiales	Desmarestiaceae	Desmarestia	*Desmarestia menziesii*
Chromista	Ochrophyta	Phaeophyceae	Ascoseirales	Ascoseiraceae	Ascoseira	*Ascoseira mirabilis*

In Coppermine Cove on Robert Island, nine species were found, which belonged to three kingdoms, five phyla, five classes, nine orders, nine families, and nine genera (Table 4.10).

Table 4.10. Taxonomic lineage of unique taxa found in Coppermine Cove on Robert Island.

Kingdom	Phylum	Class	Order	Family	Genus	Species
Animalia	Cnidaria	Staurozoa	Stauromedusae	Haliclystidae	Haliclystus	*Haliclystus antarcticus*
Animalia	Chordata	Actinopteri	Perciformes	Harpagiferidae	Harpagifer	*Harpagifer antarcticus*
Plantae	Chlorophyta	Mamiellophyceae	Mamiellales	Bathycoccaceae	Bathycoccus	*Bathycoccus prasinos*
Plantae	Rhodophyta	Florideophyceae	Gigartinales	Gigartinaceae	Sarcopeltis	*Sarcopeltis scottsbergii*
Plantae	Rhodophyta	Florideophyceae	Gracilariales	Gracilariaceae	Curdiea	*Curdiea racovitzae*
Plantae	Rhodophyta	Florideophyceae	Ceramiales	Delesseriaceae	Myriogramme	*Myriogramme manginii*
Plantae	Rhodophyta	Florideophyceae	Halymeniales	Halymeniaceae	Cryptonemia	*Cryptonemia sp.*
Chromista	Ochrophyta	Phaeophyceae	Desmarestiales	Desmarestiaceae	Desmarestia	*Desmarestia menziesii*
Chromista	Ochrophyta	Phaeophyceae	Ascoseirales	Ascoseiraceae	Ascoseira	*Ascoseira mirabilis*

The taxonomic information of six species found in the offshore around Robert Island are shown in the Table 4.11. They were in two kingdoms, four phyla, five classes, five orders, and five genera.

Table 4.11. Taxonomic lineage of unique taxa found in the offshore around Robert Island.

Kingdom	Phylum	Class	Order	Family	Genus	Species
Plantae	Chlorophyta	Mamiellophyceae	Mamiellales	Bathycoccaceae	Bathycoccus	*Bathycoccus prasinos*
Chromista	Ochrophyta	Dictyochophyceae	Florenciellales	Florenciellales incertae sedis	Pseudochattonella	*Pseudochattonella farcimen*
Chromista	Ochrophyta	Chrysophyceae	Parmales	Triparmaceae	Triparma	*Triparma laevis*
Chromista	Haptophyta	Prymnesiophyceae	Phaeocystales	Phaeocystaceae	Phaeocystis	*Phaeocystis antarctica*
Chromista	Myzozoa	Dinophyceae	Gymnodiniales	Kareniaceae	Kareniaceae gen.	*Kareniaceae sp.*

Figure 4.9 is a clustered column chart that shows the distribution of the numbers of taxonomic levels by region. According to the chart, the maximum numbers of species, genus, family, and order found in a location are the same and equal to nine. These locations were species diversity is much higher at Coppermine and Carlota Coves on Robert Island.

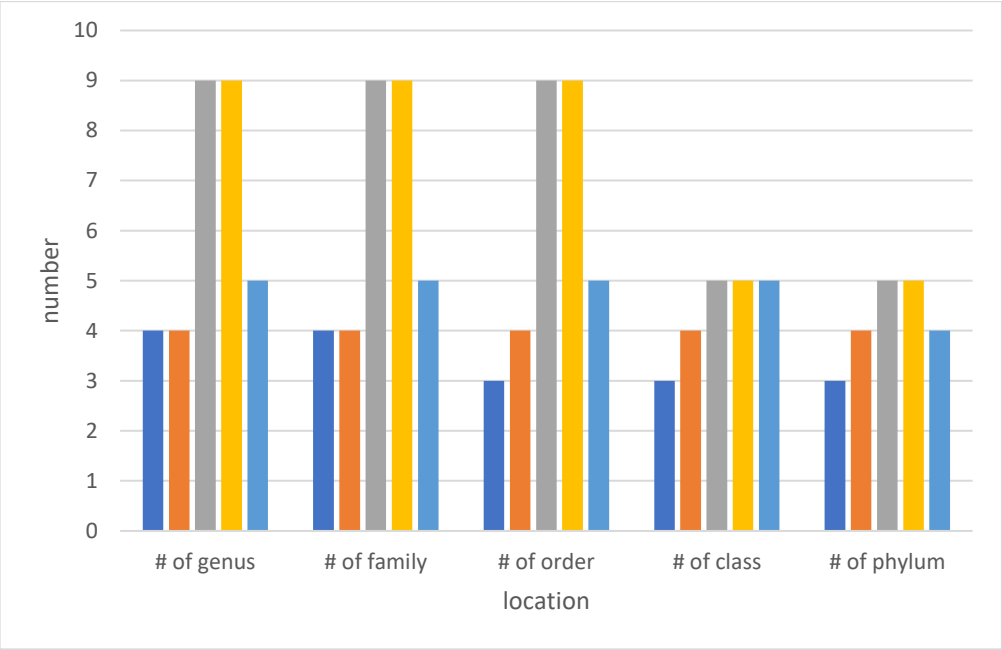


Figure 4.9. The numbers of genera, families, orders, classes, phyla, and kingdoms identified in the five locations.

The number of species found in each location is demonstrated in a column chart in Figure 4.10. The regions with the highest number of species are Carlota Cove and Coppermine Cove on Robert Island.

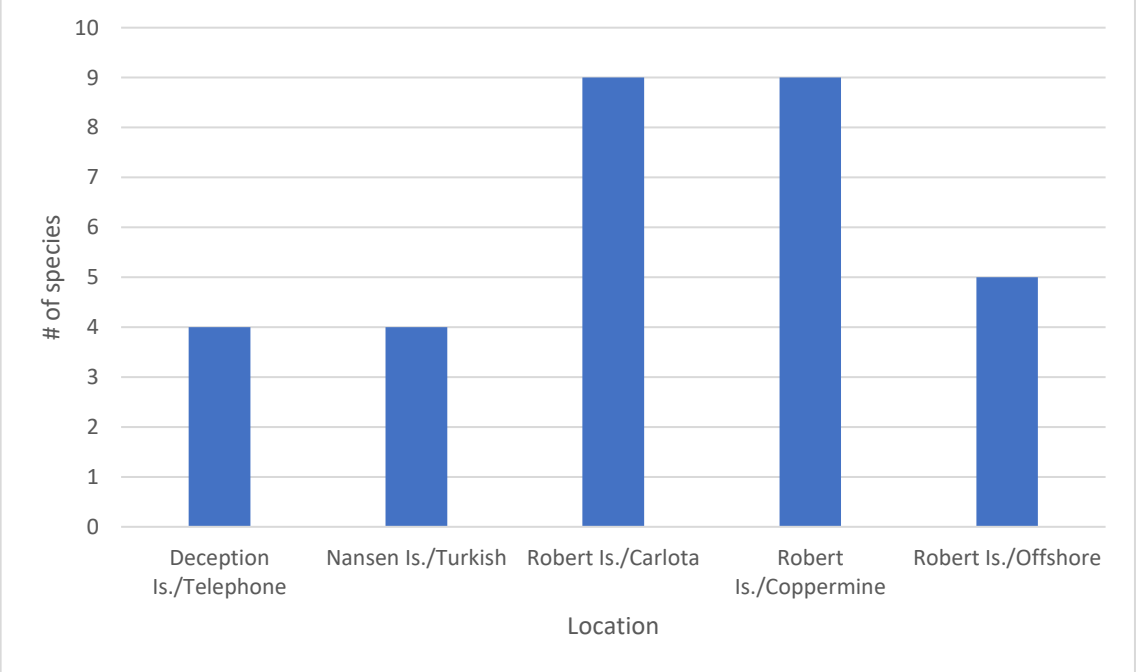


Figure 4.10. The numbers of species identified in the five locations.

Figure 4.11 is a clustered bar chart which is showing how many different sites each species is found in. *Bathycoccus prasinos* is the only species identified in all five sites. *Curdiea racovitzae* is detected in three locations through Robert Island and Nansen Island. *Sarcopeltis scottsbergii*, *Myriogramme manginii*, *Desmarestia menziesii*, *Ascoseira mirabilis*, *Phaeocystis antarctica*, *Haliclystus antarcticus*, and *Crytonemia sp.* were detected in two of the five locations. The remaining species (Others in the Figure 4.11) are identified in one region. Species names and the regions where they are defined are presented in the Table 4.12.

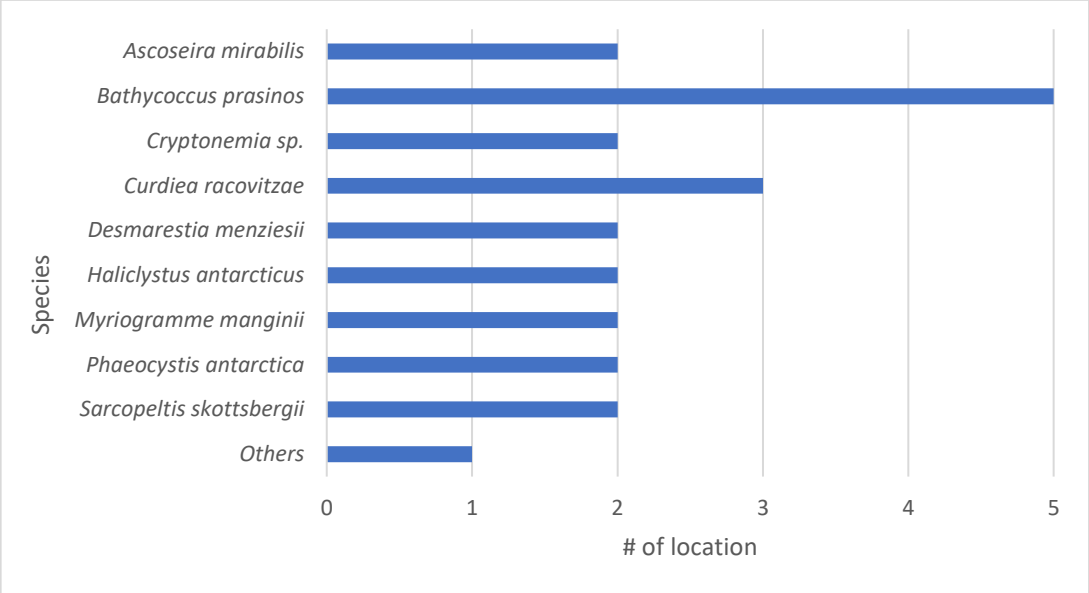


Figure 4.11. The number of locations where each species found.

Table 4.12. The species were identified in only one region.

Species	**Location**
Ctenocalanus citer	Telephone Cove
Harpagifer antarcticus	Coppermine Cove
Kareniaceae sp.	Offshore
Litoditis marina	Carlota Cove
Margalefidinium polykrikoides	Turkish Point
Petalonia fascia	Telephone Cove
Pseudochattonella farcimen	Offshore
Pylaiella washingtoniensis	Telephone Cove
Triparma laevis	Offshore

Figure 4.12 shows the hierarchical clustering for the species diversity between the filters. The species diversities in the filters of Deception Island (AD 1.1 and AD 1.2) are highly similar, and the filters of Nansen Island (AN 1 and AN 2) also have similar species diversity and clustered together. The percentages of similarity were about 80% for both islands. The filters of Robert Island (AR 1.2, AR 3, AR 2.1, AR 5.1, AR 5.2, AR 1.1, AR 4.1, and AR 1.3) are clustered together except the filter AR 9. The similarity rate is more than eighty percent according to the species whose existence has been detected. The similarity percentage is 100% for filters AR 2.1 and AR 5.1. In other words, the same species were identified in both filter samples. The filter AR 9 which was collected from offshore of Robert Island is clustered with the filters from Nansen Island.

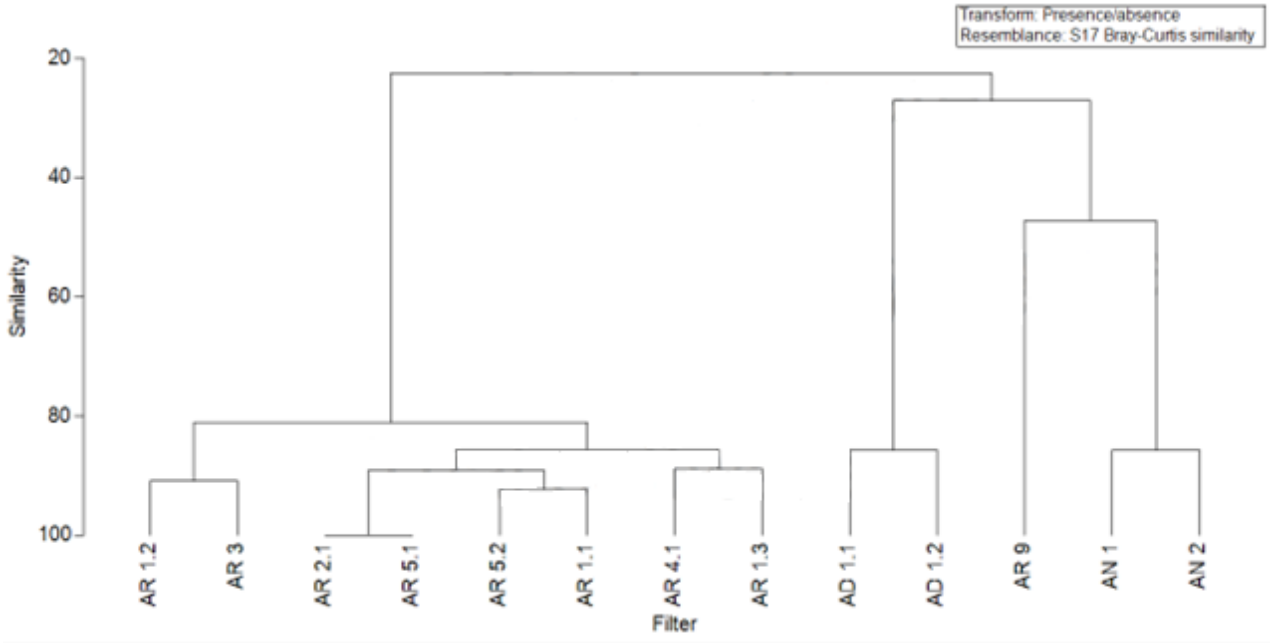


Figure 4.12. The dendrogram shows the similarity of the species diversity found in the filters according to the results obtained from WoRMS database.

The similarities of species diversity between the PCR replicates are shown in Figure 4.13. According to the dendrogram, PCR replicates of the water samples from Deception Island (5-1, 4-7, 4-8, 4-10, and 5-2) is clustered together. 4-7, 4-8, 4-10, and 5-2 include the same species (percentage of similarity is 100%), and 5-1 has an extra species than the others. The PCR replicates obtained from the eDNA samples collected from Nansen Island (4-1, 4-3, 4-4, 4-5, and 4-6) and offshore of Robert Island (3-7, 3-8, and 3-10) are clustered together. PCR replicates of the other regions in Robert Island except the offshore also show similarity among themselves in terms of species diversity. PCR replicates of the negative control (3-4, 3-5, and 3-6) can be seen in the left side of the dendrogram; they are not clustered neither among themselves nor with other replicates.

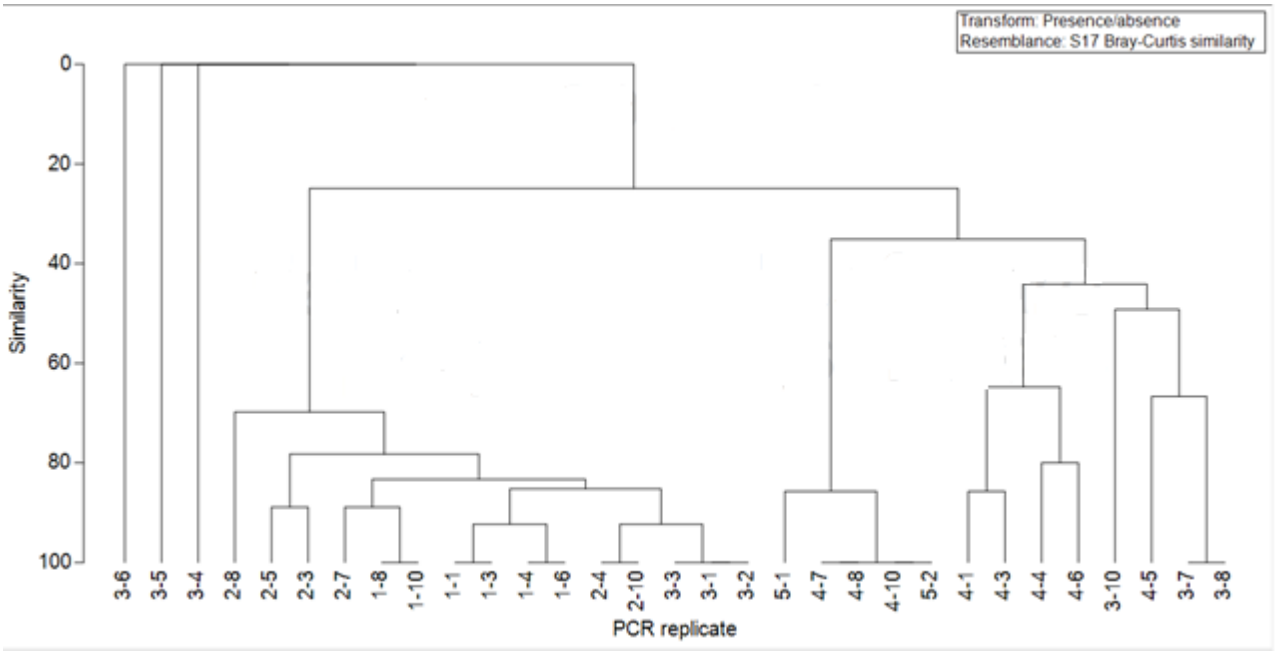


Figure 4.13. The dendrogram displays the similarity of the species diversity found in the PCR replicates according to the results obtained from WoRMS database.

As it is known that WoRMS contains taxonomic information of only marine species, we expect to detect marine species in our eDNA sample. When we look at the species groups the identified species belong to, we see eight seaweeds, seven algae, a fish, a jellyfish, and a copepod. All species live in the marine environment; thus, expectations were met. Additionally, although offshore Robert Island is closer in the distance to Carlota and Coppermine Coves, it showed high similarities with Nansen Island and Deception Island regarding identified taxa in WoRMS. The taxa distribution in each location presents that Turkish Point on Nansen Island and offshore of Robert Island have the taxonomic information of only algae species, unlike the other locations. The absence of fish taxa in Turkish Point may be explained by the presence of *Margalefidinium polykrikoides*, a species known to cause fish deaths by producing red tides (Son et al., 2011). However, another probable reason may be that no fish DNA got into the filter due to the surrounding algae density.

As another species, to note, *Sarcopeltis skottsbergii* is an endemic seaweed of southern of South America (Castro-varela et al., 2022) and was identified in both Carlota and Coppermine Coves on Robert Island, with 462 reads in the taxonomic assignment processed in the WoRMS database. One of the reasons for the occurrence of this seaweed native to South America in Antarctica may be that this species came with water currents from South America towards this region. Another reason could be shipping activities; shipping can facilitate the spread of non-native species through ballast water and hull marine growth (K. E. Costello et al., 2022).

4.5. Species Identification by GenBank

Species diversity in Antarctic Peninsula was identified using the information taken from the GenBank database via an in-house script. There were 391 unique sequences in the data when the percent identity cut-off is 97%, and a name at a taxonomical category was assigned if that name was found in more than 85% of the first 100 hits in GenBank. The unique sequences were detected in 32 PCR replicates of 14 filters, including the negative controls. A total of 43 taxa were ascertained, but seven were discarded from the species data because they were found in the negative controls. Overall, 36 taxa were identified from GenBank database.

Several pie charts were built in order to demonstrate the frequencies of the sequences of the detected taxa from GenBank database. Figure 4.14 displays the species identified using GenBank database corresponding to their number of reads. *Petalonia fascia,* a seaweed, had 922 reads (26% of the total number of reads), making it the species with the highest number of reads in this dataset. In addition, another species with a relatively higher number of reads is *Bathycoccus prasinos*, an algae*;* it had 502 reads (14% of the total number of reads). The third highest number of reads was 462, and belonged to *Sarcopeltis skottsbergii*, a seaweed (13% of the total number of reads).

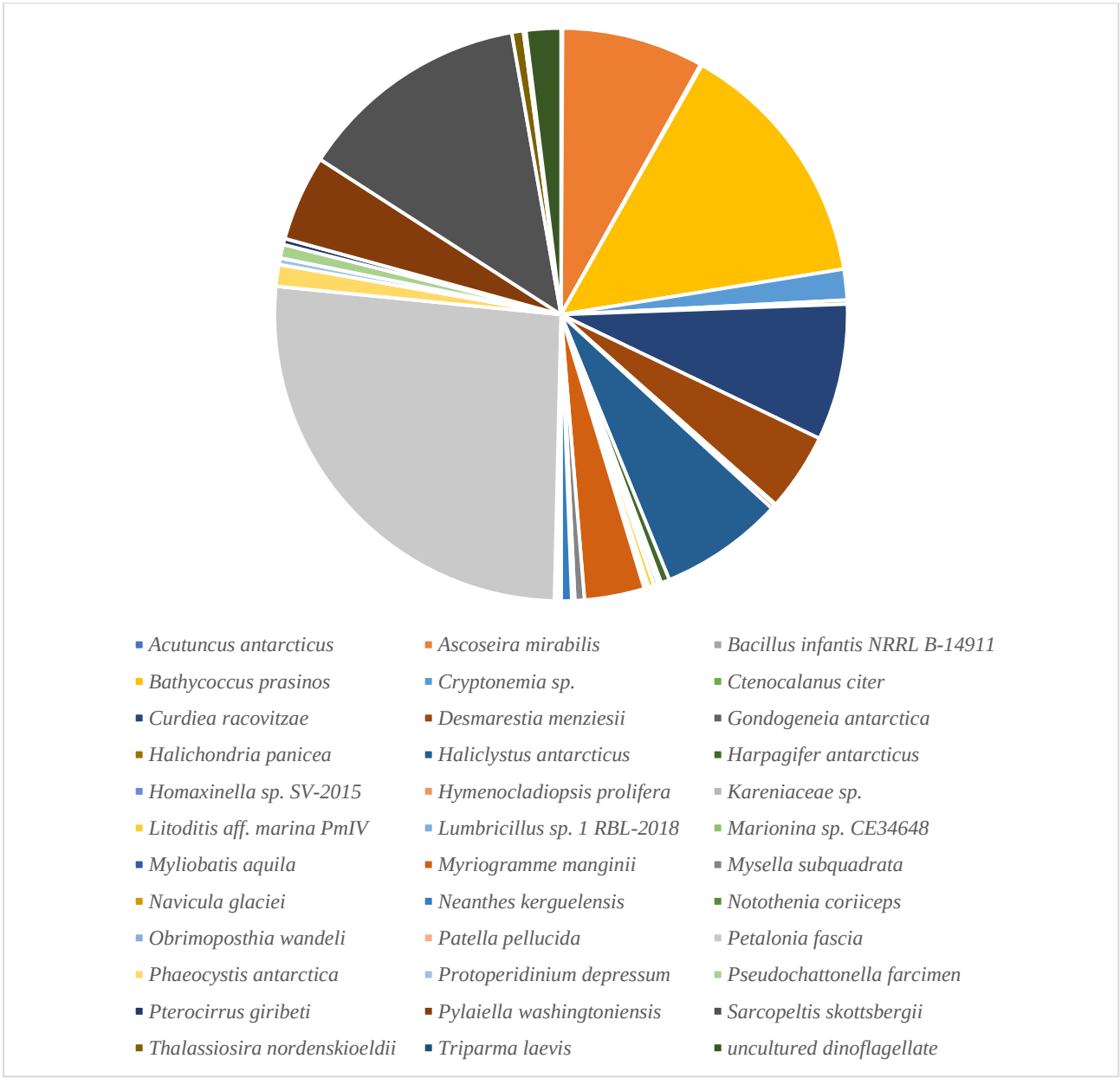


Figure 4.14. The pie chart shows the number of reads of sequences corresponding to each species identified in GenBank.

Figure 4.15 displays the relative frequencies of the number of reads of genera in the BLAST+ output corresponding to the information obtained from the GenBank database. The most frequent sequences are associated with *Petalonia*, *Bathycoccus*, and *Sarcopeltis*. The numbers of reads are 922, 502, and 462, respectively. Additionally, there are 77 not available (N/A) sequences at the genus level.

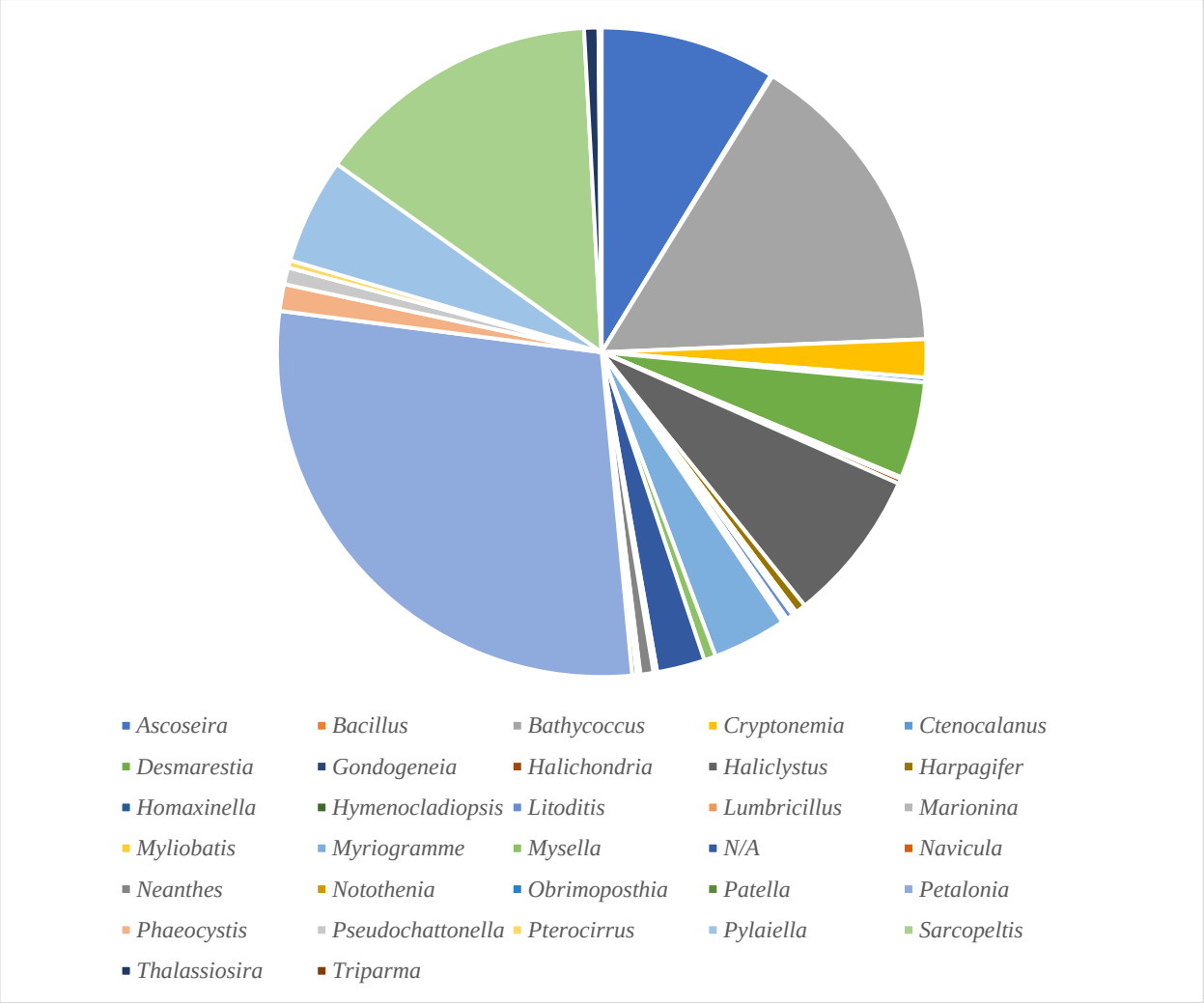


Figure 4.15. The pie chart shows the number of reads of sequences corresponding to each genus identified in GenBank.

Figure 4.16 demonstrates the families with the number of reads of sequences found in the GenBank database. The pie chart indicates that Scytosiphonaceae (algae) has the maximum number of reads, and it is 922. Palmariaceae (seaweed) and Bathycoccaceae (algae) families become second and third in terms of the read-number. Palmariaceae has 544 reads, and Bathycoccaceae has 502 reads. 382 sequence-reads remain N/A at the family level.

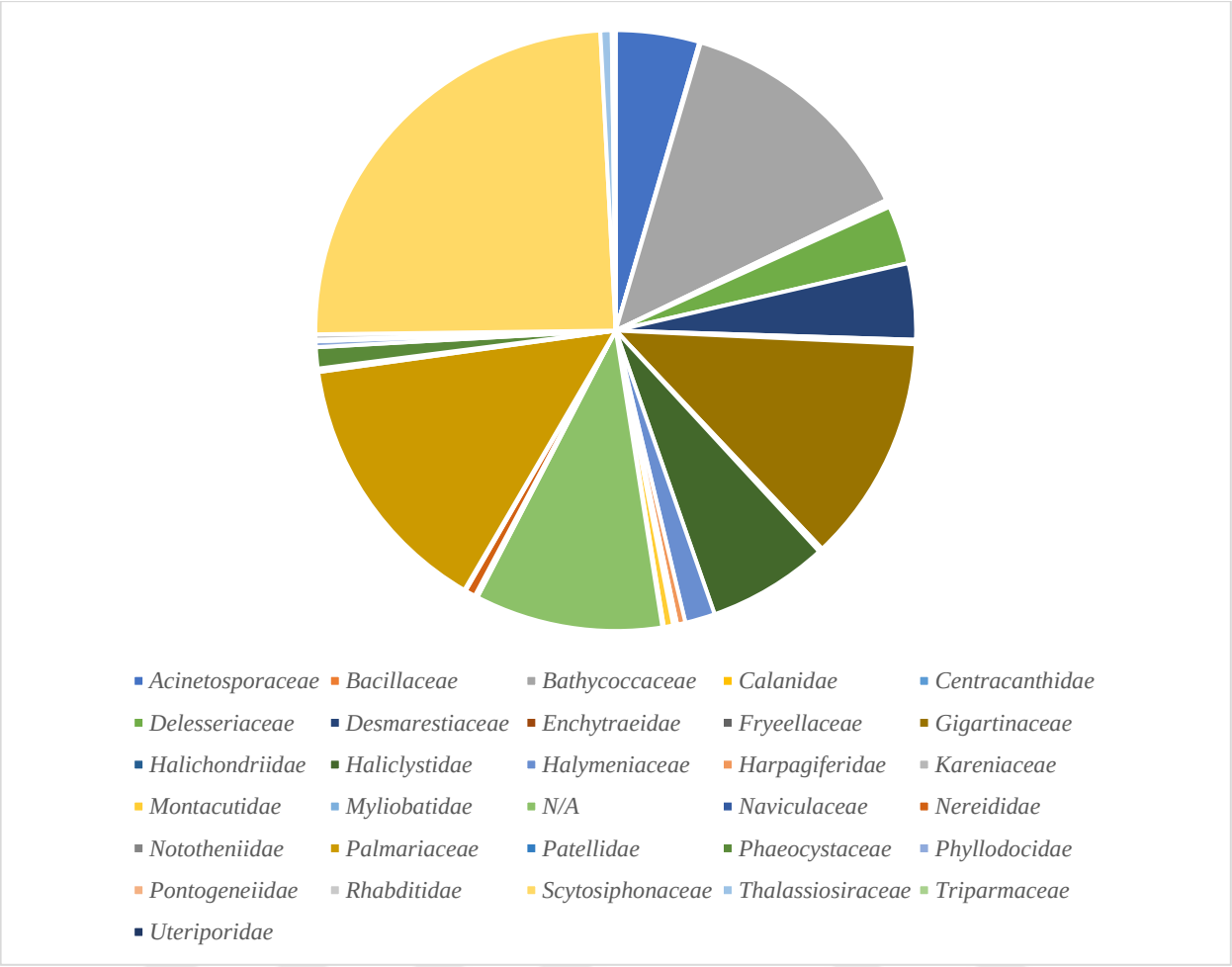


Figure 4.16. The pie chart shows the number of reads of sequences corresponding to each family identified in GenBank.

Distribution of the counts of the sequences that indicate orders found in GenBank can be seen in Figure 4.17. The maximum number of reads is 1092 and belongs to the order of Ectocarpales (seaweed). Palmariales (seaweed) has the highest second number of reads that is 544, and Mamiellales (seaweed) has the highest third number of reads which is 502. There are 76 N/A reads for the order level.

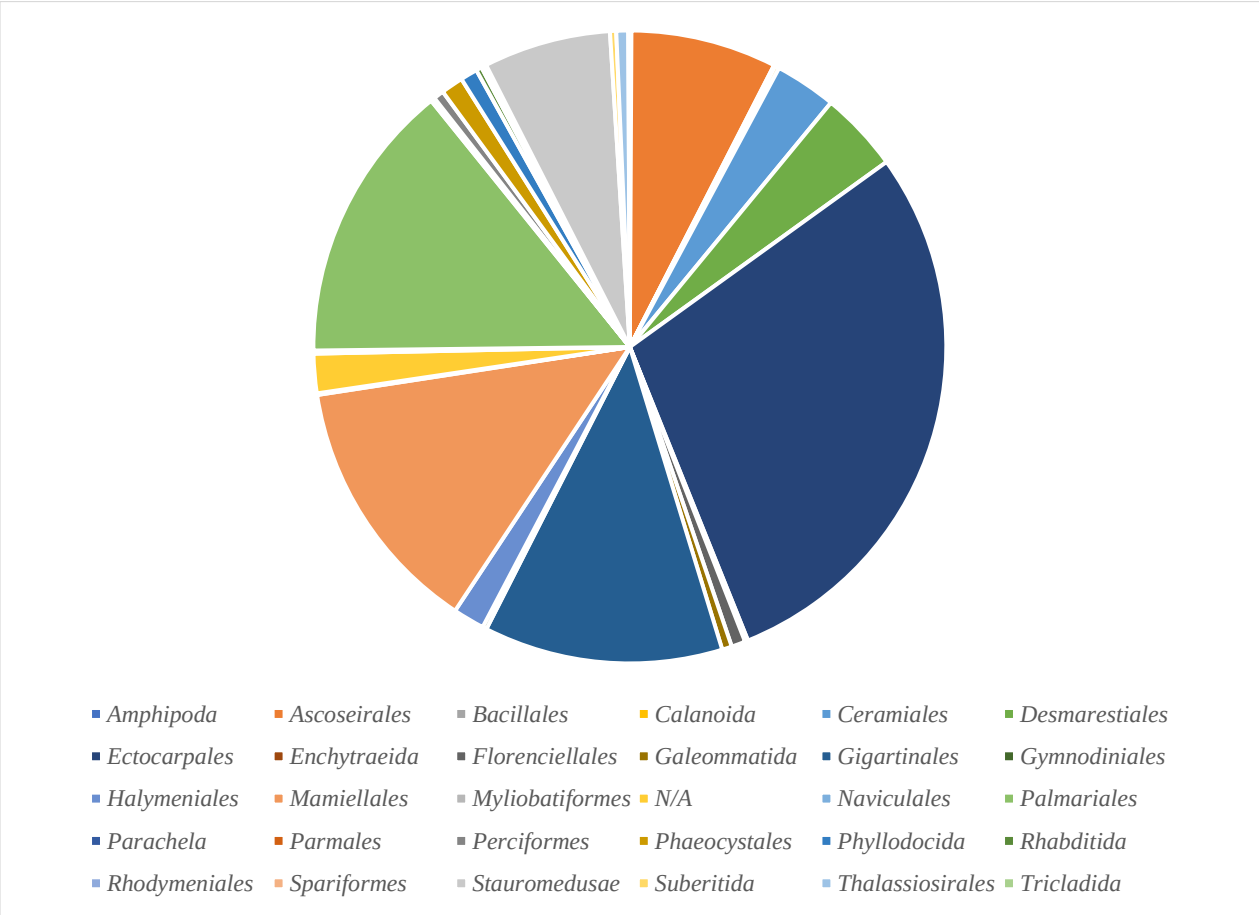


Figure 4.17. The pie chart shows the number of reads of sequences corresponding to each order detected in GenBank.

The pie chart below (Figure 4.18) was built to see the relative frequencies of sequence reads in terms of classes. The read numbers of the two classes with the highest number of reads are close to each other. Phaeophyceae (algae) has 1529 reads, while Florideophyceae (seaweed) has 1500 reads. Mamiellophyceae (seaweed) has the third most number of reads which is 502. Besides, there are 43 N/A reads in the class data.

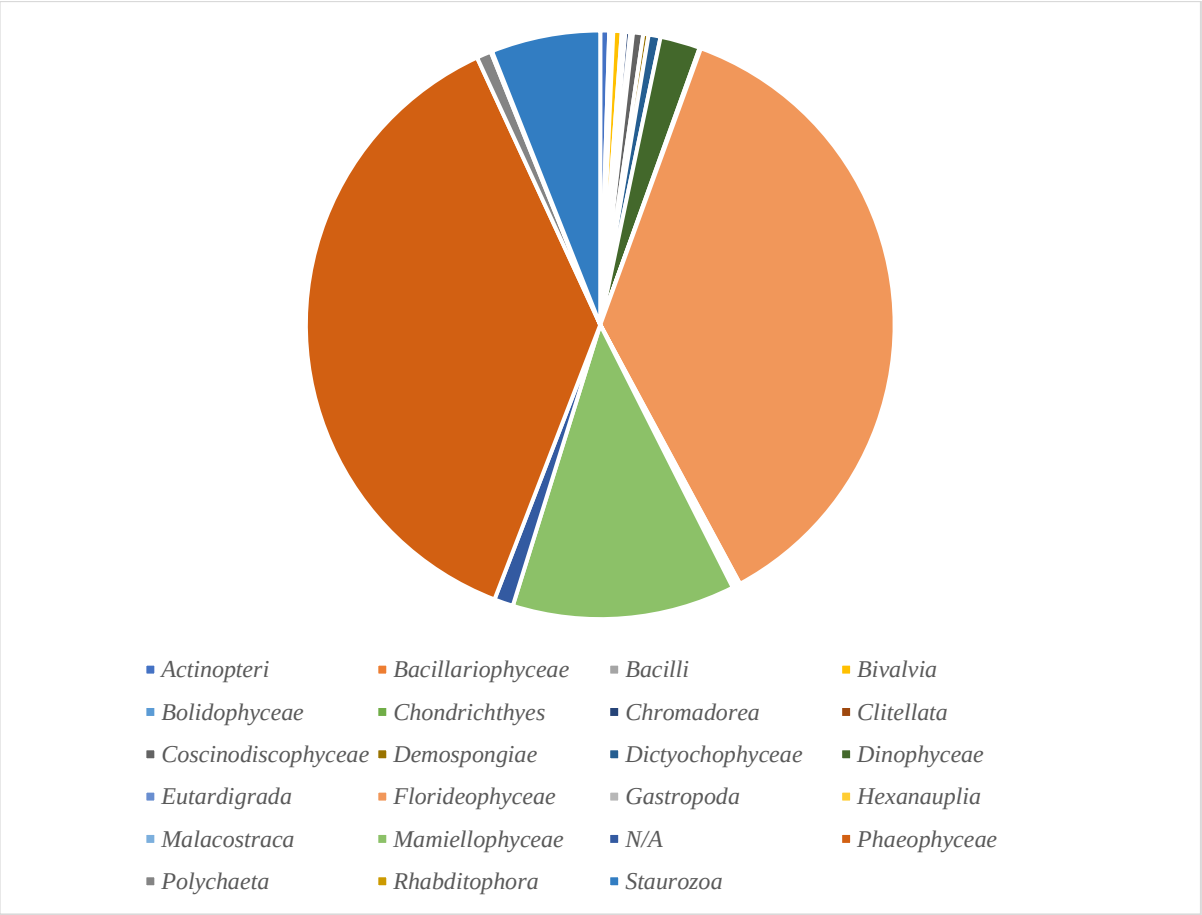


Figure 4.18. The pie chart shows the number of reads of sequences corresponding to each class found in GenBank.

Figure 4.19 indicates the phyla with the number of reads of sequences found in the GenBank database. The pie chart shows that Rhodophyta (algae) has the maximum number of reads which is 1500. Chlorophyta (algae) is the second phylum (1447), and Cnidaria (jellyfish) is the third phylum (246), corresponding to their number of reads. 1651 sequence reads remain N/A at the phylum level.

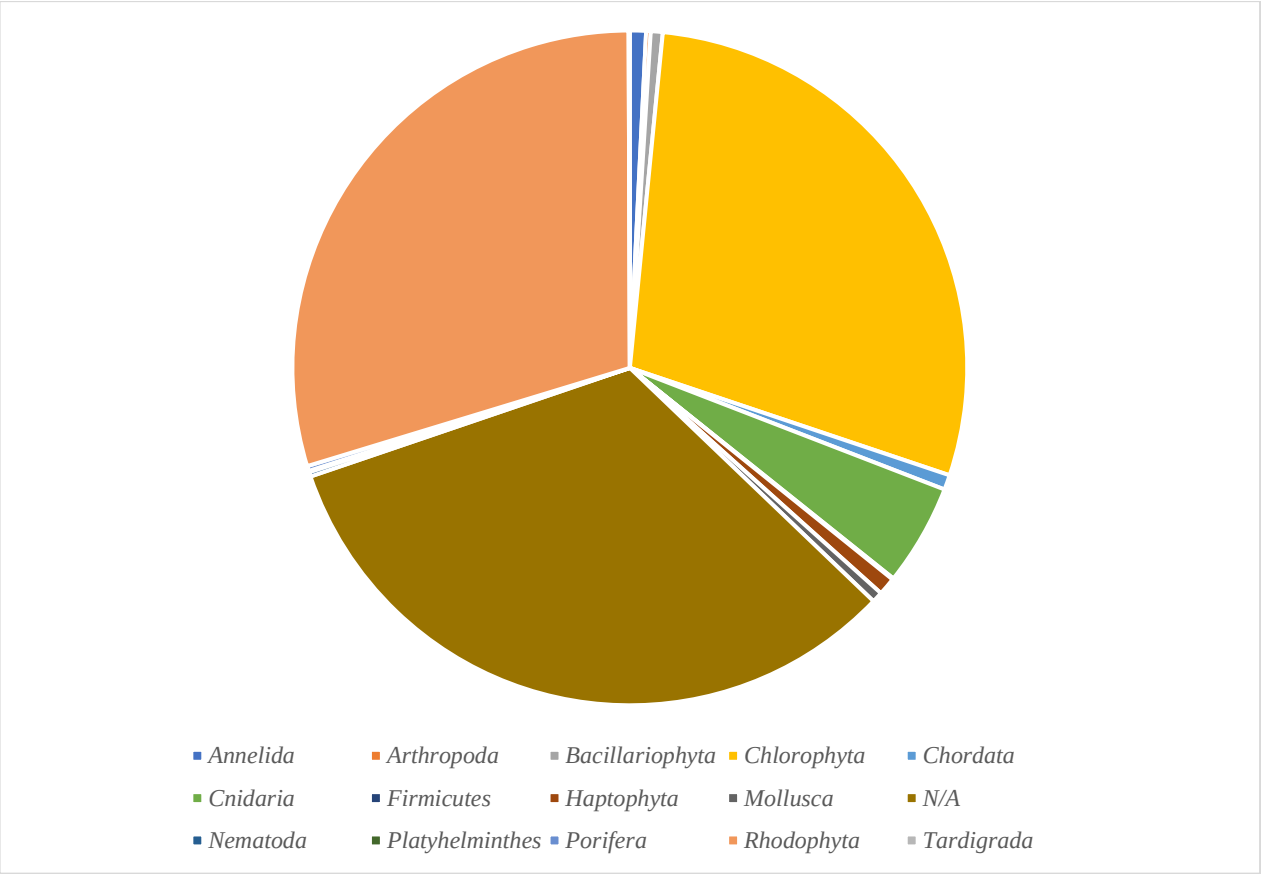


Figure 4.19. The pie chart shows the number of reads of sequences corresponding to each phylum identified in GenBank.

The dataset obtained from GenBank includes the sequence reads of five kingdoms. The relative frequencies of the sequence reads are shown in Figure 4.20. Plantae have 2947 reads, the maximum number of counts (58.2% of the total count), Animalia species comprised 387 sequence reads (7.6% of the total count), and 1651 sequence reads were stated as N/A (32.6%).

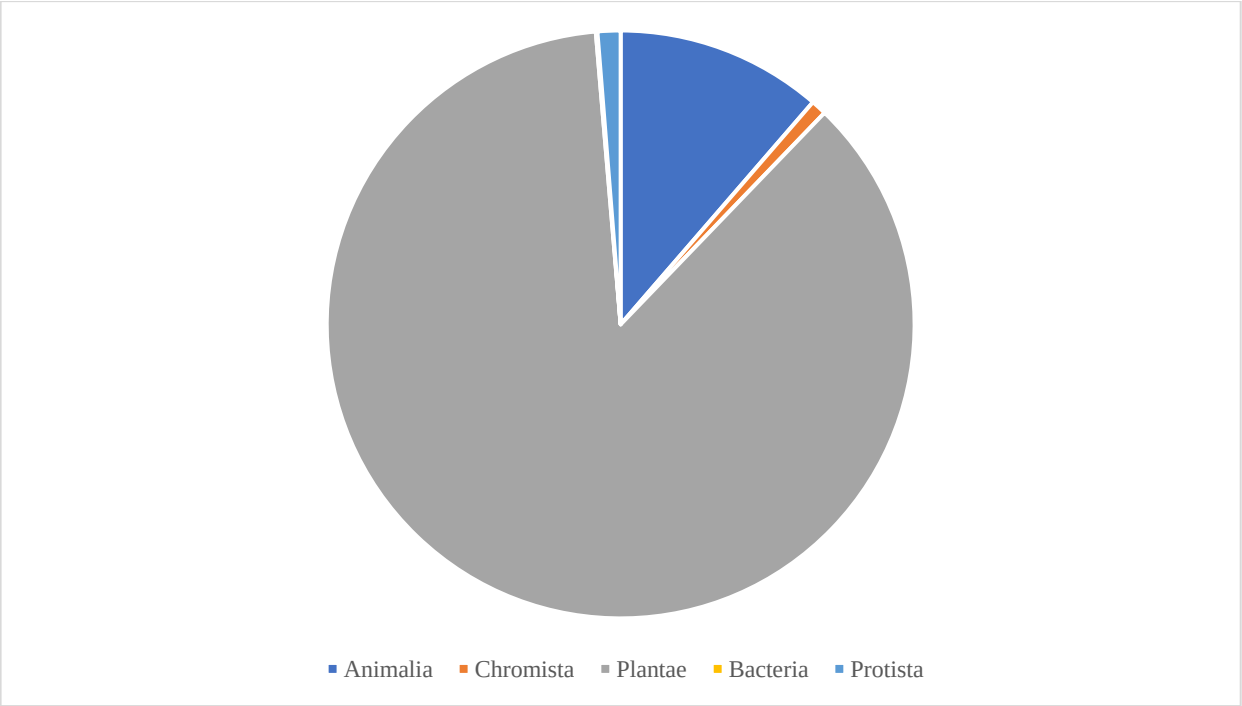


Figure 4.20. The pie chart shows the number of reads of sequences corresponding to each kingdom identified in GenBank.

Coppermine Cove on Robert Island has the most species richness, with 19 unique taxa; similarly, 17 taxa were found in Carlota Cove on Robert Island (Figure 4.21). There were detected ten taxa in common in both locations.

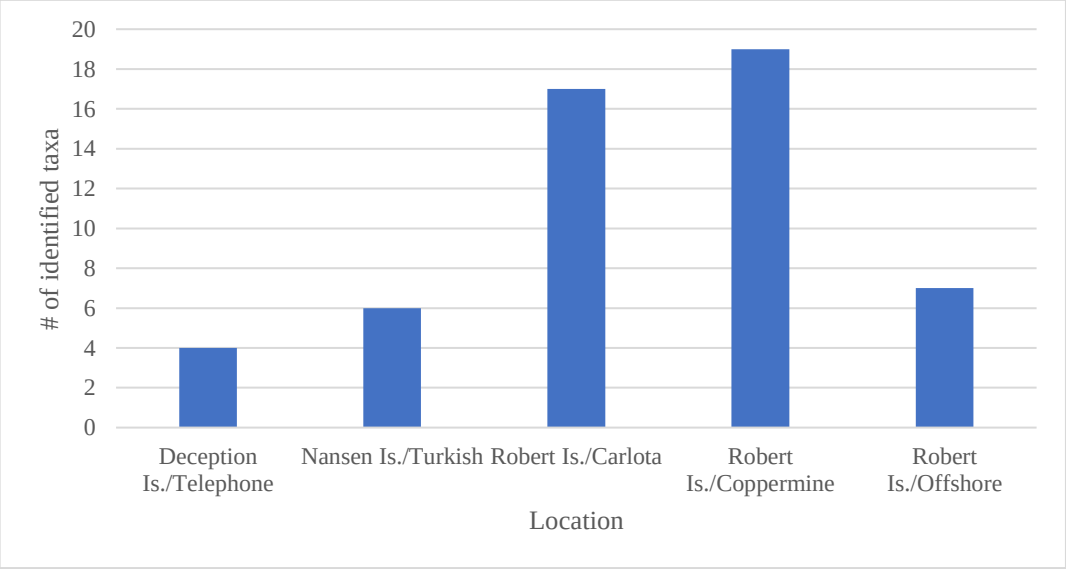


Figure 4.21. The column graph represents the number of unique taxa identified in each location based on the GenBank database.

The dendrogram in Figure 4.22 were formed to observe the similarity in species diversity determined for each filter. As stated earlier, AR 6.2 is the negative control filter, and did not cluster with the other filters. The filters used in Deception Island (AD 1.1 and AD 1.2) clustered together, and shared 40% of similarity in terms of taxa they contain. The same result can be seen for the filters used in Nansen Island (AN 1 and AN 2). The filter AR 9 is clustered with the filters AN 1 and AN 2. The filters obtained from the coastal regions of Robert Island (AR 4.1, AR 5.2, AR 1.3, AR 2.1, AR5.1, AR 3, AR 1.1, and AR 1.2) contain at least 60% species diversity similarity, and they were clustered together.

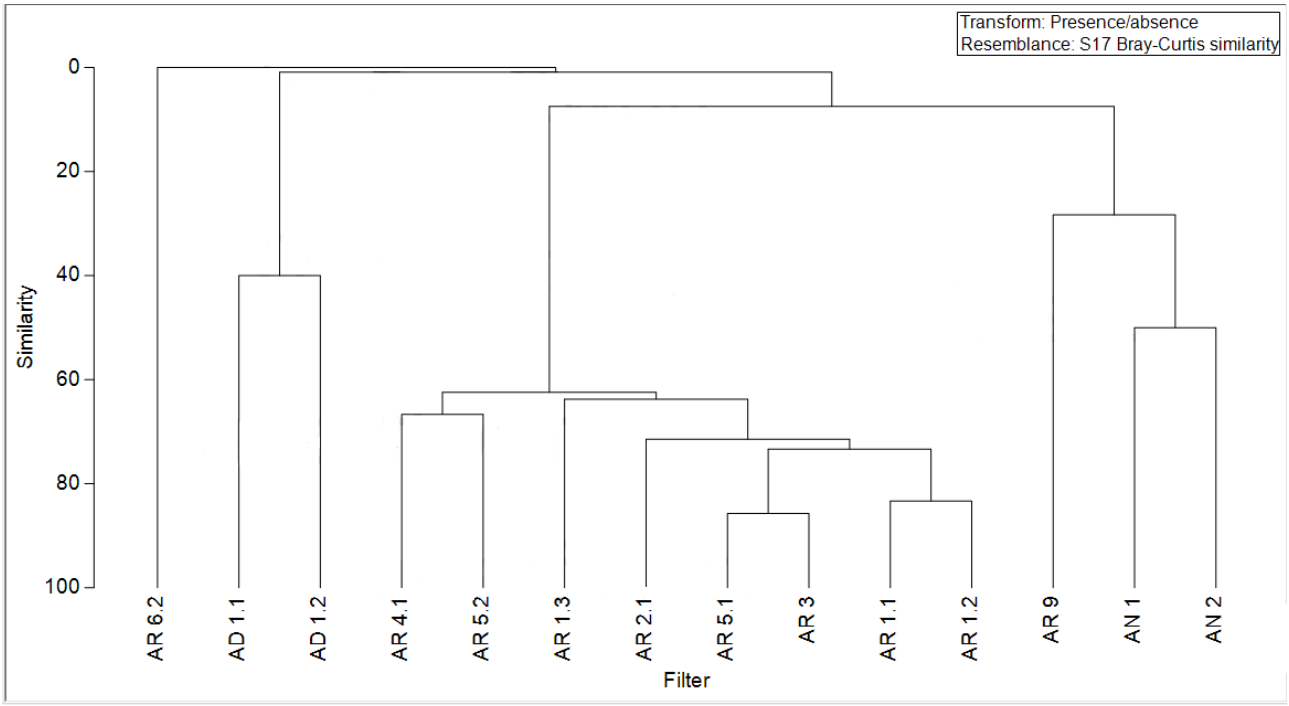


Figure 4.22. The dendrogram indicates the similarity of the species diversity found in the different filters according to the results obtained from GenBank.

A dendrogram that shows the similarity in species diversity between the PCR replicates was also constructed (Figure 4.23). 3-4, 3-5, and 3-6 are the PCR replicates of the negative control. PCR replicates 5-1 and 5-2 have the exact same taxa (similarity percentage is 100%). Altough 2-4 is the PCR replicate of the sample from filter AR 2.1 and 2-10 is the PCR replicate of the sample obtained from the filter AR 5.1, both PCR replicates involved identical taxa.

The PCR replicates of eDNA samples collected from Nansen Island (4-1, 4-2, 4-3, 4-4, 4-5 and 4-6) were clustered together. 4-3, 4-4 and 4-6 show the 100% similarity in species diversity. 4-4 and 4-6 are from the filter AN 2, but 4-3 is from the filter AN 1.

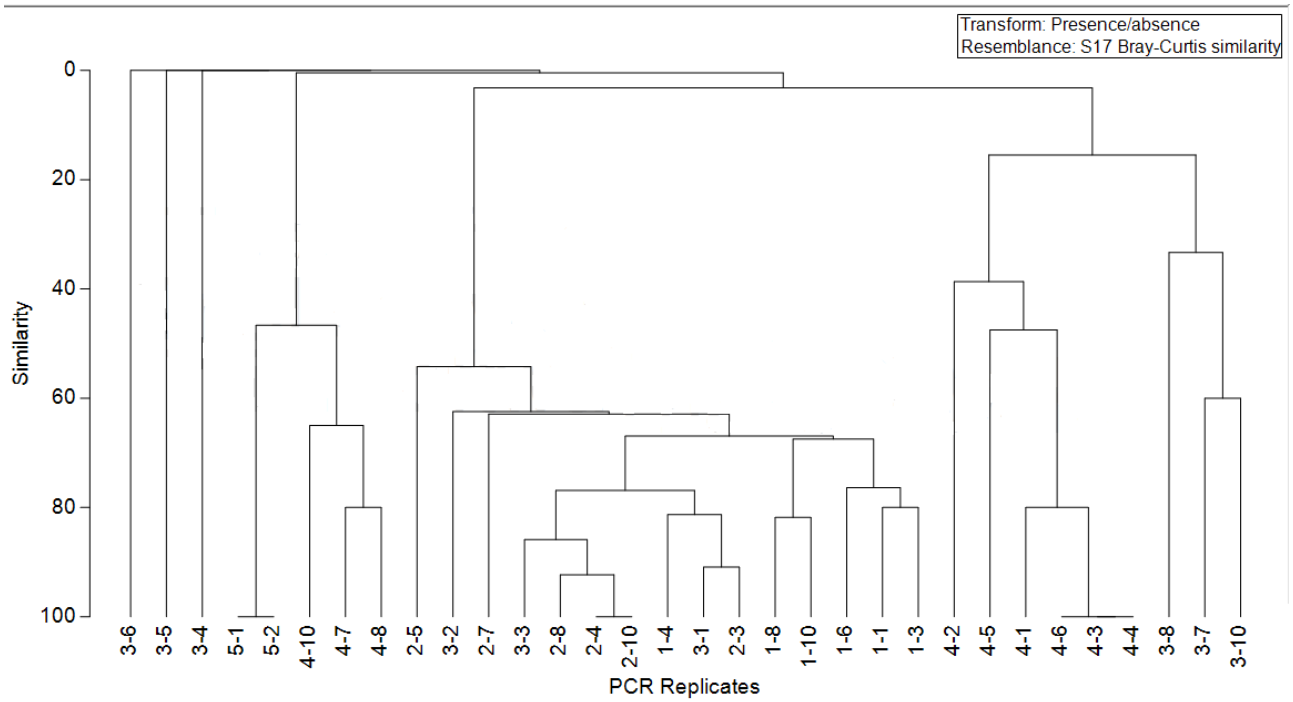


Figure 4.23. The dendrogram indicates the similarity of the species diversity found in the different PCR replicates according to the results obtained from GenBank.

GenBank includes sequences of all kinds of species and is the most widely used database for biodiversity analysis. Thus, observing a relatively higher number of taxa in GenBank, when compared to the other databases is expected. The identified taxa were classified into 11 algae, eight seaweeds, three fish, five marine worms, two sponges, three shellfish, a water bear, a copepod, a jellyfish, and a bacteria species. The total read number of algae and seaweeds is 3152, which equals about 90% of total reads. This finding displays a relatively high abundance of algal species in our data set.

Since we used the CO1 gene, a mitochondrial gene, in our study, we expect only eukaryotic organisms to be detected. However, we see that one bacterial species was found in the data obtained from the GenBank database. There could be several different reasons for this unexpected result. The very low number of sequence reads of *Bacillus infantis* (3) supports that this may be a PCR error. Most primers targeted to amplify the CO1 region can also amplify many prokaryotic genes (Zafeiropoulos et al., 2021; Collins et al., 2019). In addition, a misidentification in the GenBank database or not taking into account the query coverage value in the BLAST result may have caused the detection of a prokaryotic organism.

4.6. Comparison of Species Identification in MIDORI, WoRMS and GenBank

All data acquired from the three databases (MIDORI, WoRMS, and GenBank) were combined to see how many different species have been detected cumulatively. The number of unique taxa identified using MIDORI, WoRMS, and GenBank databases are 8, 18, and 36, respectively. In total, 23 species have been detected in at least two databases. Hence, there are 39 unique taxa in total, according to the combined information of the three databases. There was only one species common across the three databases: *Ctenocalanus citer*, a copepod. The scientific names of taxa and their distribution corresponding to the databases can be seen in the following Venn diagram (Figure 4.24).

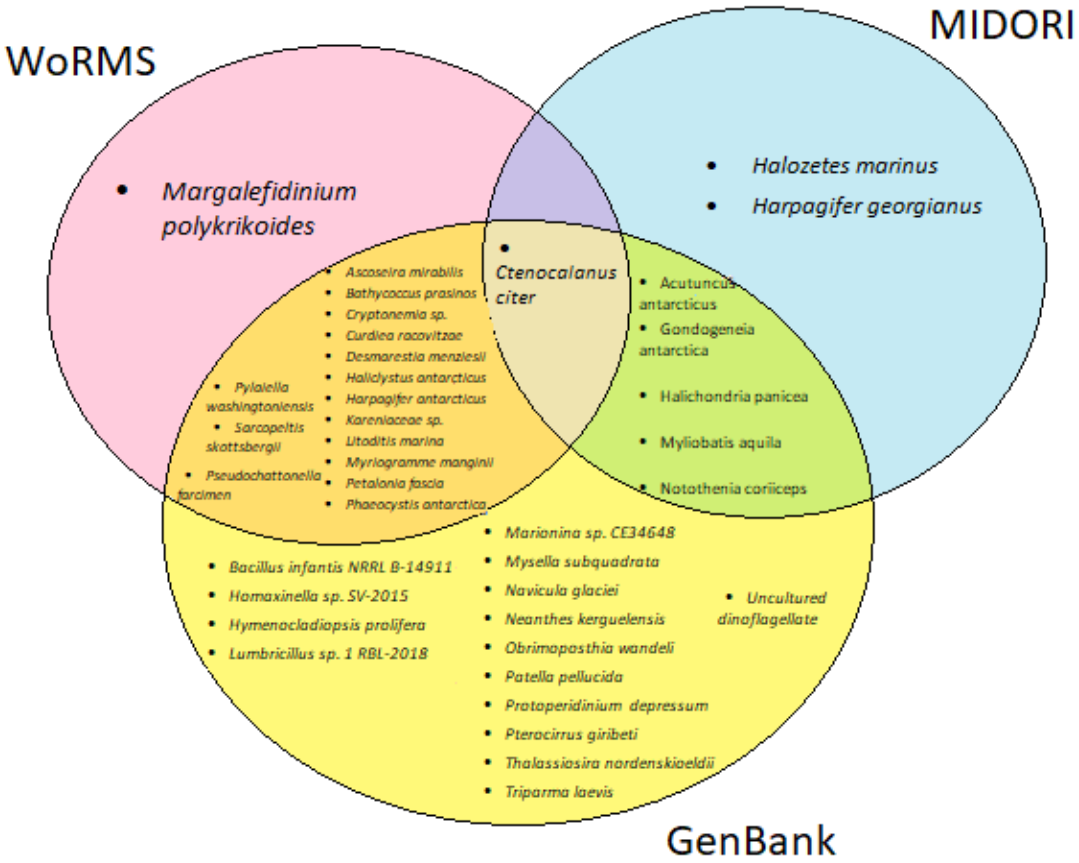


Figure 4.24. The Venn diagram shows the distribution of the taxa identified in the total dataset according to the three databases.

A stacked column chart was constructed to observe the relative number of reads of sequences of unique taxa found in the three databases separately. The graph in Figure 4.25 demonstrates the similarities and helps to compare the differences in the three databases.

When evaluated in terms of both species diversity and read numbers, the data from the WoRMS and GenBank databases showed similarity, but the MIDORI database has a quite different distribution from these two. According to the graph, 55% of the total reads are *Halozetes marinus*, and 17% of the total reads are *Harpagifer georgianus* in MIDORI. For WoRMS and GenBank databases, the frequencies are similar since they have 16 species in common. About 27% of the total reads are *Petalonia fascia* in both databases. Moreover, *Bathycoccus prasinos* makes up about 15% of the total data, and *Sarcopeltis skottsbergii* represents about 14% of the total reads in both databases.



Figure 4.25. The stacked column graph demonstrates the read counts of species found in MIDORI, WoRMS, and GenBank databases.

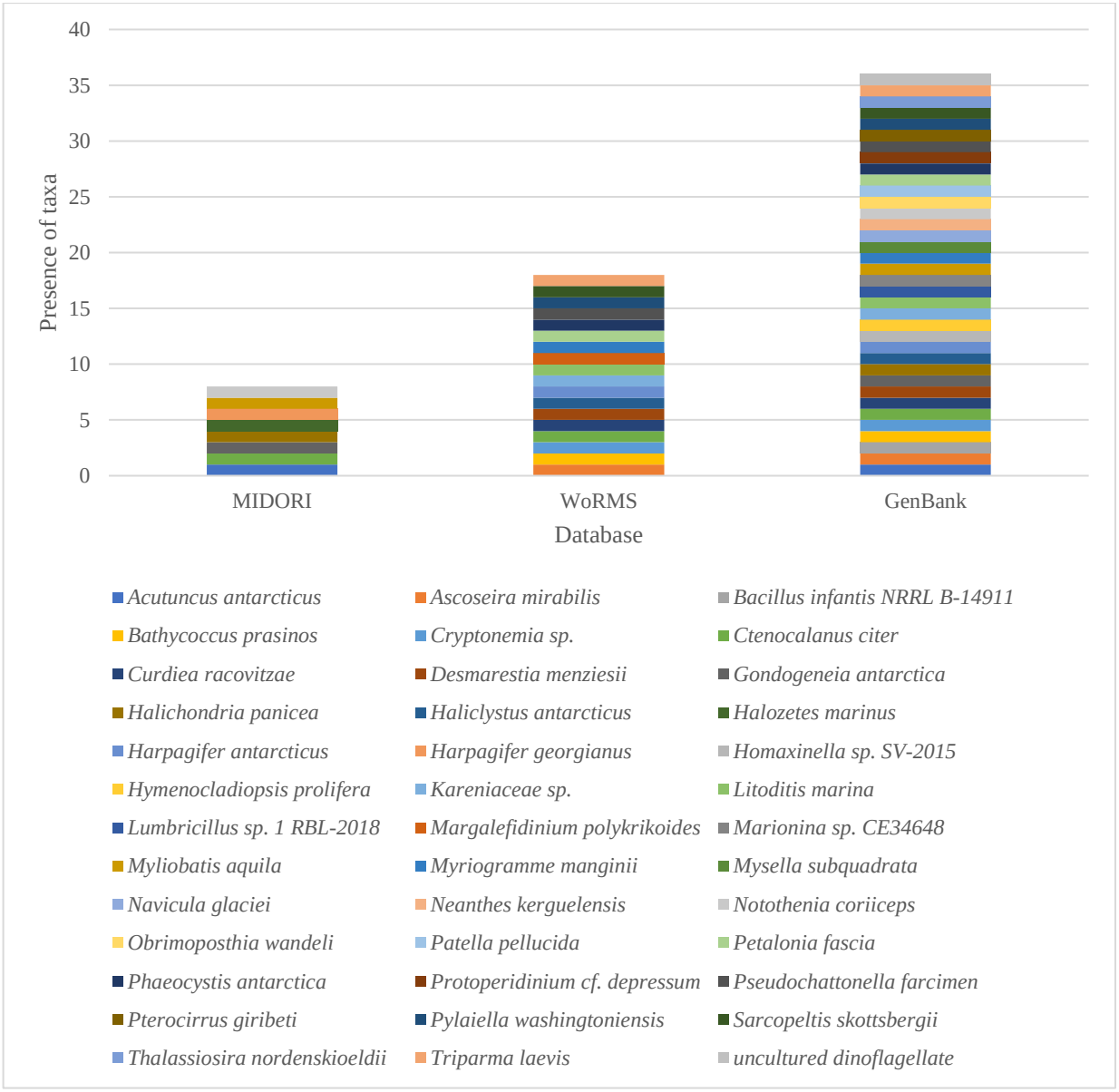


Figure 4.26. The stacked column graph displays the presence of taxa found in MIDORI, WoRMS, and GenBank databases.

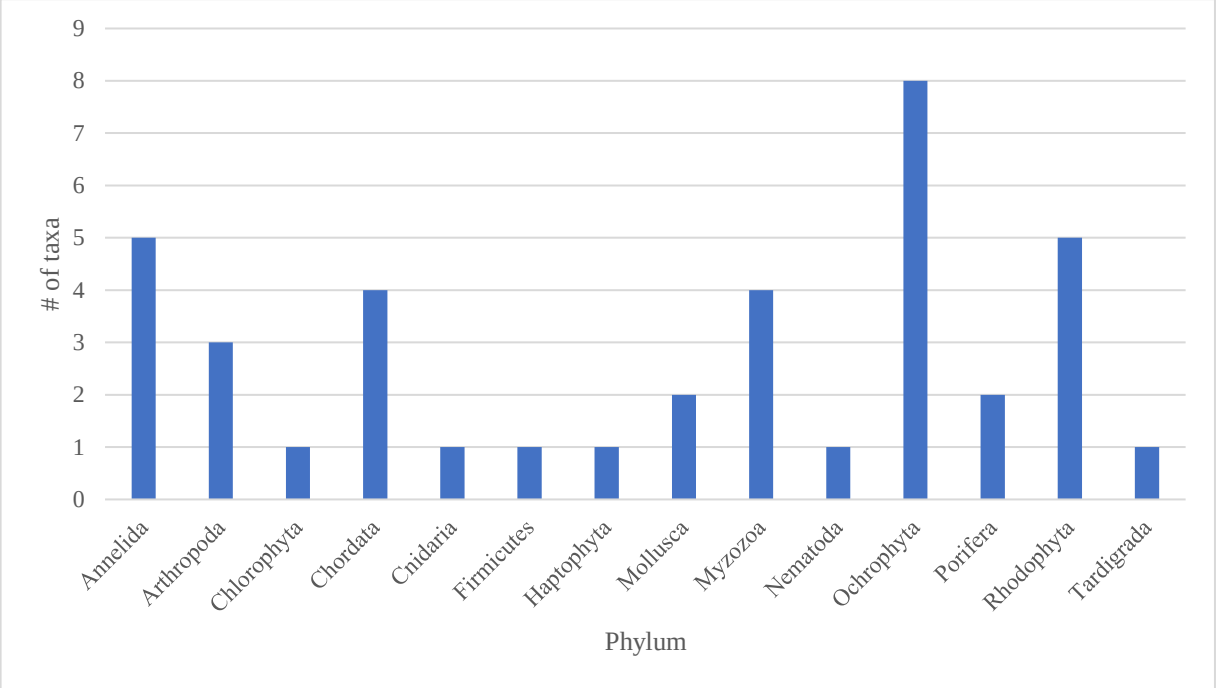


Figure 4.27. The column graph shows the number of different taxa contained in the phylum detected, according to the data obtained from the three databases: MIDORI, WoRMS, and GenBank.

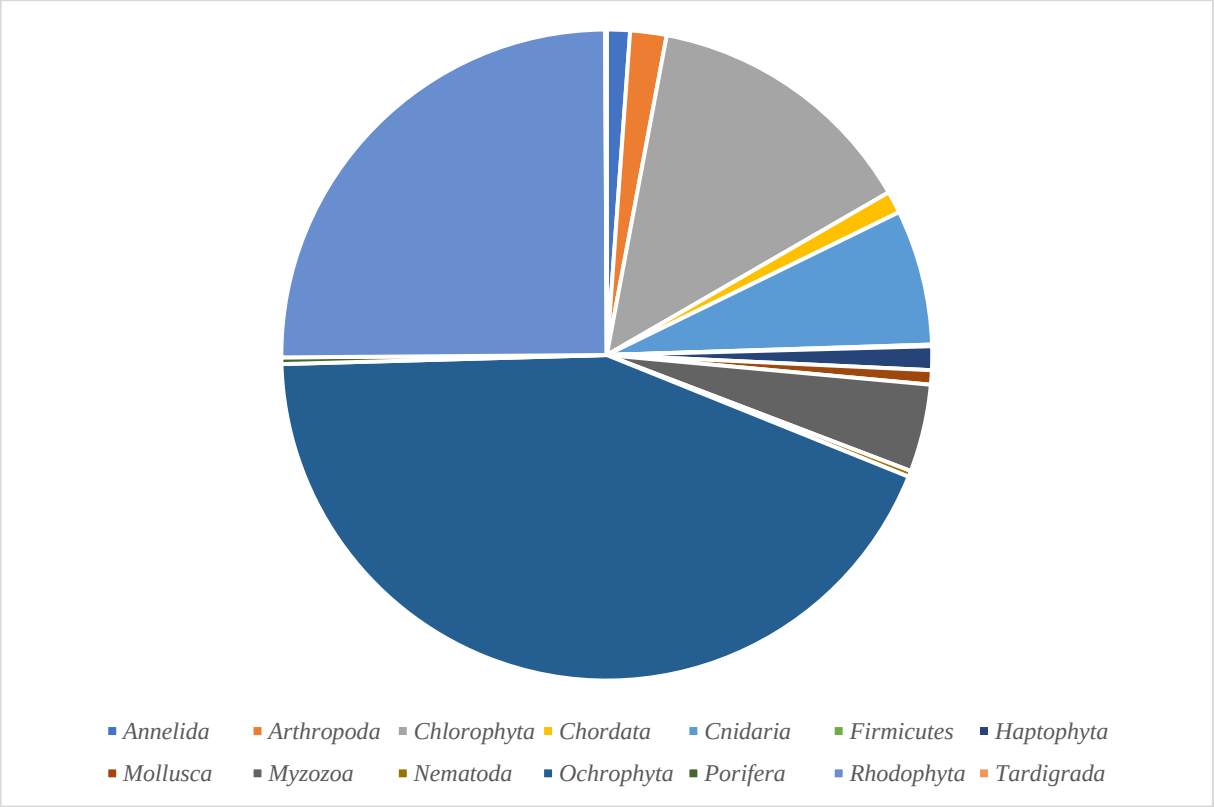


Figure 4.28. The pie chart displaying the number of sequence reads of each phylum found in the three databases; MIDORI, WoRMS, and GenBank.

39 different taxa were identified in the data collected from three databases. The most diverse taxa were identified in GenBank database (36 taxa), which is reliable since the GenBank database contains sequence information of a wide array of species such as Eukaryota, Prokaryota, etc. (Leray et al., 2019). In addition, again as expected, the least diverse taxa assemblage were obtained from MIDORI (eight taxa) since the MIDORI database is designed for alignment and taxonomic assignment of metazoan species (Machida et al., 2017). In WoRMS database, 18 taxa were identified, which is consistent with the fact that WoRMS database involves taxonomic information on marine species (M. J. Costello et al., 2013). When all taxa were grouped among themselves, the following result was obtained: 12 algae, eight seaweeds, four fish, five marine worms, two sponges, three shellfish, a water bear, a copepod, a jellyfish, a mite, and a bacteria species. These species belonged to 14 phyla; Ochrophyta, Annelida, and Rhodophyta were the most found phyla in our data set. On the other hand, the most abundant phyla were Ochrophyta, Rhodophyta, and Chlorophyta, regarding the number of sequence reads. The reason why the most abundant species are algal organisms can be because of the presence of an algal bloom in the Antarctic Peninsula when the eDNA samples were collected. The term algal bloom refers to the overgrowth of algae and the rapid increase of the algae population as substances such as nitrogen or phosphorus enter the water system. When looking at the distribution of the taxa found in each database according to the locations, it is seen that no species have been detected in Nansen Island and offshore Robert Island in MIDORI. This result is reasonable because the data obtained from WoRMS and GenBank demonstrate that the taxa found in Nansen Island and offshore Robert Island belong to groups of algae and seaweeds.

As mentioned earlier, we used a trained RDP Classifier algorithm in MIDORI software with the MIDORI CO1 reference database, but we used the BLAST+ alignment method with WoRMS and GenBank databases. In a study conducted to compare RDP Classifier and BLAST hit method for CO1 data of Arthropods and Chordates, it was found that the RDP classifier is faster than the top BLAST hits approach and shows a much lower false positive rate (Porter & Hajibabaei, 2018). They demonstrate that RDP Classifier gives more accurate results than the widely used top BLAST hits method for taxonomic assignment of CO1 metabarcodes. In this study, we identified eight animal species in MIDORI with the RDP Classifier program but 16 animal species in GenBank with the BLAST hit method (all species are also found in WoRMS with the BLAST hits approach). There were six species in common in both databases. In contrast to the findings of Porter, we can not claim that the RDP Classifier gives more accurate results than the BLAST. The number of sequence reads was significantly lower in RDP Classifier than in BLAST+ output. A low number of sequence reads is also not ideal for increased reliability of the RDP Classifier.

One of the most interesting findings of our study was the identification of *Sarcopeltis skottsbergii* in WoRMS and GenBank databases. *Sarcopeltis skottsbergii* is an endemic organism and has been found only in southern of South American waters. *Sarcopeltis skottsbergii* was only detected in Carlota and Coppermine Coves on Robert Island, according to two databases. The fact that Robert Island is close to the southern coast of South America supports the idea that this species may have spread to this location by water currents or shipping activities. In addition to all these possibilities, since this is an eDNA study and we do not have any observational information, it is also likely not this species itself, but its DNA was carried to Antarctic Peninsula by water currents. The fact that the number of sequence reads of *Sarcopeltis skottsbergii* is significantly higher than the other taxa in the dataset indicates that this is not a PCR error or a false positive error caused by the BLAST+. Detection of *Bacillus infantis,* a bacterium, in GenBank is also unexpected. CO1 barcode region is a mitochondrial gene, meaning the bacteria do not have this gene. Thus, it is a false positive result probably derived from a PCR error or insufficiency of filtering the data.

Apart from *Bacillus infantis*, 13 taxa; *Bathycoccus prasinos* (algae), *Cryptonemia sp.* (seaweed), *Halichondria panicea* (sponge), *Halozetes marinus* (mite), *Kareniaceae sp.* (algae), *Litoditis marina* (algae), *Marionina sp.* (worm), *Myliobatis aquila* (fish), *Patella pellucida* (shellfish), *Protoperidinium depressum* (algae), *Pseudochattonella farcimen* (algae), *Pylaiella washingtoniensis* (seaweed), and *Thalassiosira nordenskioeldii* (algae) had not been documented before in Antarctica when their geographic distributions were checked in the WoRMS, AlgaeBase, and Ocean Biodiversity Information System (OBIS) databases (Table 4.13). Considering the geographical distribution of these 13 species, *Halozetes marinus*, *Litoditis marina, Protoperidinium depressum*, and *Thalassiosira nordenskioeldii* were documented in the regions close to Antarctica so that their DNA molecules can spread through the water currents into Antarctic Peninsula or they could have existed in Antarctica without having been detected previously. The rest of the species have not been detected in the areas close to Antarctica. Among these, *Myliobatis aquila* is widespread around the world, however its southernmost distribution is South Africa. *Patella pellucida* and *Pseudochattonella farcimen* were documented only in the North Sea, and *Pylaiella washingtoniensis* was documented only in the North American waters. It would not be expected for the DNA of these species to naturally arrive in Antarctica, and their detection in this study could potentially indicate the movement of their DNA through ballast waters of ships or the first evidence of their existence in Antartica, pending confirmation. In addition, no information about the location distribution of *Pterocirrus giribeti* (worm) could be found in the three databases. It should also be noted that, although *Sarcopeltis skottsbergii* is considered as an endemic species in South America (Castro-Varela et al., 2022), it has been recorded before in Antarctica according to the AlgaeBase.

Table 4.13. The table shows whether the 39 species identified in this study were previously recorded in Antarctica.

Species	Has it been recorded in Antarctica before?	Database
Acutuncus antarcticus	Yes	OBIS
Ascoseira mirabilis	Yes	AlgaeBase
Bacillus infantis	--	--
Bathycoccus prasinos	No	AlgaeBase
Cryptonemia sp.	No	--
Ctenocalanus citer	Yes	WoRMS
Curdiea racovitzae	Yes	AlgaeBase
Desmarestia menziesii	Yes	AlgaeBase
Gondogeneia antarctica	Yes	WoRMS
Halichondria panicea	No	OBIS
Haliclystus antarcticus	Yes	WoRMS
Halozetes marinus	No	--
Harpagifer antarcticus	Yes	OBIS
Harpagifer georgianus	Yes	OBIS
Homaxinella sp.	Yes	OBIS
Hymenocladiopsis prolifera	Yes	AlgaeBase
Kareniaceae sp.	No	--
Litoditis marina	No	--
Lumbricillus sp.	Yes	Wo
Margalefidinium polykrikoides	Yes	AlgaeBase
Marionina sp.	No	--
Myliobatis aquila	No	--
Myriogramme manginii	Yes	AlgaeBase
Mysella subquadrata	Yes	OBIS
Navicula glaciei	Yes	AlgaeBase
Neanthes kerguelensis	Yes	WoRMS
Notothenia coriiceps	Yes	OBIS
Obrimoposthia wandeli	Yes	WoRMS
Patella pellucida	No	--
Petalonia fascia	Yes	AlgaeBase

Phaeocystis antarctica	Yes	AlgaeBase
Protoperidinium depressum	No	--
Pseudochattonella farcimen	No	--
Pterocirrus giribeti	--	--
Pylaiella washingtoniensis	No	--
Sarcopeltis skottsbergii	Yes	AlgaeBase
Thalassiosira nordenskioeldii	No	--
Triparma laevis	Yes	AlgaeBase
Uncultured dinoflagellate	--	--

5. LIMITATIONS OF THE STUDY

eDNA studies have many limitations due to the fact that the method is novel and relatively few studies has been undertaken on this method. Unfortunately, it was inevitable that these limitations would also affect our study. The most important of these limits is that the eDNA samples do not stay long in the environment and degrade in about two weeks, thus, the biodiversity we found only gives us the information of the species present at the time of sampling. Therefore, the environmental conditions such as water temperature, salinity, acidity, and nutrients at the time the samples were collected also affect both the eDNA samples and the species detected in the environment. For example, the presence of an algal bloom in Antarctica at the time of sampling caused algae species to dominate in our study. Therefore, the results found should be evaluated according to that period of the year. As a second limitation, we can point to the sequencing method, having used the Illumina Miseq platform, even though providing longer reads, resulted in a lower number of reads *per* sample. The relatively low number of sequences potentially affects the reliability of the study. Therefore, the use of a sequencing machine like Illumina Nextseq or Novaseq that can generate more reads could increase the resolution of the study.

6. CONCLUSIONS AND RECOMMENDATIONS

In this study, we identified species diversity in Antarctic Peninsula using eDNA metabarcoding technique and compared the results obtained from three databases, MIDORI, WoRMS, and GenBank. We found a total of 39 taxa, and out of these 39 taxa, 34 are identified at the species level, three to the genus level, one to the family level, and one to the class level. We identified 18 animal species when the data of three databases were combined, but our samples primarily include planktonic algae and seaweeds corresponding to the sequence reads. Based on our findings, the finding of 14 new taxa in Antarctica comprises the first records for the continent. According to the results, Coppermine and Carlota Coves on Robert Island are the regions with the most species diversity.

Despite the fact that all three databases contain information about the sequences of taxa belonging to different groups, it is a problem that only one species appears in common in all three datasets. Besides, 166 unidentified OTUs in MIDORI and 179 unidentified OTUs in WoRMS and GenBank were calculated. The fact that the species in the three databases produce such different results and the high number of unidentified OTUs shows that the CO1 reference databases are limited in their taxonomic coverage and do not contain sufficient information for Antarctica. In addition, when we examined in detail the non-common taxa in the three databases and which sequence they correspond to, we detected that some taxa have the same sequence but are named differently in different databases. For instance, *Harpagifer georgianus* and *Harpagifer antarcticus* were identified in MIDORI and GenBank databases, respectively, but they have the same sequence. Similarly, a sequence corresponding to a species named *Margalefidinium polykrikoides* in the WoRMS database is named *Uncultured dinoflagellate* in the GenBank database. All these contradictions demonstrate the inconsistencies across these popular reference databases.

Assigning taxonomic information to the thousands of sequences produced by high-throughput sequencing is difficult because many DNA reference libraries are deficient in information about particular groups of organisms and may contain incorrect sequences. A combination of several reference databases would be promising in terms of taxonomic coverage and reliability of results. Moreover, the query coverage value should also be taken into account, and an acceptable threshold should be applied in blast searches, besides the percent identity value, to avoid false positive results. eDNA studies are often like a snapshot of the moment the samples were collected. Therefore, eDNA samples should be collected from the same region at different times of the year and from different depths to conduct a more comprehensive biodiversity analysis in general, and this could also be

beneficial for this study. In our study, we used the short region of the CO1 gene as a barcode but using the long region of the CO1 gene as a barcode will increase the accuracy of the results. In addition, using multiple barcode genes in a study is recommended to increase the number of identified species, which might comprise the next step of this study.

REFERENCES

Aglieri, G., Baillie, C., Mariani, S., Cattano, C., Calò, A., Turco, G., Spatafora, D., Di Franco, A., Di Lorenzo, M., Guidetti, P., & Milazzo, M. (2020). Environmental DNA effectively captures functional diversity of coastal fish communities. *Molecular Ecology*, *August*, 1–13. https://doi.org/10.1111/mec.15661

Andrews, S. (2010). *Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data*. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Antich, A., Palacín, C., Cebrian, E., Golo, R., Wangensteen, O. S., & Turon, X. (2020). Marine biomonitoring with eDNA: Can metabarcoding of water samples cut it as a tool for surveying benthic communities? *Molecular Ecology*. https://doi.org/10.1111/mec.15641

Atherton, S., & Jondelius, U. (2020). Biodiversity between sand grains: Meiofauna composition across southern and western Sweden assessed by metabarcoding. *Biodiversity Data Journal*, *8*, 1–40. https://doi.org/10.3897/BDJ.8.E51813

Banchi, E., Ametrano, C. G., Tordoni, E., Stanković, D., Ongaro, S., Tretiach, M., Pallavicini, A., Muggia, L., Verardo, P., Tassan, F., Trobiani, N., Moretti, O., Borney, M. F., & Lazzarin, S. (2020). Environmental DNA assessment of airborne plant and fungal seasonal diversity. *Science of the Total Environment*. https://doi.org/10.1016/j.scitotenv.2020.140249

Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., & Courchamp, F. (2012). Impacts of climate change on the future of biodiversity. *Ecology Letters*, *15*(4), 365–377. https://doi.org/10.1111/j.1461-0248.2011.01736.x

Blowes, S. A., Supp, S. R., Antão, L. H., Bates, A., Bruelheide, H., Chase, J. M., Moyes, F., Magurran, A., McGill, B., Myers-Smith, I. H., Winter, M., Bjorkman, A. D., Bowler, D. E., Byrnes, J. E. K., Gonzalez, A., Hines, J., Isbell, F., Jones, H. P., Navarro, L. M., … Dornelas, M. (2019). The geography of biodiversity change in marine and terrestrial assemblages. *Science*, *366*(6463), 339–345. https://doi.org/10.1126/science.aaw1620

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., Yu, D. W., &

de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. In *Trends in Ecology and Evolution* (Vol. 29, Issue 6, pp. 358–367). Elsevier Ltd. https://doi.org/10.1016/j.tree.2014.04.003

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/BIOINFORMATICS/BTU170

Borrell, Y. J., Miralles, L., Do Huu, H., Mohammed-Geba, K., & Garcia-Vazquez, E. (2017). DNA in a bottle - Rapid metabarcoding survey for early alerts of invasive species in ports. *PLoS ONE*, *12*(9), 1–17. https://doi.org/10.1371/journal.pone.0183347

Bovo, S., Ribani, A., Utzeri, V. J., Schiavo, G., Bertolini, F., & Fontanesi, L. (2018). Shotgun metagenomics of honey DNA: Evaluation of a methodological approach to describe a multi-kingdom honey bee derived environmental DNA signature. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0205575

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*(1), 176–182. https://doi.org/10.1111/1755-0998.12428

Brandt, A., Gooday, A. J., Brandão, S. N., Brix, S., Brökeland, W., Cedhagen, T., Choudhury, M., Cornelius, N., Danis, B., De Mesel, I., Diaz, R. J., Gillan, D. C., Ebbe, B., Howe, J. A., Janussen, D., Kaiser, S., Linse, K., Malyutina, M., Pawlowski, J., … Vanreusel, A. (2007). First insights into the biodiversity and biogeography of the Southern Ocean deep sea. *Nature 2007 447:7142*, *447*(7142), 307–311. https://doi.org/10.1038/nature05827

Cahill, A. E., Pearman, J. K., Borja, A., Carugati, L., Carvalho, S., Danovaro, R., Dashfield, S., David, R., Féral, J. P., Olenin, S., Šiaulys, A., Somerfield, P. J., Trayanova, A., Uyarra, M. C., & Chenuil, A. (2018). A comparative analysis of metabarcoding and morphology-based identification of benthic communities across different regional seas. *Ecology and Evolution*, *8*(17), 8908–8920. https://doi.org/10.1002/ece3.4283

Câmara, P. E. A. S., Eisenlohr, P. V., Coelho, L. C., Carvalho-Silva, M., Amorim, E. T., Convey, P., Pinto, O. H. B., & Rosa, L. H. (2021). Fairy ring disease affects epiphytic algal assemblages

associated with the moss Sanionia uncinata (Hedw.) Loeske (Bryophyta) on King George Island, Antarctica. *Extremophiles*, *25*(5–6), 501–512. https://doi.org/10.1007/S00792-021-01246-9

Canini, F., Geml, J., D'Acqui, L. P., Selbmann, L., Onofri, S., Ventura, S., & Zucconi, L. (2020). Exchangeable cations and pH drive diversity and functionality of fungal communities in biological soil crusts from coastal sites of Victoria Land, Antarctica. *Fungal Ecology*, *45*. https://doi.org/10.1016/J.FUNECO.2020.100923

Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., Narwani, A., MacE, G. M., Tilman, D., Wardle, D. A., Kinzig, A. P., Daily, G. C., Loreau, M., Grace, J. B., Larigauderie, A., Srivastava, D. S., & Naeem, S. (2012). Biodiversity loss and its impact on humanity. *Nature*, *486*(7401), 59–67. https://doi.org/10.1038/nature11148

Castro-varela, P., Celis-pla, P. S. M., Figueroa, F. L., & Rubilar, M. (2022). *Highly Ef fi cient Water-Based Extraction of Biliprotein R- Phycoerythrin From Marine the Red- Macroalga Sarcopeltis skottsbergii by Ultrasound and High-Pressure Homogenization Methods*. *9*(June), 1–14. https://doi.org/10.3389/fmars.2022.877177

Cheang, C. chiu, Lee, B. yee, Ip, B. H. yeung, Yiu, W. hong, Tsang, L. ming, & Ang, P. O. (2020). Fish and crustacean biodiversity in an outer maritime estuary of the Pearl River Delta revealed by environmental DNA. *Marine Pollution Bulletin*, *161*(October), 111707. https://doi.org/10.1016/j.marpolbul.2020.111707

Chown, S. L., Clarke, A., Fraser, C. I., Cary, S. C., Moon, K. L., & McGeoch, M. A. (2015). The changing form of Antarctic biodiversity. *Nature 2015 522:7557*, *522*(7557), 431–438. https://doi.org/10.1038/nature14505

Clarke, A. (2008). Antarctic marine benthic diversity: patterns and processes. *Journal of Experimental Marine Biology and Ecology*, *366*(1–2), 48–55. https://doi.org/10.1016/J.JEMBE.2008.07.008

Coissac, E., Riaz, T., & Puillandre, N. (2012). Bioinformatic challenges for DNA metabarcoding of plants and animals. *Molecular Ecology*, *21*(8), 1834–1847. https://doi.org/10.1111/j.1365-294X.2012.05550.x

Coleine, C., Albanese, D., Onofri, S., Tringe, S. G., Pennacchio, C., Donati, C., Stajich, J. E., & Selbmann, L. (2020). Metagenomes in the Borderline Ecosystems of the Antarctic Cryptoendolithic Communities. *Microbiology Resource Announcements*, *9*(10). https://doi.org/10.1128/MRA.01599-19

Coleine, C., Stajich, J. E., Zucconi, L., Onofri, S., Pombubpa, N., Egidi, E., Franks, A., Buzzini, P., & Selbmann, L. (2018). Antarctic cryptoendolithic fungal communities are highly adapted and dominated by Lecanoromycetes and Dothideomycetes. *Frontiers in Microbiology*, *9*(JUN). https://doi.org/10.3389/FMICB.2018.01392

Collins, R. A., Bakker, J., Wangensteen, O. S., Soto, A. Z., Corrigan, L., Sims, D. W., Genner, M. J., & Mariani, S. (2019). Non-specific amplification compromises environmental DNA metabarcoding with COI. *Methods in Ecology and Evolution*, *10*(11), 1985–2001. https://doi.org/10.1111/2041-210X.13276

Costello, K. E., Lynch, S. A., Mcallen, R., Riordan, R. M. O., & Culloty, S. C. (2022). Assessing the potential for invasive species introductions and secondary spread using vessel movements in maritime ports. *Marine Pollution Bulletin*, *177*(December 2021), 113496. https://doi.org/10.1016/j.marpolbul.2022.113496

Costello, M. J., Bouchet, P., Boxshall, G., Fauchald, K., Gordon, D., Walter, T. C., Vanhoorne, B., Decock, W., & Appeltans, W. (2013). *Global Coordination and Standardisation in Marine Biodiversity through the World Register of Marine Species (WoRMS) and Related Databases*. *8*(1). https://doi.org/10.1371/journal.pone.0051629

Czechowski, P., Clarke, L., Cooper, A., & Stevens, M. (2017). Ground-truthing Phylotype Assignments for Antarctic Invertebrates. *DNA Barcodes*, *5*(1), 1–13. https://doi.org/10.1515/DNA-2017-0001

Davey, M. P., Norman, L., Sterk, P., Huete-Ortega, M., Bunbury, F., Loh, B. K. W., Stockton, S., Peck, L. S., Convey, P., Newsham, K. K., & Smith, A. G. (2019). Snow algae communities in Antarctica: metabolic and taxonomic composition. *New Phytologist*, *222*(3), 1242–1255. https://doi.org/10.1111/nph.15701

De Broyer, C., P. Koubbi, H.J. Griffiths, B. Raymond, C. Udekem d'Acoz, A.P. Van de Putte,

B. Danis, B. David, S. Grant, J. Gutt, C. Held, G. Hosie, F. Huettmann, A. P. and Y. R. (Eds). (2014). *Biogeographic Atlas of the Southern Ocean*. Scientific Committee on Antarctic Research, Cambridge.

Deagle, B. E., Clarke, L. J., Kitchener, J. A., Polanowski, A. M., & Davidson, A. T. (2018). Genetic monitoring of open ocean biodiversity: An evaluation of DNA metabarcoding for processing continuous plankton recorder samples. *Molecular Ecology Resources*, *18*(3), 391–406. https://doi.org/10.1111/1755-0998.12740

Dias, P. J., Fotedar, S., Munoz, J., Hewitt, M. J., Lukehurst, S., Hourston, M., Wellington, C., Duggan, R., Bridgwood, S., Massam, M., Aitken, V., de Lestang, P., McKirdy, S., Willan, R., Kirkendale, L., Giannetta, J., Corsini-Foka, M., Pothoven, S., Gower, F., … Snow, M. (2017). Establishment of a taxonomic and molecular reference collection to support the identification of species regulated by the Western Australian Prevention List for Introduced Marine Pests. *Management of Biological Invasions*, *8*(2), 215–225. https://doi.org/10.3391/mbi.2017.8.2.09

Djurhuus, A., Pitz, K., Sawaya, N. A., Rojas-Márquez, J., Michaud, B., Montes, E., Muller-Karger, F., & Breitbart, M. (2018). Evaluation of marine zooplankton community structure through environmental DNA metabarcoding. *Limnology and Oceanography: Methods*, *16*(4), 209–221. https://doi.org/10.1002/lom3.10237

Edgar, R. C. (2010). *Search and clustering orders of magnitude faster than BLAST*. *26*(19), 2460–2461. https://doi.org/10.1093/bioinformatics/btq461

Elbrecht, V., & Leese, F. (2017). Validation and development of COI metabarcoding primers for freshwater macroinvertebrate bioassessment. *Frontiers in Environmental Science*, *5*(APR), 1–11. https://doi.org/10.3389/fenvs.2017.00011

Fraser, C. I., Connell, L., Lee, C. K., & Cary, S. C. (2018). Evidence of plant and animal communities at exposed and subglacial (cave) geothermal sites in Antarctica. *Polar Biology*, *41*(3), 417–421. https://doi.org/10.1007/s00300-017-2198-9

Gielings, R., Fais, M., Fontaneto, D., Creer, S., Costa, F. O., Renema, W., & Macher, J. N. (2021). DNA Metabarcoding Methods for the Study of Marine Benthic Meiofauna: A Review. *Frontiers in Marine Science*, *8*(September), 1–13. https://doi.org/10.3389/fmars.2021.730063

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A.,

Spear, S. F., McKee, A., Oyler-McCance, S. J., Cornman, R. S., Laramie, M. B., Mahon, A. R., Lance, R. F., Pilliod, D. S., Strickler, K. M., Waits, L. P., Fremier, A. K., Takahara, T., Herder, J. E., & Taberlet, P. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, *7*(11), 1299–1307. https://doi.org/10.1111/2041-210X.12595

Grant, R. A., & Linse, K. (2009). Barcoding Antarctic biodiversity: Current status and the caml initiative, a case study of Marine invertebrates. *Polar Biology*, *32*(11), 1629–1637. https://doi.org/10.1007/S00300-009-0662-X

Günther, B., Knebelsberger, T., Neumann, H., Laakmann, S., & Martínez Arbizu, P. (2018). Metabarcoding of marine environmental DNA based on mitochondrial and nuclear genes. *Scientific Reports*, *8*(1), 1–13. https://doi.org/10.1038/s41598-018-32917-x

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218

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

Jeunen, G. J., Knapp, M., Spencer, H. G., Lamare, M. D., Taylor, H. R., Stat, M., Bunce, M., & Gemmell, N. J. (2019). Environmental DNA (eDNA) metabarcoding reveals strong discrimination among diverse marine habitats connected by water movement. *Molecular Ecology Resources*, *19*(2), 426–438. https://doi.org/10.1111/1755-0998.12982

Lacerda, A. L. d. F., Proietti, M. C., Secchi, E. R., & Taylor, J. D. (2020). Diverse groups of fungi are associated with plastics in the surface waters of the Western South Atlantic and the Antarctic Peninsula. *Molecular Ecology*, *29*(10), 1903–1918. https://doi.org/10.1111/MEC.15444

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

Leray, M., Knowlton, N., Ho, S., Nguyen, B. N., & Machida, R. J. (2019). *GenBank is a reliable resource for 21st century biodiversity research*. *116*(45), 22651–22656. https://doi.org/10.1073/pnas.1911714116

Levine, R. E., Zhang, Y., Leng, Y., Snow, D. D., Cassada, D., Durso, L. M., & Li, X. (2019). Microbial Transformation of A Sulfonamide Antibiotic Under Various Background Nutrient Conditions. *Bulletin of Environmental Contamination and Toxicology*, *103*(6), 808–813. https://doi.org/10.1007/s00128-019-02727-6

Liu, Y., Wikfors, G. H., Rose, J. M., McBride, R. S., Milke, L. M., & Mercaldo-Allen, R. (2019). Application of Environmental DNA Metabarcoding to Spatiotemporal Finfish Community Assessment in a Temperate Embayment. *Frontiers in Marine Science*, *6*(November). https://doi.org/10.3389/fmars.2019.00674

Machida, R. J., Leray, M., Ho, S.-L., Knowlton, N., & Characteristic Metazoa, S. (2017). *Data Descriptor: Metazoan mitochondrial gene sequence reference datasets for taxonomic assignment of environmental samples Background & Summary*. https://doi.org/10.1038/sdata.2017.27

Ogaki, M. B., Teixeira, D. R., Vieira, R., Lírio, J. M., Felizardo, J. P. S., Abuchacra, R. C., Cardoso, R. P., Zani, C. L., Alves, T. M. A., Junior, P. A. S., Murta, S. M. F., Barbosa, E. C., Oliveira, J. G., Ceravolo, I. P., Pereira, P. O., Rosa, C. A., & Rosa, L. H. (2020). Diversity and bioprospecting of cultivable fungal assemblages in sediments of lakes in the Antarctic Peninsula. *Fungal Biology*, *124*(6), 601–611. https://doi.org/10.1016/J.FUNBIO.2020.02.015

Ogram, A., Sayler, G. S., & Barkay, T. (1987). The extraction and purification of microbial DNA from sediments. *Journal of Microbiological Methods*, *7*(2–3), 57–66. https://doi.org/10.1016/0167-7012(87)90025-X

Pearman, J. K., Chust, G., Aylagas, E., Villarino, E., Watson, J. R., Chenuil, A., Borja, A., Cahill, A. E., Carugati, L., Danovaro, R., David, R., Irigoien, X., Mendibil, I., Moncheva, S., Rodríguez-Ezpeleta, N., Uyarra, M. C., & Carvalho, S. (2020). Pan-regional marine benthic cryptobiome biodiversity patterns revealed by metabarcoding Autonomous Reef Monitoring Structures. In *Molecular Ecology* (Vol. 29, Issue 24). https://doi.org/10.1111/mec.15692

Peck, J. R., & Welch, J. J. (2004). Adaptation and species range. *Evolution*, *58*(2), 211–221. https://doi.org/10.1111/J.0014-3820.2004.TB01639.X

Porter, T. M., & Hajibabaei, M. (2018). Automated high throughput animal CO1 metabarcode classification. *Scientific Reports*, *8*(1), 1–10. https://doi.org/10.1038/s41598-018-22505-4

Qu, W., Zhou, Y., Zhang, Y., Lu, Y., Wang, X., Zhao, D., Yang, Y., & Zhang, C. (2012). *program for checking PCR primer specificity*. *40*(June), 205–208. https://doi.org/10.1093/nar/gks552

Ransome, E., Geller, J. B., Timmers, M., Leray, M., Mahardini, A., Sembiring, A., Collins, A. G., & Meyer, C. P. (2017). The importance of standardization for biodiversity comparisons: A case study using autonomous reef monitoring structures (ARMS) and metabarcoding to measure cryptic diversity on Mo'orea coral reefs, French Polynesia. *PLoS ONE*, *12*(4), 1–19. https://doi.org/10.1371/journal.pone.0175066

Rippin, M., Lange, S., Sausen, N., & Becker, B. (2018). Biodiversity of biological soil crusts from the Polar Regions revealed by metabarcoding. *FEMS Microbiology Ecology*, *94*(4). https://doi.org/10.1093/FEMSEC/FIY036

Rondon, M. R., August, P. R., Bettermann, A. D., Brady, S. F., Grossman, T. H., Liles, M. R., Loiacono, K. A., Lynch, B. A., MacNeil, I. A., Minor, C., Tiong, C. L., Gilman, M., Osburne, M. S., Clardy, J., Handelsman, J., & Goodman, R. M. (2000). Cloning the soil metagenome: A strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and Environmental Microbiology*, *66*(6), 2541–2547. https://doi.org/10.1128/AEM.66.6.2541-2547.2000

Rosa, L. H., de Sousa, J. R. P., de Menezes, G. C. A., da Costa Coelho, L., Carvalho-Silva, M., Convey, P., & Câmara, P. E. A. S. (2020). Opportunistic fungi found in fairy rings are present on different moss species in the Antarctic Peninsula. *Polar Biology*, *43*(5), 587–596. https://doi.org/10.1007/S00300-020-02663-W

Ruppert, K. M., Kline, R. J., & 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. *Global Ecology and Conservation*, *17*, e00547.

https://doi.org/10.1016/j.gecco.2019.e00547

Sawaya, N. A., Djurhuus, A., Closek, C. J., Hepner, M., Olesin, E., Visser, L., Kelble, C., Hubbard, K., & Breitbart, M. (2019). Assessing eukaryotic biodiversity in the Florida Keys National Marine Sanctuary through environmental DNA metabarcoding. *Ecology and Evolution*, *9*(3), 1029–1040. https://doi.org/10.1002/ece3.4742

Schroeder, A., Stanković, D., Pallavicini, A., Gionechetti, F., Pansera, M., & Camatti, E. (2020). DNA metabarcoding and morphological analysis - Assessment of zooplankton biodiversity in transitional waters. *Marine Environmental Research*, *160*(February). https://doi.org/10.1016/j.marenvres.2020.104946

Siegert, M., Atkinson, A., Banwell, A., Brandon, M., Convey, P., Davies, B., Downie, R., Edwards, T., Hubbard, B., Marshall, G., Rogelj, J., Rumble, J., Stroeve, J., & Vaughan, D. (2019). The Antarctic Peninsula under a 1.5°C global warming scenario. *Frontiers in Environmental Science*, *7*(JUN), 102. https://doi.org/10.3389/FENVS.2019.00102/BIBTEX

Son, Y. B., Ishizaka, J., Jeong, J., Kim, H., & Lee, T. (2011). *Cochlodinium polykrikoides Red Tide Detection in the South Sea of Korea using Spectral Classification of MODIS Data*. *46*, 239–263.

Soto, D. F., Fuentes, R., Huovinen, P., & Gómez, I. (2020). Microbial composition and photosynthesis in Antarctic snow algae communities: Integrating metabarcoding and pulse amplitude modulation fluorometry. *Algal Research*, *45*(March 2019), 101738. https://doi.org/10.1016/j.algal.2019.101738

Steyaert, M., Priestley, V., Osborne, O., Herraiz, A., Arnold, R., & Savolainen, V. (2020). Advances in metabarcoding techniques bring us closer to reliable monitoring of the marine benthos. *Journal of Applied Ecology*, *57*(11), 2234–2245. https://doi.org/10.1111/1365-2664.13729

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

Turon, X., Antich, A., Palacín, C., Præbel, K., & Wangensteen, O. S. (2020). From

metabarcoding to metaphylogeography: separating the wheat from the chaff. *Ecological Applications*, *30*(2). https://doi.org/10.1002/eap.2036

Valsecchi, E., Bylemans, J., Goodman, S. J., Lombardi, R., Carr, I., Castellano, L., Galimberti, A., & Galli, P. (2020). Novel universal primers for metabarcoding environmental DNA surveys of marine mammals and other marine vertebrates. *Environmental DNA*, *2*(4), 460–476. https://doi.org/10.1002/edn3.72

Vaughan, D., Marshall, G., Connolley, W., Parkinson, C., Mulvaney, R., Hodgson, D. (2003). Recent rapid regional climate warming on the Antarctic Peninsula. In *Climate Change* (pp. 243–273).

Von Ammon, U., Wood, S. A., Laroche, O., Zaiko, A., Lavery, S. D., Inglis, G. J., & Pochon, X. (2019). Linking Environmental DNA and RNA for Improved Detection of the Marine Invasive Fanworm Sabella spallanzanii. *Frontiers in Marine Science*, *6*. https://doi.org/10.3389/fmars.2019.00621

Wang, Q., Garrity, G. M., Tiedje, J. M., Cole, J. R., & Al, W. E. T. (2007). *Naı ¨ve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy □ †*. *73*(16), 5261–5267. https://doi.org/10.1128/AEM.00062-07

Wangensteen, O. S., Palacín, C., Guardiola, M., & Turon, X. (2018). DNA metabarcoding of littoral hardbottom communities: High diversity and database gaps revealed by two molecular markers. *PeerJ*, *2018*(5). https://doi.org/10.7717/peerj.4705

Wood, S. A., Pochon, X., Laroche, O., von Ammon, U., Adamson, J., & Zaiko, A. (2019). A comparison of droplet digital polymerase chain reaction (PCR), quantitative PCR and metabarcoding for species-specific detection in environmental DNA. *Molecular Ecology Resources*, *19*(6), 1407–1419. https://doi.org/10.1111/1755-0998.13055

Yoccoz, N. G. (2012). The future of environmental DNA in ecology. *Molecular Ecology*, *21*(8). https://doi.org/10.1111/j.1365-294X.2012.05505.x

Zafeiropoulos, H., Gargan, L., Hintikka, S., Pavloudi, C., & Carlsson, J. (2021). *Bacteria are everywhere, even your CO1 marker gene data!*

Zoccarato, L., Pallavicini, A., Cerino, F., Fonda Umani, S., & Celussi, M. (2016). Water mass dynamics shape Ross Sea protist communities in mesopelagic and bathypelagic layers. *Progress in Oceanography*, *149*, 16–26. https://doi.org/10.1016/J.POCEAN.2016.10.003