UiT The Arctic University of Norway

Faculty of Biosciences, Fisheries and Economics - Norwegian College of Fishery Science

# Diving into the diet of sperm whales (*Physeter macrocephalus*) off the northern Norwegian coast: Insights from fecal DNA metabarcoding

Rasmus Buhl Søiland

Master's thesis in Biology FSK-3860 May 2024

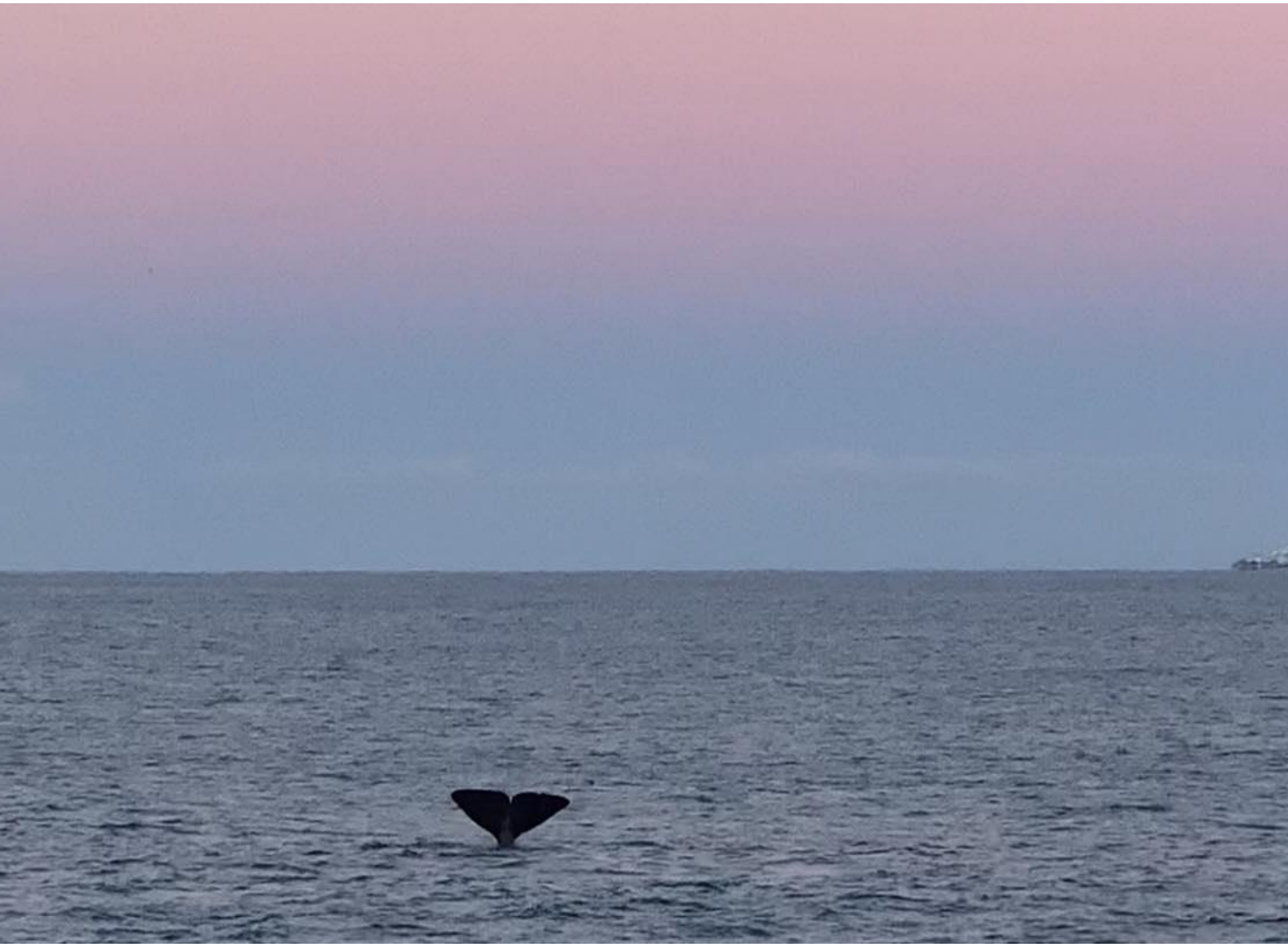


*Photo: Cecilie Bauer-Schäfer*

# Abstract

Sperm whales are elusive creatures; they spend much of their time out of sight, deep in the oceans. This makes them difficult to study, and previously, studies of their diet have mainly been stomach content analyses of stranded or caught individuals. However, DNA metabarcoding has the potential for solving this problem. This thesis examines the diet of sperm whales in the Norwegian Sea, of which we know very little, and aims to investigate the diet by examining fecal samples utilizing DNA metabarcoding and using surface water samples as a reference. Two types of fecal samples were utilized: ethanol fecal samples, which are ethanol that was filtered from sample-containing jars, as well as regular fecal samples, which are extracted DNA from the fecal matter. For comparison, seven reference surface water samples were selected. For this study, the fecal samples were split into five different sample groups with the closest related water samples. The samples were amplified and sequenced with two primer sets: Leray XT for amplifying and sequencing a fragment of the mitochondrial Cytochrome c Oxidase subunit I (COI) gene, and MiFish-U for amplifying and sequencing a region of the 12S ribosomal RNA gene. The focus point in this study was metazoans, and more specifically, fish species. The 12S marker registered 20 unique fish taxa for fecal samples, of which four were exclusive to fecal samples. The genetic marker COI identified nine unique fish taxa for fecal samples, of which five were exclusive to fecal samples. Surface water samples generally exhibited more unique taxa than fecal samples. Analysis of similarities (ANOSIM) proved significant dissimilarities between the sample types for the overall dataset (including all taxonomic kingdoms), group one in the 12S dataset, group four and the combined group in the COI-*Metazoa* dataset, and the combined group for the COI-fish dataset. Similarity Percentage (SIMPER) analysis was carried out on the significant groups and illustrated which species were responsible for most of the dissimilarity between the groups. Based on all of the results, it seemed that benthic fish were a smaller part of the diet of sperm whales than previously believed, and that *Gadus morhua* (Atlantic cod), *Sprattus sprattus* (sprat), *Clupea harengus* (Atlantic herring), *Trisopterus esmarkii* (Norway pout), *and Scomber scombrus* (Atlantic mackerel) were likely to have been a part of the diet of the sperm whales in the Norwegian Sea in the area of Bleik Canyon, either as primary or secondary prey. Notably, *T. esmarkii* and *S. sprattus* had not been registered in the diet of sperm whales before. However, a definitive conclusion on the prey of sperm whales in the Bleik canyon area in the Norwegian Sea cannot be reached, but this study provided insight into the fecal sampling method and the potential diet of sperm whales.

# Table of Contents

# List of Tables



# List of Figures

# Acknowledgements

I would like to show my sincere gratitude to everyone whose invaluable support and contributions helped me make my project a reality. Firstly, I express my appreciation to the University of Copenhagen for facilitating my (KU-ID: CTS410) exchange with the University of Tromsø. I am deeply thankful for my third supervisor at the University of Copenhagen, Morten Tange Olsen, for putting me into contact with the right people in Tromsø. Special thanks to my first supervisor at the University of Tromsø, Kim Præbel, for the guidance, numerous meetings, and for entrusting me with the sperm whale diet project. I am also grateful to Audun Rikardsen for his assistance as second supervisor. I want to pay my respects to the entire Research Group for Genetics for their valuable input in group meetings and for welcoming me with open arms into the group. I want to acknowledge Adwoa Sarfowaa for her guidance on the use of Powersoil Pro kits, Melissa Michelle Brandner for supervising me on PCR, pooling, and library preparation, and Andrea Iselin Elvheim, who was always ready to help whenever I had problems or questions in the laboratory. Thanks to Tiu Similä from Whale2Sea for always being willing to help with questions about the samples they had taken, and to Zoë Luccia Juliette Morange, and Emma Vogel for helping me with questions about the fecal and surface water samples, which they had taken part in sampling. Thank you to Ove Mikal Pedersen for jumping into the sea of whale poop, to collect the fecal samples. Furthermore, thanks to Daniel Kumazawa Morais and Mads Kristian Reinholdt Jensen for helping me with metabarcoding theory, and a special thanks to Daniel Kumazawa Morais for assisting me with handling data and creating plots in RStudio. Lastly, a big thanks to all the people who have helped me with feedback on my thesis: Kim Præbel, Audun Rikardsen, Morten Tange Olsen, Daniel Kumazawa Morais, and Melissa Michelle Brandner.

# 1 Introduction

Global temperatures are increasing, and studies indicate that we are on the way to surpassing 1.5°C above preindustrial temperatures (Matthews & Wynes, 2022). It was established in the Paris Agreement in 2016 that we should stay well below 2,0°C above preindustrial temperatures and preferably stay below 1,5°C. Nonetheless, the commitment to stay below these levels faces a critical challenge since current projections suggest that the limit will be surpassed within the next 10 years (Matthews & Wynes, 2022). However, the temperature increase is not uniform across the globe. Notably, the temperature increased at a rate three–four times faster in the Arctic compared to the rest of the world in the period 1979–2021 (Rantanen et al., 2022). In marine ecosystems, these temperature changes have wide-ranging impacts (Gissi et al., 2021) that are yet to be fully understood. But impacts on ecosystem composition and food chains across the globe have been observed (Gibert, 2019). Understanding the diet of top predators becomes crucial in this context. Disruptions in a food chain, such as a species migrating northward because of climate change, could lead to a scenario where the top predator experiences a scarcity of its usual prey. Consequently, the predator may adapt by changing its prey preferences, thereby influencing other food chains. Alternatively, the predator disappears entirely from an area, following the prey distribution. The reduction of ice cover in the Arctic due to climate change has resulted in the largest toothed whale, the sperm whale (*Physeter macrocephalus*, Linnaeus, 1758), now having an expanded habitat range towards the poles (van Weelden et al., 2021). This shift might impact the Arctic food web and increase competition for other top predators such as beluga whales (*Delphinapterus leucas*) and narwhals (*Monodon monoceros)*, which are endemic to the Arctic region. Thus, a comprehensive investigation of the dietary preferences of sperm whales is important to understand their potential impact on the Arctic food web (Posdaljian et al., 2022). Sperm whales play a crucial ecological role by recycling nutrients. They forage in the deep seas and bring those nutrients up to the surface through their excretion. This process can enhance phytoplankton production, which takes place in the photic zone. So not only are the sperm whales at the top of the food chain, but they also provide nutrition for the lower tiers of it (Pearson et al., 2023).

## 1.1 General biology of sperm whales

Sperm whales are mysterious creatures that are hard to study because they spend most of their lives deep in the oceans. With a total length of up to 21 meters, they are the largest toothed predator in the world (Maio et al., 2022). There is a great difference in the size of males and females since the sperm whale is a sexually dimorphic species, and males can be up to 40% longer and three times heavier than females (Eguiguren, 2023). Male individuals can dive to depths of up to 1900 meters during foraging activities and can remain submerged for up to 60 minutes (Towers et al., 2019), but they may have the potential to dive much deeper and remain submerged for even longer (Towers et al., 2019). Sperm whales have a global distribution from the ice-free waters at the poles all the way to the equator (Figure 1), and they generally inhabit productive waters, such as those along continental shelves, where it is easier for them to find prey (Whitehead, 2018).

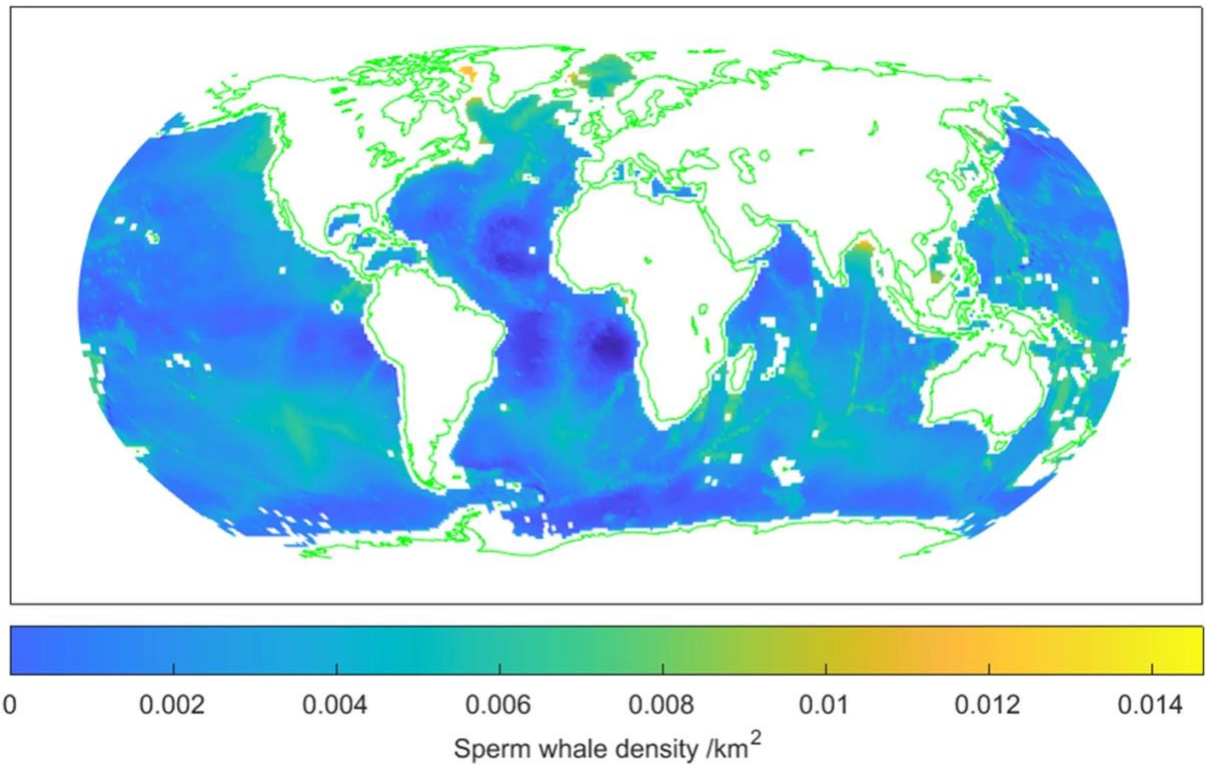



*Figure 1 - World map showing the habitat range of sperm whales. Estimated population density was derived from a weighted stepwise regression analysis of survey densities in relation to oceanographic variables. Land areas and ocean areas less than 1000 meters deep are colored white (Whitehead & Shin, 2022).*

Sperm whales are social animals and stay in groups called natal units in mid- or low latitudes on both sides of the hemisphere (Kobayashi et al., 2020). Females usually stay in the natal units throughout their lives, whereas males depart from the group to travel to higher latitudes outside of the females' range before reaching sexual maturity between 6 and 16 years of age

(Kobayashi et al., 2020). The males form groups with other males of roughly the same age, called bachelor schools. When the mating season approaches, the males that have reached sexual maturity migrate to lower latitudes to meet the females (Kobayashi et al., 2020). The social behavior decreases with age in the male sperm whales, and older individuals have been observed in solitude at high latitudes (Kobayashi et al., 2020).

## 1.2 Whaling

Historically, there has been hunting for whales (called whaling). Whaling on sperm whales started in the early 17th century (Whitehead & Shin, 2022), mainly for the very valuable sperm oil and ambergris. Sperm oil, derived primarily from the spermaceti organ and the melon located in the head along with the blubber, served multiple purposes for humans, such as oil for lamp illumination, lubricant for various machinery, and as a rust preventative (York, 2017). Another valuable component obtained from sperm whales was ambergris, a substance formed in their digestive system. Ambergris was widely used in perfumes and as incense (Brito et al., 2016).

It is hard to estimate how large the population size of sperm whales was prior to the onset of whaling, yet recent assessments suggest that the global population of sperm whales counted approximately two million individuals (Whitehead & Shin 2022). The population size was more than halved due to the whaling industry, and it is estimated that around 736,000 individuals were left in 1993 (Whitehead & Shin 2022). Today, the sperm whale has been classified as vulnerable according to the International Union for Conservation of Nature's (IUCN) Red List of Threatened Species (IUCN, 2023). In the 1970s, commercial whaling was banned which led to a recovering sperm whale population, with estimates indicating an increase to approximately 845,000 individuals globally in 2022 (Whitehead & Shin, 2022).

## 1.3 Previous studies on sperm whale diet

The dietary preferences of sperm whales remain a subject of ongoing investigation. Historically, it has been difficult to examine their diet since they spend much of their lives deep in the oceans. Therefore, examination of stomach content from deceased whales has been the primary method of diet analysis, but that has been limited to stranded whales after the ban on whaling on sperm whales (Wild et al., 2020). In most regions of the world, their diet has been found to mainly consist of cephalopods, particularly squid (Wild et al., 2020). However, an older study examining the stomach content of sperm whales caught in Icelandic

waters (Roe, 1969), revealed that fish may be a main component of the diet in some areas of the world. The most common fish species found in the sperm whale stomachs in Icelandic waters were *Cyclopterus lumpus* (lumpsucker), *Sebastes sp*. (redfish), and *Lophius piscatorius* (anglerfish) (Roe, 1969). All the fish found in the sperm whale stomachs were benthic, with the exception of *Pollachius virens* (saithe), which was only found in 2 out of 57 stomachs (Roe, 1969). The study found that 45% of the sperm whale stomachs containing food exclusively consisted of fish, and only 2% consisted solely of squid. In total, there were fish in 98% of the sperm whale stomachs, which was a much higher percentage than for any other area the data was compared with. The comparison areas for the study were the Azores, Kuriles, Kamchatka, Japan, Antarctic (pelagic), and South Georgia and South Africa, all at lower northward latitudes than Iceland. The location with the second highest percentage of stomachs with fish was the Kuril Islands, with a percentage of stomachs with fish of 30%. Antarctica had one of the lowest percentages of stomachs containing fish, at only 13% and 5% in two different studies (Roe, 1969). A more recent study on diet employed isotope analysis on sperm whales in the Gulf of Alaska (Wild et al., 2020) Here it was found that the combined *Anoplopoma fimbria*/*Squalus suckleyi* (sablefish/dogfish) group made up the largest proportion of the sperm whale diet in the Gulf of Alaska at 35%, followed by *Rajidae sp.* (skates) at 28% and *Sebastes sp.* at 12%.

Not much is known about the diet of sperm whales in the Norwegian Sea, where this study has its focus, although some information exists from strandings and caught whales. In 1971, a total of 12 sperm whales were caught off of Andøya in Northern Norway, and their stomachs were investigated (Bjørke, 2001). The study found that four of the sperm whale stomachs solely contained squid beaks, six contained a combination of squid beaks and fish, and two of the stomachs were empty. However, there was no information on the species of fish and cephalopods. In another study, three male sperm whales were stranded on the island of Andøya in Northern Norway between March 28 and April 21, 2020 (Similä et al., 2022). It was found that the stomachs contained four different fish species: *C. lumpus*, *L. piscatorius*, *Gadus morhua* (Atlantic cod), and *Chondrichthyes sp*. (cartilaginous fish). Furthermore, two different squid species were found: *Gonatus sp.* (armhook squid), and *Histioteuthis sp*. (cock-eyed squid). Notably, *L. piscatorius*, *G. morhua*, *Chondrichthyes sp*., and *Histioteuthis sp.* had not been seen before in the diet of sperm whales in Norwegian waters, adding four new species to their known diet. Furthermore, there were unidentifiable fish remains in all three of the whale stomachs. Although the sample size was very limited, the findings indicate that fish

might be an important part of the diet of sperm whales in the Norwegian Sea, but to what extent remains to be investigated.

Commercial fishing has an influence on the diet of sperm whales in the area of Bleik Canyon in the Norwegian Sea during some periods of the year (Oyarbide et al., 2023), since longline fishing attracts sperm whales, which will engage in depredation, which for toothed whales is the act of removing fish from fishing gear. It is influencing the diet of sperm whales since the whales probably find it easier to eat fish that cannot escape, and furthermore, the longlines also bring the fish closer to the surface, whereas they can avoid their deep dives. Sperm whales have also been detected following fishing vessels for multiple days (Oyarbide et al., 2023). It can therefore be assumed that a sperm whale that has been close to a longline fishing vessel have been depredating and might have a great proportion of the species targeted by the fishing vessel in its stomach. More research is needed to understand the dietary habits of sperm whales completely, and other methods, such as DNA metabarcoding, might be helpful tools to get a clearer picture of what sperm whales eat.

## 1.4 Introduction to environmental DNA

Analysis of environmental DNA is a fairly new technique but has found widespread application in various research domains and is being used for an increasing number of purposes. The fundamentals behind environmental DNA are that organisms continuously leave behind traces of DNA, which can come in many forms, such as urine, feces, scales, hairs, skin, etc. This makes it possible to gather DNA from multiple organisms in one sample. There are many ways to obtain environmental samples. There are, for example, natural samplers such as leeches (Schnell et al., 2018) and mussels (Weber et al., 2023), as well as manually sampling by vacuuming air through a filter (Clare et al., 2021) or filtering a water sample (Thomsen & Willerslev, 2015). Before the implementation of eDNA analysis, it was required to physically observe species or their remains to verify that they were or had been in an area. While this approach was not a problem for abundant species which could easily be observed, it posed challenges for rare species that might evade detection, particularly in the presence of human observers (Rishan et al., 2023). The implementation of eDNA has significantly alleviated these challenges. One notable application of eDNA analysis lies in its efficiency for determining communities within aquatic ecosystems, especially in the vast deep-sea environment, which makes up the largest ecosystem on Earth. Previously, deep-sea exploration required several fishing expeditions to discover which species were present at

these depths. However, with the advancement of eDNA, water samples taken at various depths with a Conductivity-Temperature-Depth (CTD) profiler can be analyzed with eDNA to detect species in the given area (Yoshida et al., 2023). This represents a great progress in our ability to study the deep seas without the previous logistical challenges.

There are several applications for eDNA, but it also has limitations, which must be considered when doing eDNA research. For instance, the absence of a species in a water sample does not automatically imply its absence in the area (Ficetola et al., 2015). On the other hand, presence of a species in a sample does not automatically indicate that it has been in the area, as the DNA could have been transported to the sampling location by currents. The presence of a species may also be due to contamination either in the field or during laboratory procedures, which is why it is important to work as sterile as possible when handling eDNA samples (Roussel et al., 2015). These limitations highlight the need for caution in the interpretation of eDNA results, calling for a nuanced and critical approach to avoid inaccurate conclusions.

## 1.5 Diet analysis methods

When analyzing diet, two methods of DNA analysis have been important: species-specific quantitative PCR (qPCR) and metabarcoding. Species-specific qPCR provides a targeted approach, offering insights into specific species in the diet, and has been used on various sample types, including fecal matter (Murray et al., 2011), stomach content (Töbe et al., 2010), and oral swabs (Young et al., 2023). On the other hand, metabarcoding is a broad-scale analysis method that can be used with different genetic markers, depending on the organisms of interest. There are universal genetic markers, such as different fragments of the mitochondrial Cytochrome c Oxidase subunit I (hereafter referred to as COI) gene, e.g. Leray XT (Geller 2013; Leray et al. 2013; Wangensteen et al. 2018), and more specific genetic markers, such as ceph18S for targeting Cephalopoda (de Jonge et al., 2021), and 12S ribosomal RNA (hereafter referred to as 12S) for targeting fish species (Cawthorn et al., 2012). The metabarcoding method has been used for eDNA samples of several different types such as esophagus samples (Díaz-abad et al., 2022), fecal samples (Cabodevilla et al., 2023), and stomach content (Siegenthaler et al., 2019; Urban et al. 2022). Fecal samples from animals such as sperm whales, which have liquid feces (Roman & McCarthy, 2010), can be considered eDNA since the fecal samples will consist of a mixture of fecal matter and seawater.

Historically, studying an animal's diet necessitated finding prey in its stomach (Similä et al., 2022), in the fecal matter (Trites & Joy, 2005), or observing feeding behavior (Matthews et al., 2020). Additionally, methods such as tagging, stable isotope analysis (Tykot, 2004), and fatty acid analysis (Cordain et al., 2002) have been used in diet analysis. For a species like the sperm whale, known to forage in the deep sea, direct observation of feeding is a significant challenge. Furthermore, the digestion process makes it hard to identify species in their fecal matter (Trites & Joy, 2005). Compounded by the protected status of sperm whales, investigations into stomach contents are limited to stranded individuals, introducing potential biases as dying individuals may exhibit different foraging behaviors than their healthy counterparts (Pierce et al., 2004). Another potential bias in stomach content analysis of sperm whales is the potential retention of cephalopod beaks in their stomachs, which might lead to an overrepresentation of cephalopods in their diet (Foskolos et al., 2020). Employing eDNA analysis of the fecal matter from sperm whales appears to be the best approach. This method has not been previously applied to sperm whales. By circumventing the limitations associated with previous methods, the eDNA approach may provide new insights into the dietary habits of sperm whales. The use of eDNA analysis of fecal samples provides an opportunity to improve our understanding of sperm whales, shedding light on aspects that were previously difficult to examine. However, sampling of fecal matter is no easy task. They are hard to track down, and when they defecate the feces must be sampled fast, since it is not firm and therefore quickly dilutes.

## 1.6 Objectives

DNA metabarcoding of fecal samples had never been used to study the diet of sperm whales before and have the potential to add new knowledge about the species composition in the diet of sperm whales. Therefore, the main objective of this study was to obtain insights into the diet of sperm whales in the Bleik Canyon area of the Norwegian Sea. This study also aimed to compare whether ethanol fecal samples (hereafter referred to as ethanol samples) were as effective for representing the diversity in the diet as regular fecal samples.

The technique used for this study was DNA metabarcoding of fecal samples collected in the seawater behind diving sperm whales. The data was explored using DNA obtained directly from the fecal matter as well as from the ethanol that the fecal samples were stored in, and surface water samples (hereafter referred to as water samples) that served as reference samples. Two sets of primers were used: the Leray-XT primer set (Wangensteen et al., 2018),

which amplified the metabarcoding fragment of the COI gene, and the MiFish-U primer set (Miya et al., 2015), which amplified the metabarcoding fragment of the 12S gene.

I hypothesized that:

i. Fish species consumed by sperm whales in the Bleik Canyon area of the Norwegian Sea are primarily benthic species.

This is based on the fact that sperm whales mostly make deep dives when foraging (Watwood et al., 2006), and that previous studies on the sperm whale diet in Icelandic waters and the Norwegian Sea indicated a diet consisting of mainly benthic prey (Roe, 1969; Similä et al., 2022).

ii. Fecal samples will yield higher amounts of unique taxa than water samples.

This is hypothesized since the fecal matter is believed to bring in new species from the deep ocean to the already present species at the surface.

iii. Ethanol and fecal samples will display the same diversity of unique taxa identified by the COI and 12S primers, respectively.

This is hypothesized due to previous studies, which have shown that ethanol that had been used for preservation was a suitable source of DNA (Wang, Chen et al., 2021).

# 2 Materials and methods

## 2.1 Sampling area and design

This study utilized analysis of eDNA from 16 sperm whale fecal samples as well as from seven related water samples that acted as a control for the fecal samples. Two types of fecal samples were used: ethanol and fecal samples. Figure 2 shows the various steps in the process, from sampling to results.

All samples were collected by UiT colleagues and employees from the whale safari company Whale2Sea in Andenes, northern Norway. The samples were collected in the area of the Bleik Canyon in the Norwegian Sea (Figure 3), which is a known foraging ground for sperm whales in the area (Similä et al., 2022). Bleik Canyon starts about 15 kilometers off the shore of Andenes and has very steep sides. The maximum depth in the 40–50 km long canyon is 3

kilometers in the outer end, and it is 20 kilometers across at its widest (Rødland & Bjørge, 2015). The samples included in this project were taken between June 5, 2021, and August 17, 2023 (Table 1).

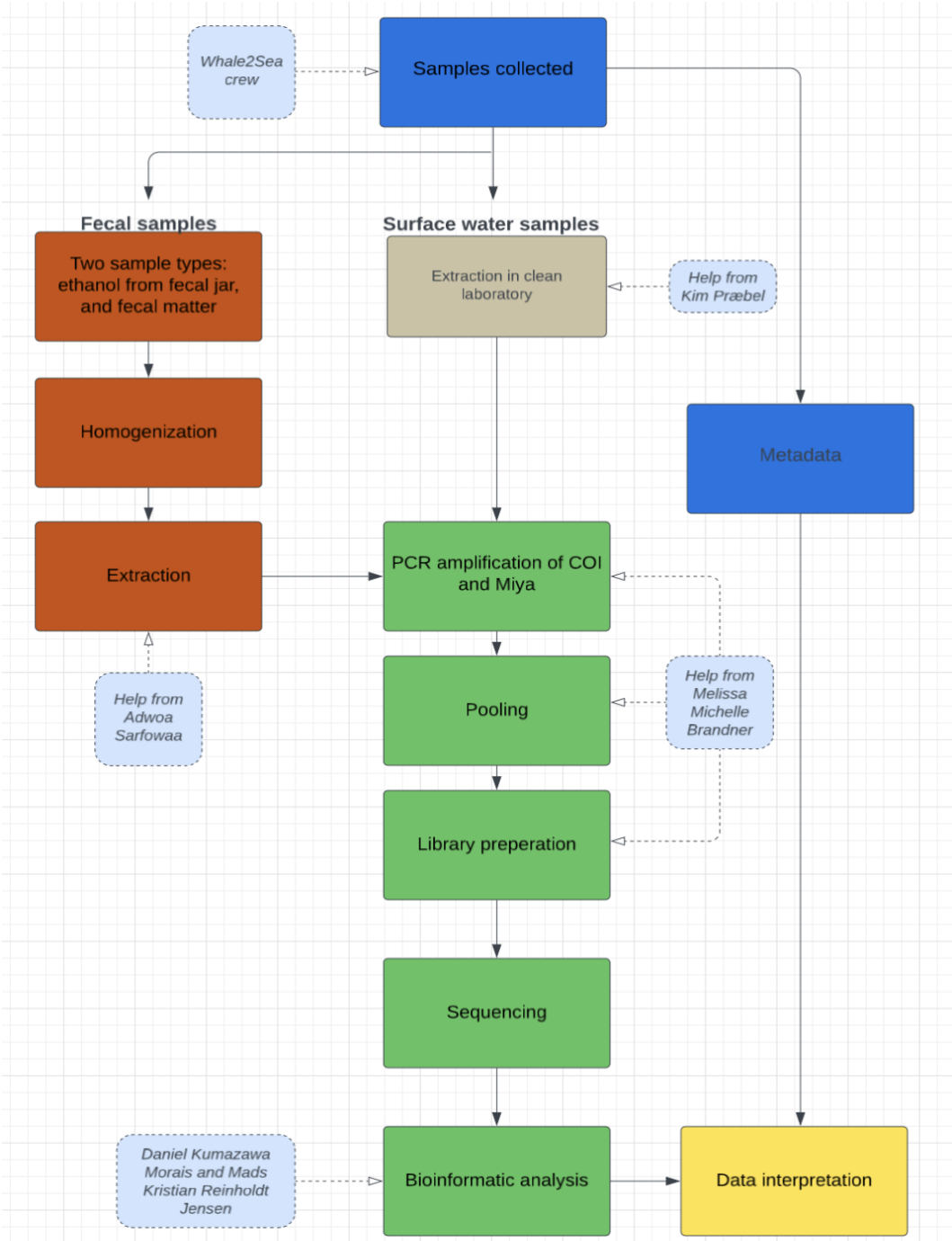



*Figure 2 - Flow chart showing all the methods included in this project. Light blue boxes show which people that has been involved with the various steps.*

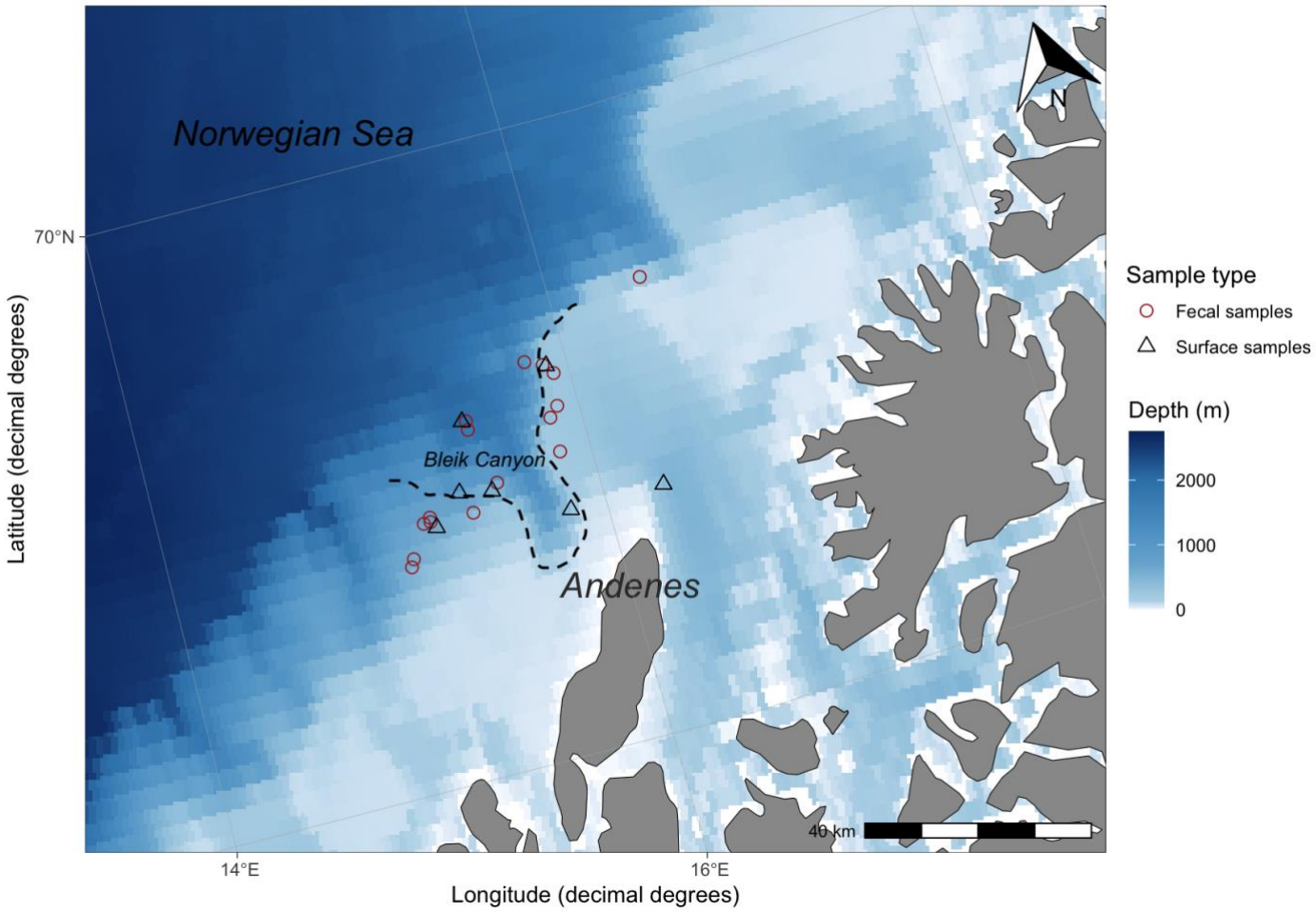



*Figure 3 - Sample locations of seven water samples (black triangles), and 16 fecal samples (brown circles). Bleik Canyon is marked with a dashed line. The map includes bathymetric data, with darker colors illustrating a greater depth than lighter colors.*

*Table 1 - Sample dates for surface water samples and fecal samples, including information on number of samples for each date.*

| Sampling date | Water samples | Fecal samples |
|---|---|---|
| 05.06.2021 | 0 | 3 |
| 09.06.2021 | 1 | 0 |
| 22.04.2022 | 1 | 0 |
| 30.05.2022 | 0 | 1 |
| 27.03.2023 | 0 | 1 |
| 28.03.2023 | 1 | 1 |
| 29.03.2023 | 1 | 4 |
| 10.06.2023 | 2 | 2 |
| 13.08.2023 | 1 | 1 |
| 14.08.2023 | 0 | 1 |
| 17.08.2023 | 0 | 2 |
| **Total** | **7** | **16** |

## 2.2 Fecal sampling

In total, 16 fecal samples were collected from sperm whales in the Norwegian Sea from 2021 to 2023 (Table 1). Three of the samples were from 2021, all taken on June 5. Only one sample was taken in 2022, on May 30. The majority of the samples were from 2023, where 12 samples were taken from March 27 to August 17. The sperm whales were located by using a rigid inflatable boat (RIB). Various research was carried out on the sperm whales, such as biopsies and tagging. Sometimes the sperm whales defecated as they were diving, or as a reaction to tagging or biopsy sampling. In the case of defecation, one person jumped into the water with a plankton net and swirled it through the fecal matter, making sure it did not leave the net again (Appendix A). On the boat, the fecal matter was flushed into a jar with 96% ethanol, and additional 96% ethanol was added to the jar for immediate preservation of the DNA. The coordinates for the sampling locations were noted when the samples were taken. However, there was no note of the coordinates for the third fecal sample on 5/6/2021, therefore, the most likely location has been used instead (Suppl. Table 1). Afterwards, the samples were sent to the Norwegian College of Fishery Science at the University of Tromsø (UiT) for storage in a -20°C freezer, where they were stored until further processing.

Fecal samples were paired up with the surface water samples that they had the closest timewise relation to (Table 3). Water samples were used as a reference to the fecal samples to be able to see which species were already present at the surface. Five sample groups were made: sample group one (SG1) with reference water samples from the exact diving site where the fecal sample was taken; sample group two (SG2) with water samples taken the same day as the fecal samples; sample group three (SG3) with a one-day difference between the fecal and the water sample; sample group four (SG4) with a four-day difference; and sample group five (SG5) with more than a month's difference. Furthermore, a combined sample group (SGC) that contained sample groups SG1–SG5 was made as well. The sample groups were made to have as reliable data as possible. SG1 had surface water samples that were taken right after the fecal sample and in the same area and it was therefore more reliable than the rest of the groups. The reliability of the samples decreases as the time between fecal samples and water samples increases. SG5 only consisted of one fecal sample and one surface water sample, and the dataset was therefore too small to be used for ANOSIM and SIMPER analysis. Thus, SG5 was only used for analysis in SGC. These sample groups will be referred to throughout the thesis.

*Table 2 - Grouping of fecal samples to the closest related surface water samples. There were 7 water samples, 16 fecal samples, and 16 ethanol samples. SG1 had water samples that had been sampled close to the diving site of the whale that had been fecal sampled. For SG2, water and fecal samples were taken on the same day, SG3 had one day difference between the sample types, SG4 had four days difference, and SG5 had 39 days difference. SGC is a group comprised of SG1–SG5.*

| **Sample groups** | | | | |
|---|---|---|---|---|
| Fecal samples | Ethanol sample | Surface water sample | Group number | Group ID |
| Pma23004F | Pma23004E | MOBY23003A-C | 1 | SG1 |
| Pma23007F | Pma23007E | MOBY23007A-C | 1 | SG1 |
| Pma23008F | Pma23008E | MOBY23008A-C | 1 | SG1 |
| Pma23009F | Pma23009E | MOBY23013A-B | 1 | SG1 |
| Pma23002F | Pma23002E | MOBY23002A-C | 2 | SG2 |
| Pma23003F | Pma23003E | MOBY23003A-C | 2 | SG2 |
| Pma23005F | Pma23005E | MOBY23003A-C | 2 | SG2 |
| Pma23006F | Pma23006E | MOBY23003A-C | 2 | SG2 |
| Pma23001F | Pma23001E | MOBY23002A-C | 3 | SG3 |
| Pma23010F | Pma23010E | MOBY23013A-B | 3 | SG3 |
| Pma21001F | Pma21001E | MOBY21008A-C | 4 | SG4 |
| Pma21002F | Pma21002E | MOBY21008A-C | 4 | SG4 |
| Pma21003F | Pma21003E | MOBY21008A-C | 4 | SG4 |
| Pma23011F | Pma23011E | MOBY23013A-B | 4 | SG4 |
| Pma23012F | Pma23012E | MOBY23013A-B | 4 | SG4 |
| Pma22001F | Pma22001E | MOBY22007A-C | 5 | SG5 |
| All | All | All | 1-5 | SGC |

## 2.3 Surface water sampling

Water samples were taken and sampled by the colleagues from UiT and the crew at Whale2Sea following a surface water sample protocol (Appendix B) with some modifications. A clean 10 L bucket was used as water sampler in the surface water at 0–0.5 meters depth. The water was then moved directly to sterile plastic bags, which were sealed, and they were stored until arrival at land. The samples were immediately sampled when at land. The water samples were sampled at seven different locations in total from 2021–2023 (Table 1). Coordinates for the sampling locations were noted when the samples were taken; however, exact coordinates were missing for the sample taken on August 13, 2023, and approximate coordinates have been used instead (Suppl. Table 2). For most sample sites, three replicates of the water samples were made; however, one site had only two replicates. Furthermore, air blanks were taken at six sites in 2021. The water samples were filtered through three 0.22 µm Sterivex filters using 50 ml syringes (Appendix B). Afterwards, the

filters were placed in a marked falcon tube and then put into a ziplock bag, which was stored in a -20°C freezer until shipment to UiT. Upon arrival at UiT, the samples were stored in a dedicated eDNA freezer at -80°C until DNA extraction.

## 2.4 Extraction of fecal matter

A protocol was created to prepare the fecal samples for extraction (Appendix C), as there was no one available. Two methods were used for the extraction of DNA from the fecal samples. One method involved the extraction of DNA from the ethanol in which the fecal matter was stored. The other method involved the extraction of DNA from the fecal matter itself. The sample jars containing ethanol and fecal matter were placed in an undisturbed freezer for 24 hours, allowing the fecal matter to precipitate. Then, ethanol from the sample jars was extracted with a syringe and filtered through a 0.22 µm Sterivex filter for each sample. To check for contamination, two ethanol blanks were made on two different dates by filtering clean ethanol through a Sterivex filter at the same place where the filtering of the fecal samples took place. Thereafter, the jars containing the fecal matter were taken to laboratory B-310 (NFH-building at UiT) for homogenization and extraction. The fecal matter was homogenized with a mini chopper (POINT POCH7037). Afterwards, the homogenized matter was poured into falcon tubes and centrifuged to compress the fecal matter at the bottom. The remaining ethanol was gently poured out, making sure no fecal matter left the falcon tube. Aluminum foil was wrapped around the top of the tubes, and small holes were poked in the foil to allow the remaining ethanol to evaporate overnight. Some of the samples were not completely dried the next day and were placed in an oven at 56°C for 30 minutes to accelerate the evaporation process. The two largest samples required four hours in the oven for complete evaporation of the ethanol. The Sterivex filters used for filtering the ethanol were cut open with a pipe cutter, and the filters were cut into small pieces with a clean scalpel. Thereafter, DNA from the Sterivex filters as well as from the dry fecal matter was extracted with a DNeasy Qiagen Powersoil Pro kit, following the protocol (Appendix D). Since more than 12 samples were handled simultaneously, the vortexing time was increased by 10 minutes. After following the PowerSoil Pro Kit protocol, the DNA concentrations were checked with a Qubit fluorometer. Then 15 µl of the DNA from each sample were transferred to a PCR plate. The plate was double-bagged and stored in a -20°C freezer until it could be taken through PCR.

## 2.5 Surface water - extraction

The water samples were transferred to an over-pressured eDNA clean lab at UiT. All extractions were carried out in the clean lab wearing clean suits to avoid as much contamination as possible. Cleaning was carried out following the clean lab routines protocol before starting the extractions (Appendix E). The samples were slowly thawed in a fridge overnight. The following day, extraction could begin. The extraction was a two-day process, and for each extraction 23 samples could be extracted. An extraction blank was made for each round of extraction to detect possible contamination. The Qiagen DNEasy Blood and Tissue Kit protocol was followed with slight modifications to make it better suited for water samples (Appendix F). The modifications for use on eDNA samples were a buffer composition of 50 µl ProK and 450 µl ATL per sample (step 5). Furthermore, the incubation time was set to 20 hours for all the rounds of samples (step 10), and all samples were eluted in 75 µl AE buffer (step 36).

## 2.6 PCR amplification, pooling, and library preparation

After extraction, the samples were ready for polymerase chain reaction (PCR), which was carried out with the assistance of Melissa Michelle Brandner. The aliquots were pipetted into a PCR plate for one-step PCR amplification. Two different primer sets were used on all samples: Leray XT on the COI gene (Wangensteen et al. 2018) (Appendix E), and MiFish-U 12S on the 12S rRNA gene (Miya et al., 2015) (Appendix F). The PCR mix consisted of 10 µl AmpliTaq Gold Master Mix, 0.16 µl Bovine Serum Albumin (20 µg/µl), 5.84 µl $H_2O$ (COI) or 4.84 µl $H_2O$ (12S), 1 µl each of forward and reverse tagged primer (5 µM), and 2 µl of DNA template (COI) or 3 µl of DNA template (12S). The PCR mix was prepared in bulk by multiplying by the number of samples.

After PCR, 10 µl from each sample was pooled, and the 12S and COI were kept separated. Then the pools were purified with MinElute to clean up after the PCR, and removing DNA fragments below 70 base pairs, which further concentrated the amplified fragments. In the end, the DNA concentration of the pooled samples was checked with a Qubit fluorometer using a broad-range DNA quantification kit. The concentration was 266 ng/µl for COI and 374 ng/µl for 12S.

Library preparations were performed using a QIAseq® 1-Step Amplicon Library kit (Appendix G). After this, they were sent for 250PE (COI) or 150PE (12S) illumina sequencing on a NovaSeq sequencing platform at Novogene, which is a commercial sequencing company.

## 2.7 Bioinformatics

After sequencing, the raw data were processed by Daniel Kumazawa Morais and Mads Kristian Reinholdt Jensen from the Research Group for Genetics using the default MetaBarFlow pipeline (Sigsgaard et al., 2022) on resources provided by Sigma2, the National Infrastructure for High-Performance Computing and Data Storage in Norway. We used the Divisive Amplicon Denoising Algorithm 2 (DADA2) (Callahan et al., 2016) and implemented an R package to infer amplicon sequence variants (ASVs) from the raw sequences, remove low-quality reads, and remove chimera sequences that occurred due to errors in PCR amplification. Afterwards, the ASVs were taxonomically assigned using the Basic Local Alignment Search Tool (BLAST) algorithm (Altschul et al., 1990) on the Taxonomy Database in the National Center for Biotechnology Information (NCBI) nucleotide reference database (Sayers et al., 2021). After taxonomic assignment by BLAST, the whole dataset was curated manually, and final identification was determined for each ASV. Some ASVs could be identified at a lower taxonomic level based on the alternatives for each ASV. Then, similar final identifications were clustered together, creating a dataset with unique taxa, which was used as a unit for the majority of the plots in this thesis.

## 2.8 Plots and statistical analysis

Plots and statistical analyses were carried out in RStudio version 12.1-402 (Appendix H). The bathymetric map was made with the *ggOceanMaps* package (Vihtakari, 2024) in RStudio, as well as *ggplot2* and *ggrepel.* Coordinates from fecal samples and water samples were plotted onto the map. The Phyloseq package was used to handle the data files. Rarefaction curves were plotted using the *rarecurve* function from the Vegan package, to illustrate if the sequencing depth was sufficient for representation of unique taxa in each sample. To examine if the number of samples were sufficient for illustrating the diversity of unique taxa, species accumulation curves were plotted. The *specaccum* function from the *Vegan* package was used to plot species accumulation curves for each sample type at 100 permutations, and boxplots were plotted for each sample on the rarefaction curves. For illustrating dissimilarities between water, ethanol, and fecal samples, non-metric multidimensional scaling (nMDS) plots were created using the *metaMDS* function from the *Vegan* package, with Bray-Curtis as the chosen dissimilarity index. Ellipses were plotted on the nMDS plots with the *stat_ellipse* function from the *ggplot2* package. To examine if the dissimilarities observed from the nMDS plots were significant, analysis of similarities (ANOSIM) was performed using the *anosim* function from the *vegan* package and was carried out on all sample groups as well as the combined

groups. R-values range from -1 to 1, with high values indicating more dissimilarity, values close to zero indicating similarity, and values below zero indicating more dissimilarity within groups than between. P-values below 0.05 were deemed significant. Similarity percentage (SIMPER) analyses were made with the *simper* function from the Vegan package and were performed on groups that were deemed significant by the ANOSIM analyses. SIMPER was utilized to understand which unique taxa were most responsible for the observed dissimilarity between fecal and water samples.
Bubble plots were made with the *geom_point* function from the *ggplot2* package, and the bubbles illustrated relative abundance of unique taxa with intervals of 0.1. Stacked barplots were made with *ggplot2,* as a way to illustrate relative abundances of unique taxa for the different sample types. The plots were styled using *RColorBrewer* and *Viridis*.

# 3 Results

This study investigated the sperm whale diet by extracting DNA from 16 fecal samples using the fecal matter and the ethanol in which the fecal matter was stored for DNA preservation. Additionally, seven water samples were analyzed, each with two to three replicates, which were used as references for the fecal samples. Two types of genetic markers were used for sequencing: COI and 12S.

The COI metabarcoding dataset consisted of 32,430 ASVs with a total of 8,486,350 reads (Table 2). Taxonomical assignment was possible for 2,778 ASVs with a certainty of at least 98%. After filtering out obvious contaminants, such as Homo sapiens, land-living mammals, livestock, and land-living plants, as well as removing irrelevant surface water samples, 2,653 ASVs were left. Those ASVs contained a total of 1,629,376 reads. Of these, 319,218 reads originated from seven water samples, 544,070 reads from 16 ethanol samples, and 760,926 reads from 16 fecal samples. Furthermore, there were 11,533 reads from blanks in total (Suppl. Table 3), of which 164 were from two ethanol blanks, 7,288 from seven extraction blanks, 4,016 from six air blanks, and 66 from four PCR blanks. After curation of the dataset, zero reads from blanks were left.

The 12S dataset consisted of 260 ASVs with a total of 147,058 reads (Table 2). Taxonomical assignment was possible for 226 ASVs with a certainty of at least 98%. After filtering out contaminants and irrelevant samples, 185 ASV’s remained. Those ASVs contained a total of

44,624 reads, whereas 18,791 originated from seven water samples, 8,078 originated from 16 ethanol samples, and 15,595 originated from 16 fecal samples. Furthermore, there were 1,596 reads in total from all blanks (Suppl. Table 3). From these, 18 reads were from ethanol blanks, one read from extraction blanks, 1,098 from air blanks, and 479 reads from PCR blanks. After curation, zero reads were left from the blanks.

*Table 3 - Summary of ASVs, reads, and unique taxa for each genetic marker before and after data curation.*

| **Sequencing primer** | **COI** | | | **12S** | | |
|---|---|---|---|---|---|---|
| **Total ASVs** | 32,430 | | | 260 | | |
| **Total reads** | 8,486,350 | | | 147,058 | | |
| **Assigned ASVs** | 2,778 | | | 226 | | |
| **ASVs after curation** | 2,653 | | | 185 | | |
| **Reads after curation** | 1,629,376 | | | 44,624 | | |
| **Sample type** | **Water** | **Fecal** | **Ethanol** | **Water** | **Fecal** | **Ethanol** |
| **Unique taxa** | 91 | 149 | 106 | 24 | 19 | 12 |
| **Final curation reads per sample type** | 319,218 | 760,926 | 544,070 | 18,791 | 15,595 | 8,078 |

## 3.1 Reads and unique taxa counts

The rarefaction curves showed that the sample types —water, fecal, and ethanol samples — all reached an asymptotic course for both COI and 12S (Figure 4), indicating that the sequencing depth was sufficient for representing the taxonomic complexity of the samples. The ethanol fecal samples reached the lowest number of unique taxa. The water samples generally reached the highest number of unique taxa, but a few fecal samples in the COI dataset had as many or more unique taxa. The rarefaction curves with combined ethanol and fecal samples illustrated that they reached more unique taxa when combined than individually (Figure 5). However, the fecal samples still exhibited lower amounts of unique taxa than the water samples after being combined, with a few fecal samples that surpassed the water samples in the COI dataset.

The species accumulation curves on COI and 12S for water, ethanol, and fecal samples illustrated that none of the sample types in either dataset reached an asymptotic course (Figure 6), indicating that it would have taken more samples to reach the full diversity of unique taxa for each sample type. In the COI dataset, fecal samples reached the highest number of unique taxa, followed by ethanol and water samples. Conversely, in the 12S dataset, water samples exhibited the highest number of unique taxa, and ethanol samples the lowest.

When combining the fecal and ethanol samples, the species accumulation plots illustrated that they reached a higher number of unique taxa for both COI and 12S than individually (Figure 7). The combined samples did not reach an asymptotic course either. Notably, the combined fecal samples yielded almost twice as many unique taxa for COI compared to the water samples. But it should be noted that there were also more than twice as many fecal samples as water samples. In the 12S dataset, it was illustrated that even with less than half of the sample size, the water samples still exhibited the highest number of unique taxa, even when the fecal samples were combined (Figure 6; Figure 7).

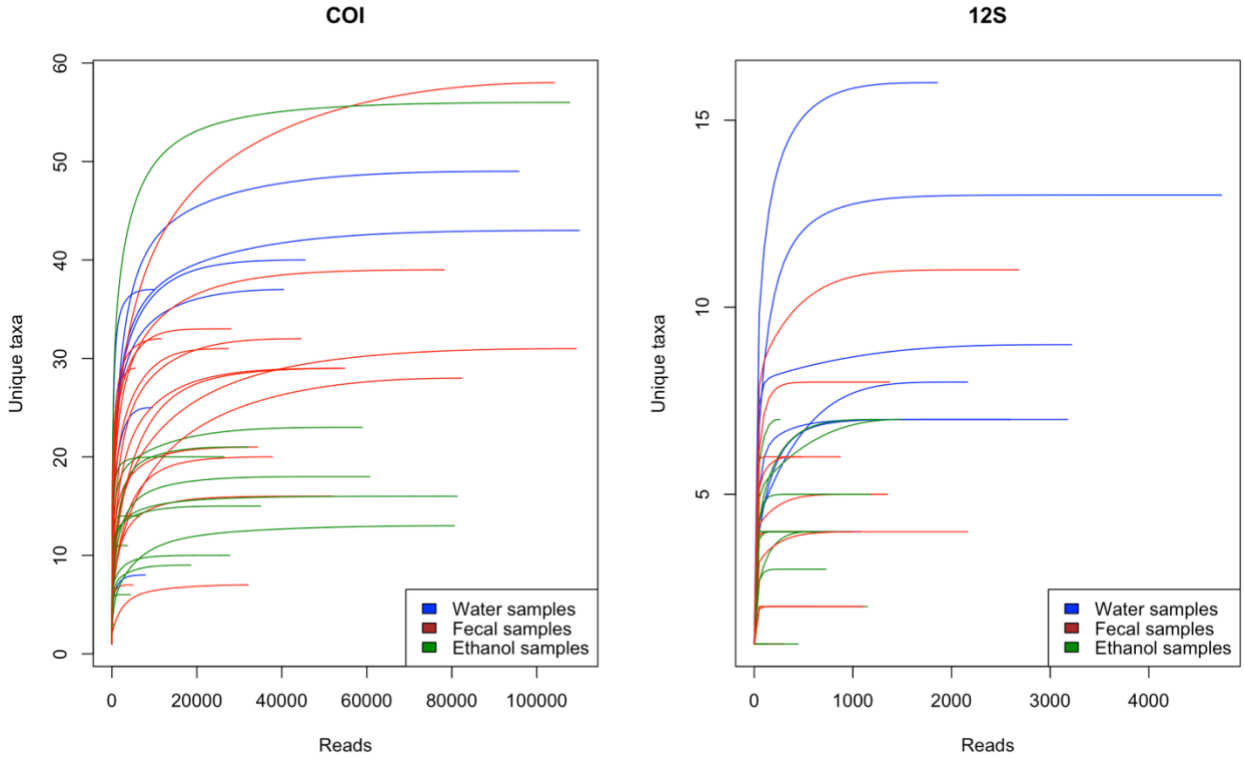



*Figure 4 - Rarefaction curves derived from seven water samples (blue), 16 fecal samples (red), and 16 ethanol samples (green) for the COI dataset (left) and the 12S dataset (right). The curves indicate the accumulation of unique taxa as a function of sequencing effort for each individual sample.*

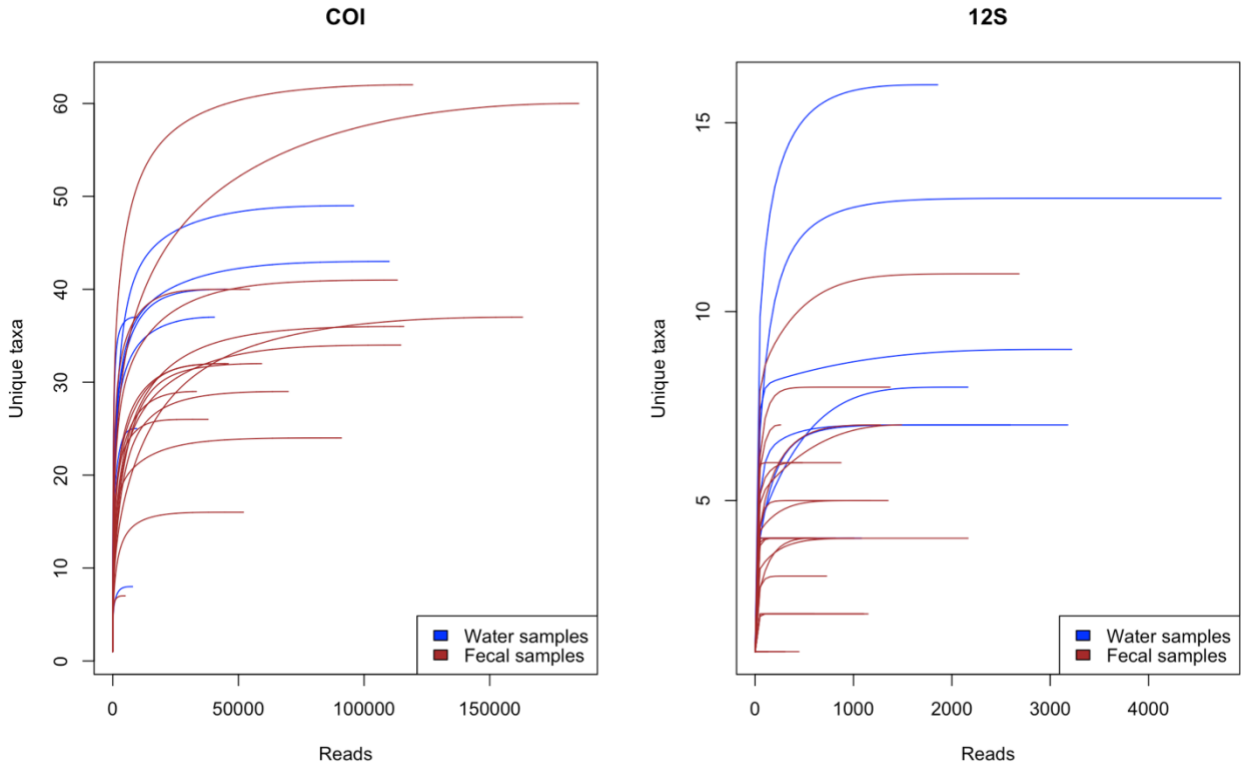



*Figure 5 - Rarefaction curves derived from seven water samples (blue) and 16 fecal samples (red), which have been summed with their corresponding ethanol samples, for COI (left) and 12S (right). The curves indicate the accumulation of unique taxa as a function of sequencing effort for each individual sample.*

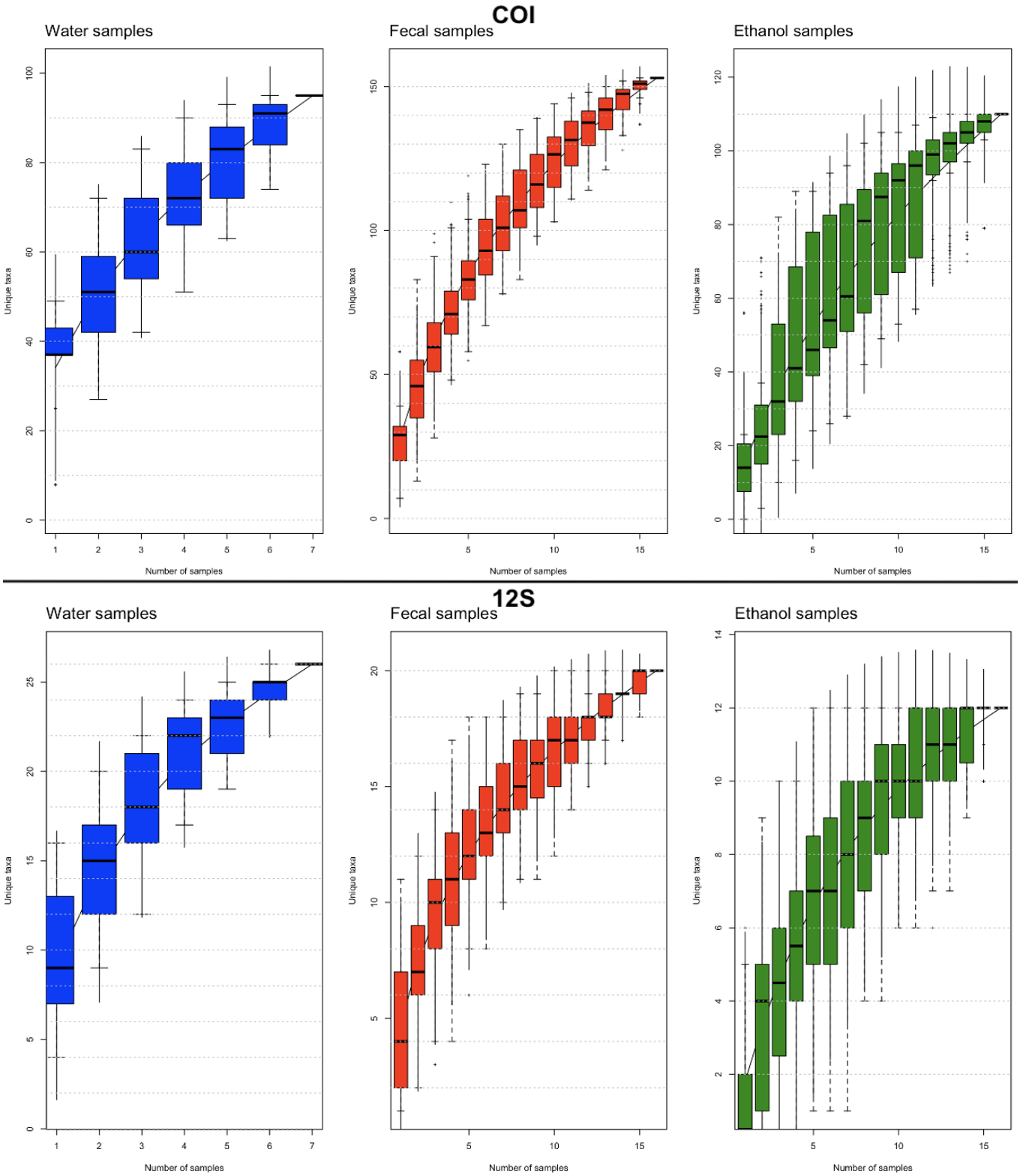



*Figure 6 - Species accumulation curves on COI (top) and 12S (bottom) data for seven water samples (blue), 16 fecal samples (red), and 16 ethanol samples (green). Each curve depicts the accumulation of unique taxa with increasing sampling effort. Box plots illustrate the variability in the number of unique taxa at each number of samples.*

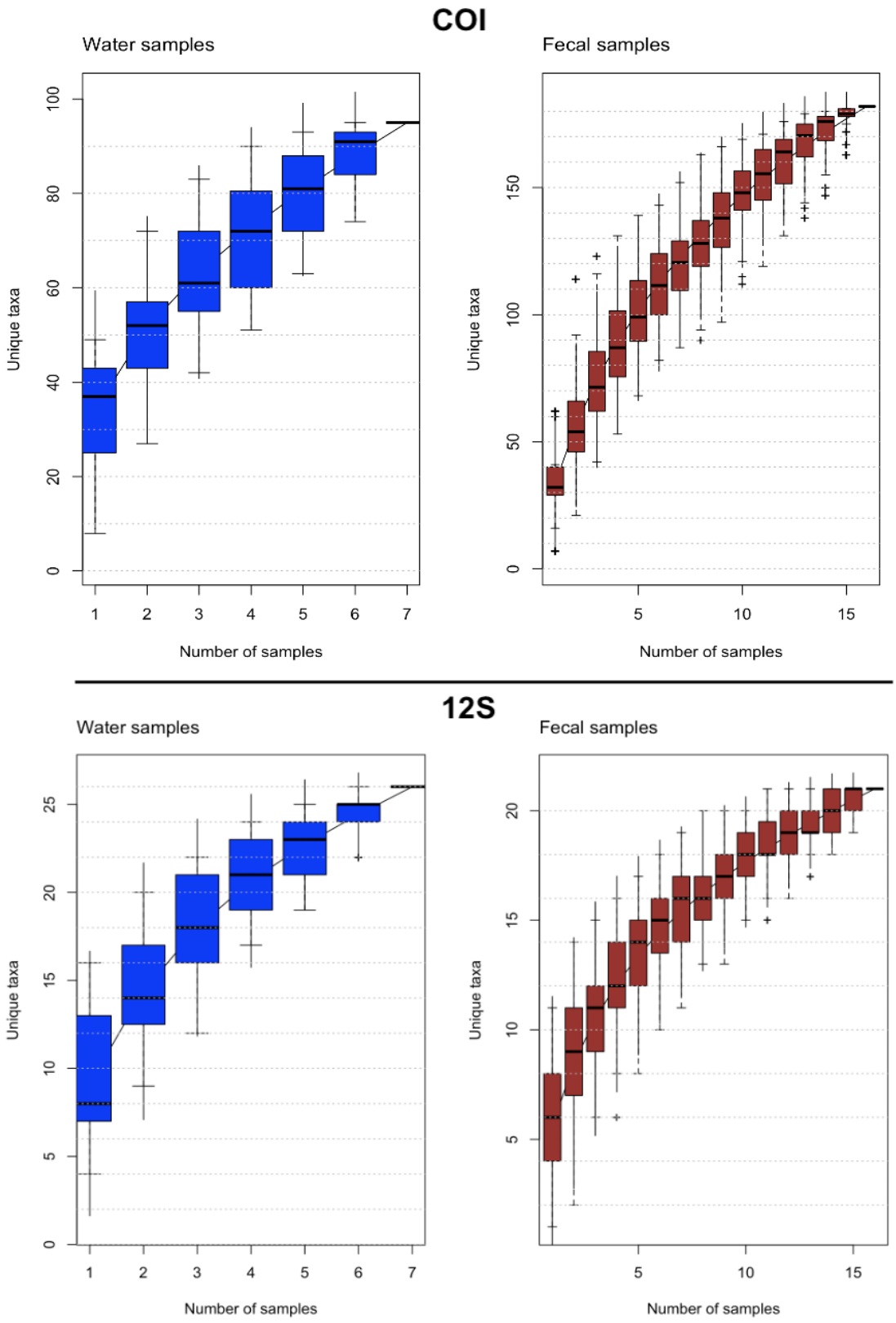



*Figure 7 - Species accumulation curves on COI (top) and 12S (bottom) data for seven surface water samples and 16 combined fecal and ethanol samples. Each curve depicts the accumulation of unique taxa with increasing sampling effort. Box plots show the variability in the number of unique taxa at each number of samples.*

## 3.2 Dissimilarity between sample types

The taxonomic diversity between sample types was examined with stacked barplots and nMDS plots. First, fecal samples and ethanol fecal samples were examined for dissimilarity with a nMDS plot of the COI dataset containing all kingdoms (Figure 8). The R-value close to 0 illustrated that there was very little dissimilarity between the groups. Therefore, the ethanol and fecal samples are treated as one group and will simply be called fecal samples hereafter.

The nMDS plot of the COI dataset with all kingdoms showed a moderate separation between water and fecal samples (Figure 9); however, the ellipses showed a fair amount of overlap, indicating a considerable amount of shared unique taxa between water and fecal samples. The ANOSIM values were calculated as the dissimilarity between water and fecal samples, and the R value indicated a moderate dissimilarity between the two sample groups, which was found to be significant. The stacked barplot of COI data with all taxonomic kingdoms (Figure 10), as well as the SIMPER analysis (Suppl. Table. 4), indicated that the dissimilarity between fecal samples and water samples primarily was driven by *Protista* and *Chlorophyta*, which made up a large proportion of the water samples. These groups were not relevant for the diet of sperm whales, and therefore a modified dataset was used for all further analysis. The only kingdom that was focused on in the modified dataset was Metazoa without sperm whales, and the primary focus point was fish species.

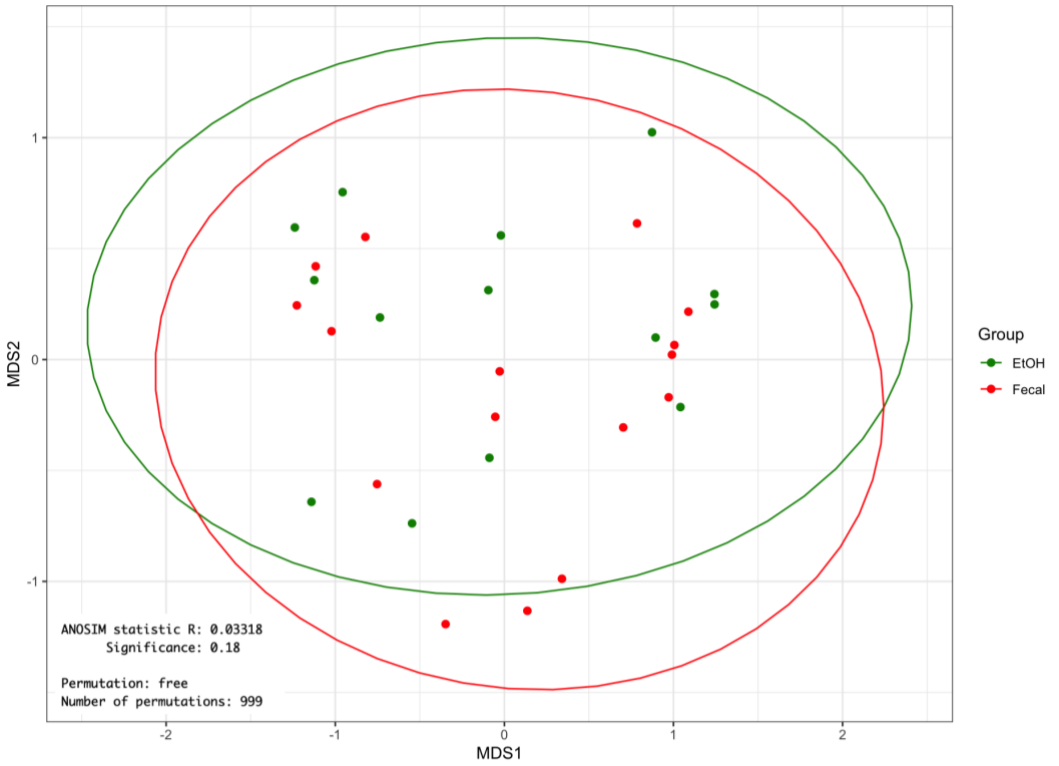



*Figure 8 - nMDS plot of the COI dataset, including all taxonomic kingdoms. Two distinct groups are represented: ethanol (EtOH) and fecal samples. The ANOSIM values are plotted in the bottom left corner of the nMDS plot.*

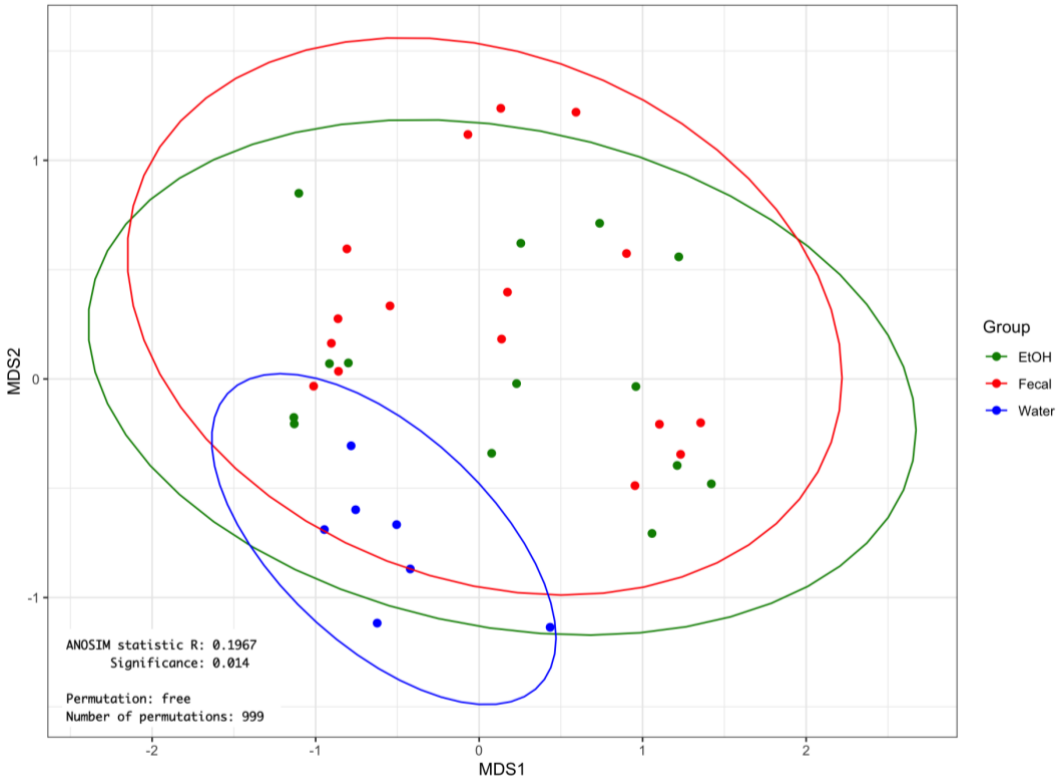



*Figure 9 - nMDS plot of the COI dataset, including all taxonomic kingdoms. Three distinct groups are represented: ethanol (EtOH), fecal, and water samples. The ANOSIM values are plotted in the bottom left corner of the nMDS plot.*

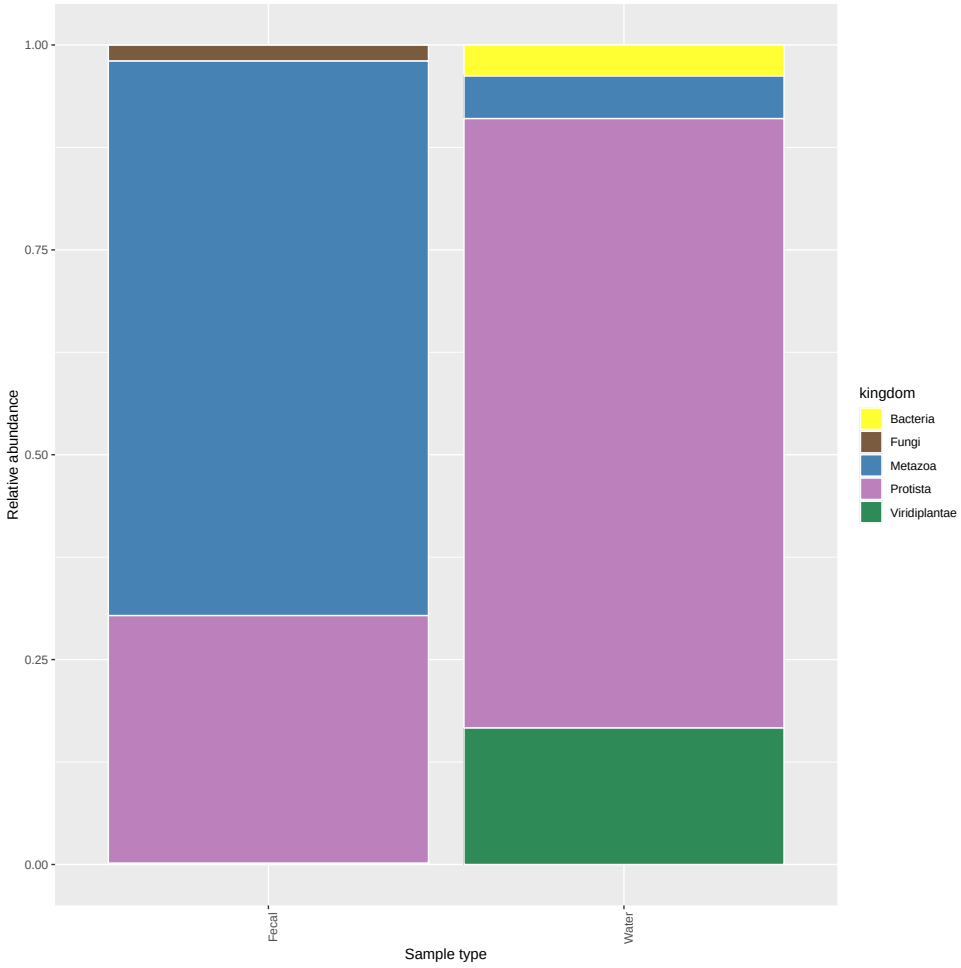



*Figure 10 - Stacked barplot of the COI dataset for all kingdoms showing the relative abundance of five taxonomical kingdoms* (Bacteria, Fungi, Metazoa, Protista, and Viridiplantae) *in 16 fecal samples combined with ethanol fecal samples (Fecal), and seven surface water samples (Water). The y-axis indicates the relative abundance of each kingdom in the sample types. The number of reads for each kingdom is written inside each group.*

## 3.3 ANOSIM analysis

All sample groups, except SG5, which was only utilized in SGC, were examined for dissimilarities with ANOSIM analyses (Table 4). Four P-values from ANOSIM were significant: SG1 from the 12S dataset, SG4 and SGC from the COI-*Metazoa* dataset, and SGC from the COI-fish dataset, indicating that there was a statistical difference in the unique taxa diversity of fecal and water samples. SIMPER analyses were carried out on the significant groups.

*Table 4 - Overview of ANOSIM analyses of SG1–SG4 as well as SGC (SG1–SG5) for 12S (fish only) and COI (metazoans and fish). The COI dataset for fish was insufficient for meaningful analyses and is marked with –, except for the combined groups. Significant P-values (P<0.05) are marked with *.*

| **ANOSIM analysis - COI & 12S** | | | | | |
|---|---|---|---|---|---|
| | **SG1** | **SG2** | **SG3** | **SG4** | **SGC** |
| 12S - fish | | | | | |
| **R-value** | 0.505 | 0.052 | 0.083 | -0.365 | 0.082 |
| **P-value** | 0.009* | 0.338 | 0.400 | 0.962 | 0.208 |
| **Permutations** | 999 | 999 | 119 | 999 | 999 |
| COI - *Metazoa* | | | | | |
| **R-value** | -0.059 | 0.755 | 0.125 | 0.402 | 0.193 |
| **P-value** | 0.567 | 0.126 | 0.400 | 0.028* | 0.039* |
| **Permutations** | 5039 | 999 | 719 | 999 | 999 |
| COI - fish | | | | | |
| **R-value** | – | – | – | – | 0.404 |
| **P-value** | – | – | – | – | 0.046* |
| **Permutations** | – | – | – | – | 999 |

## 3.4 SIMPER analysis

In the SIMPER analyses, the focus was on the species that, when accumulated, contributed to approximately 70% of the dissimilarity between fecal and water samples. The species that were relatively most abundant in the fecal samples, compared to the water samples, were highlighted in bold. The SIMPER analysis for SG4 on the COI-*Metazoa* dataset which only included species from the taxonomic kingdom *Metazoa*, illustrated that seven species contributed to 72.0% of the dissimilarity between fecal and surface water samples (Table 5). None of the species responsible for the majority of the dissimilarity for SG4 in the COI-*Metazoa* dataset were fish. For SGC in the COI-*Metazoa* dataset, it was revealed that one fish species was among the top contributors to dissimilarity (Supp. Table 5)— *Mallotus villosus* (capelin) — however, it was relatively most abundant in water samples.

The SIMPER analysis for SGC on the COI-fish dataset, which only included the taxonomic classes *Actinopteri* and *Chondrichthyes,* showed that five species contributed to 71.5% of the dissimilarity between fecal and surface water samples (Table 6). Two out of five species — *Clupea harengus* (Atlantic Herring) *and Trisopterus esmarkii* (Norway Pout) — were relatively most abundant in the fecal samples (Suppl. Figure 1). *C. harengus* and *T. esmarkii* contributed with 12.18% and 9.15% respectively, to the dissimilarity between fecal and water samples. The SIMPER analysis for SG1 on the 12S dataset illustrated that six fish species contributed to 73.8% of the dissimilarity between fecal and surface water samples (Table 7), with *C. harengus* showing the highest relative abundance in the fecal samples (Figure 13) and contributing with 24.54% to the dissimilarity between the two sample types. Furthermore, C. *harengus* was the species most responsible for the dissimilarity between fecal and water samples in SG1.

*Table 5 - SIMPER analysis on SG4 for the COI-Metazoa dataset. Cumulative sum (cumsum) values are presented for the species contributing to 72.0% of the dissimilarity between water and fecal samples. Cumulative sum contribution values (cumsum contribution) which indicate the amount of dissimilarity contributed by each species, are presented in percentage. Species contributing to dissimilarity from the fecal samples are highlighted in bold.*

| **SIMPER analysis - COI-Metazoa - SG4** | | | |
|---|---|---|---|
| Sequence number | Unique taxa | Cumsum | Cumsum contribution |
| 8 | ***Calanus finmarchicus*** | 0.157 | 15.65% |
| 154 | *Oithona similis* | 0.298 | 14.13% |
| 7,010 | *Bolinopsis sp.* | 0.416 | 11.81% |
| 122 | *Leptothecata sp.* | 0.533 | 11.75% |
| 13 | ***Nanomia cara*** | 0.628 | 9.43% |
| 23,474 | *Balaenoptera acutorostrata* | 0.687 | 5.91% |
| 64 | ***Paracalanus parvus*** | 0.720 | 3.35% |

*Table 6 - SIMPER analysis on SGC for the COI-fish dataset. Cumulative sum (cumsum) values are presented for the species contributing to 71.5% of the dissimilarity between water samples and fecal samples. Cumsum contribution values, indicating the amount of dissimilarity contributed by each species, are presented in percentage. Species contributing to dissimilarity from the fecal samples are highlighted in bold.*

| **SIMPER analysis - COI-fish - SGC** | | | |
|---|---|---|---|
| Sequence number | Unique taxa | Cumsum | Cumsum contribution |
| 4312 | *Lophius piscatorius* | 0.194 | 19.45% |
| 6752 | *Mallotus villosus* | 0.378 | 18.31% |
| 1051 | *Pollachius virens* | 0.501 | 12.38% |
| 10799 | ***Clupea harengus*** | 0.623 | 12.18% |
| 1955 | ***Trisopterus esmarkii*** | 0.715 | 9.15% |

*Table 7 - SIMPER analysis on SG1 for the 12S dataset. Cumulative sum (cumsum) values are presented for the species contributing to 73.8% of the dissimilarity between water and fecal samples. Cumsum contribution values, indicating the amount of dissimilarity contributed by each species, are presented in percentage. Species contributing to dissimilarity from the fecal samples are highlighted in bold.*

| **SIMPER analysis - 12S - SG1** | | | |
|---|---|---|---|
| Sequence number | Unique taxa | Cumsum | Cumsum contribution |
| 1 | ***Clupea harengus*** | 0.245 | 24.54% |
| 3 | *Gadus sp.* | 0.373 | 12.79% |
| 8 | *Sebastes sp.* | 0.492 | 11.84% |
| 17 | *Pleuronectidae sp.* | 0.594 | 10.23% |
| 12 | *Cyclopterus lumpus* | 0.675 | 8.15% |
| 5 | *Pollachius virens* | 0.738 | 6.29% |

## 3.5 Fish diversity

To examine which unique fish taxa were found in fecal and water samples and their relative abundance in each sample type, the diversity of unique taxa assigned to fish was examined with stacked barplots and bubble plots. The stacked barplot of the COI fish dataset illustrated that 17 unique fish taxa were found in total (Figure 11; Suppl. Figure 1), with only four unique fish taxa shared between fecal and water samples. Fecal samples contained nine unique taxa, and water samples contained 12 unique taxa. *Gadidae sp.*, *Sprattus sp.*, *L. piscatorius, and M. villosus* were barely visible on the stacked barplot for fecal samples because of their small relative abundance; however, they were all present in the fecal samples. The species with the highest relative abundance for fecal samples was *Melanogrammus*

*aeglefinus* (haddock), and for water samples, it was *P. virens*. The bubble plots of COI fish data illustrated the relative abundance of the unique fish taxa in fecal and water samples for SG1–SG3 (Figure 12), where eight different unique taxa were found for fecal samples and two for water samples. Six unique taxa were registered for the fecal samples in SG1, which were all exclusive to the fecal samples. *T. esmarkii* was relatively most abundant for fecal samples in SG1 and SG2, and *M. aeglefinus* in SG3.

For 12S, the overall registration of unique fish taxa was higher than for COI. The bubble plot of 12S data for the combined groups illustrated that 28 different unique fish taxa were found in total (Figure 13): 20 from fecal samples and 24 from water samples. Four unique taxa were exclusive to fecal samples: *Salmonidae sp., Reinhardtius hippoglossoides, Zeugopterus norvegicus, and Ciliata septentrionalis.* The unique taxa with the highest relative abundance for fecal samples was *C. harengus*, and for water samples it was *Gadus sp*. The bubble plots of 12S data for SG1–SG3 illustrated that 17 unique fish taxa were registered from the fecal samples in the first three sample groups (Figure 14). For SG1, *C. harengus* was the most abundant species in the fecal samples, followed by *Sebastes sp*. Furthermore, there were three unique taxa exclusive to fecal samples in SG1: *Scomber scombrus* (Atlantic mackerel), *Trisopterus sp.*, and *C. septentrionalis*. For SG2, the most abundant species in the fecal samples was also *C. harengus*, followed by *Gadus sp*. SG2 contained five unique taxa exclusive to fecal samples: *P. virens, Pleuronectidae sp.* (flounders)., *Trisopterus sp.*, *Gadidae sp.*, and *Chimaera monstrosa* (rabbitfish). And for the fecal samples in SG3, the three most abundant species were *C. lumpus, Gadidae sp.*, and *Gadus sp.* SG3 contained four unique taxa exclusive to the fecal samples: *S. scombrus, L. piscatorius, Sebastes sp.*, and *Trisopterus sp.*

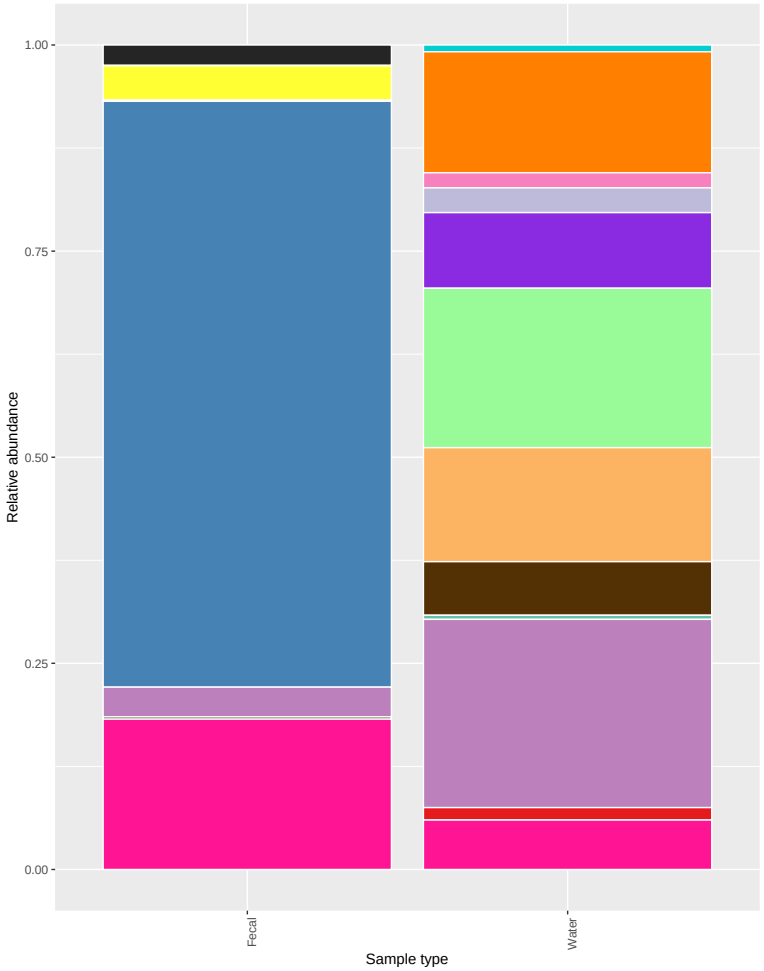



*Figure 11 - Stacked barplot of the COI-fish dataset showing the relative abundance of fish species across seven water samples and 16 fecal samples. The y-axis indicates the relative abundance of each unique taxa within the two taxonomic classes* Actinopteri *and* Chondrichthyes *for the sample types.*

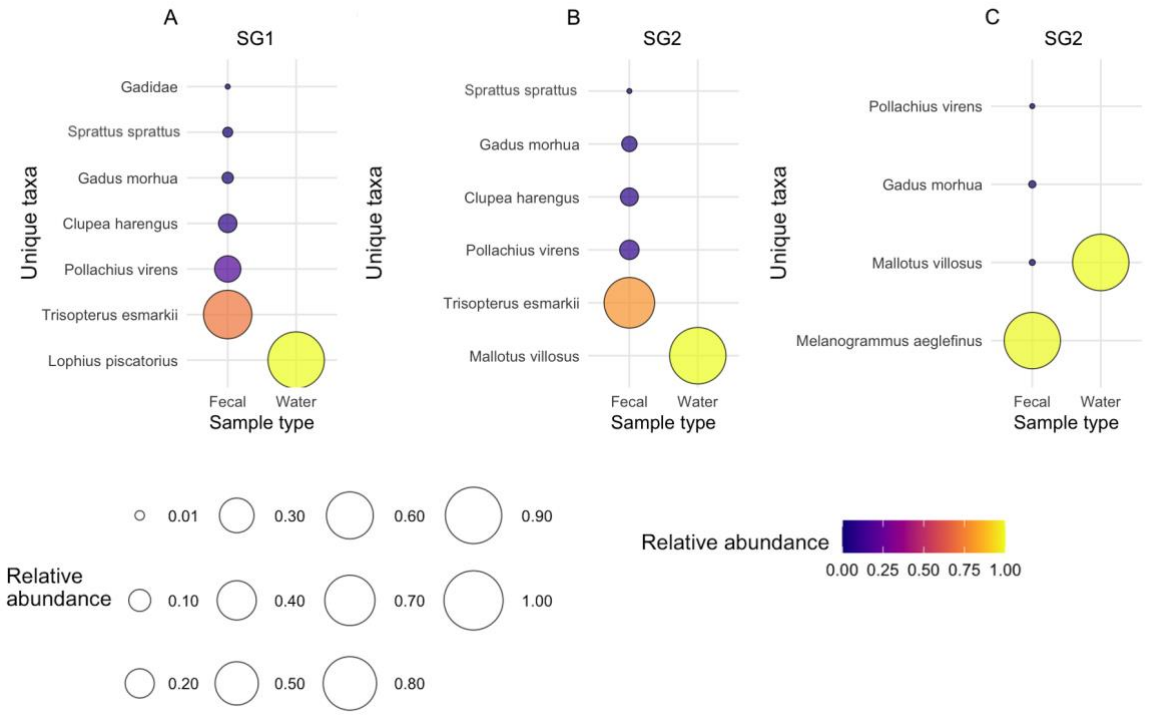



*Figure 12 - Bubble plots of the COI-fish dataset, with bubble size and color indicating the relative abundance of the unique taxa within each sample type. A) SG1, B) SG2, C) SG3.*

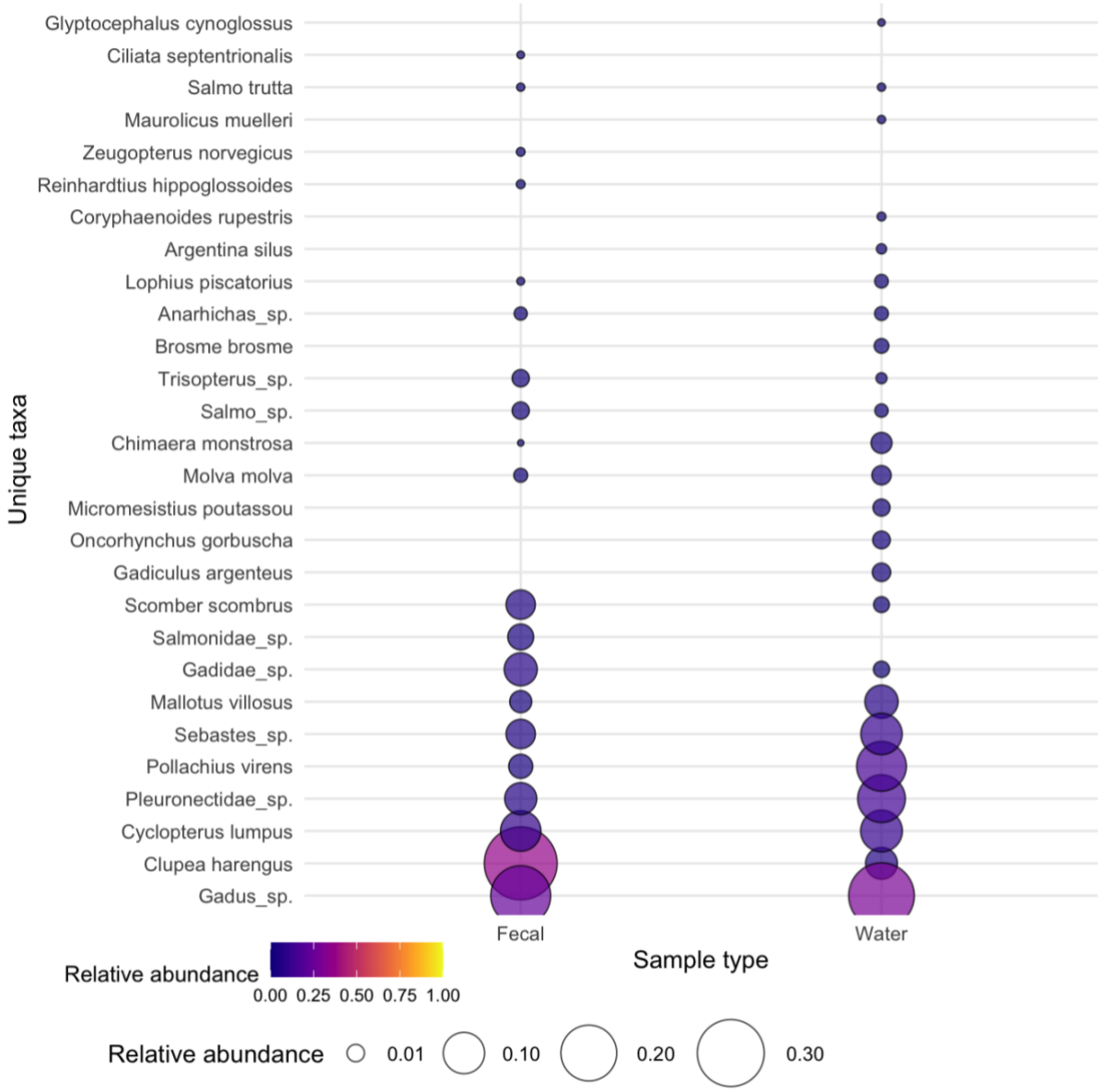



*Figure 13 - Bubble plot of the 12S dataset for SGC (SG1–SG5), with bubble size and color indicating the relative abundance of unique taxa within each sample type.*

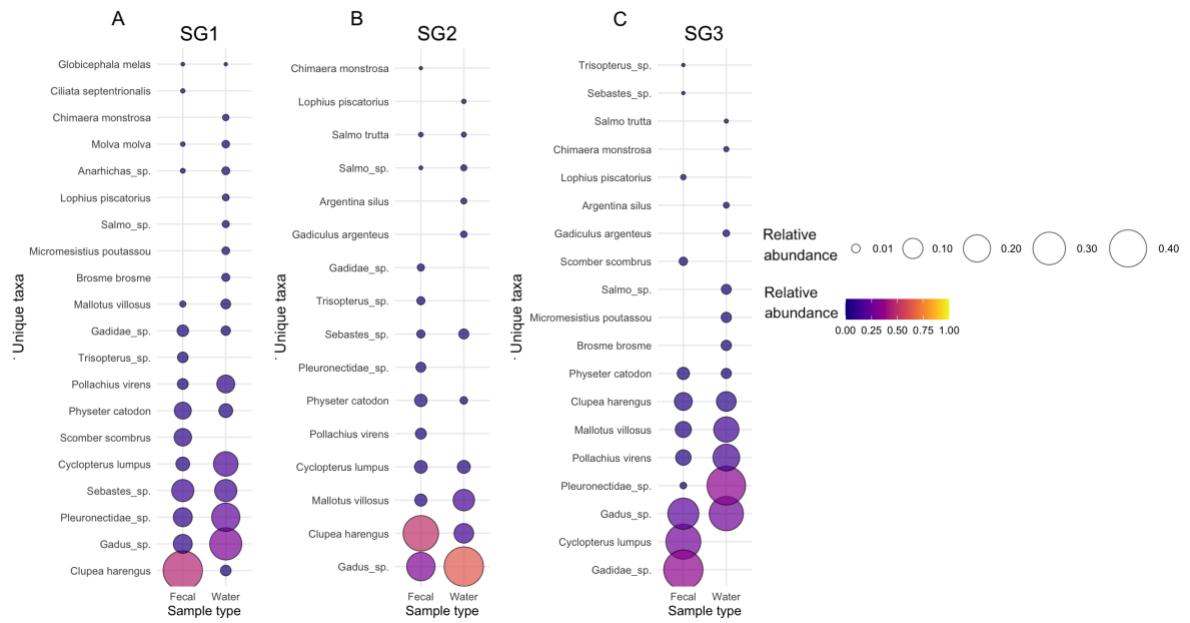



*Figure 14 - Bubble plots of the 12S dataset for the first three sample groups A) SG1, B) SG2, and C) SG3. Bubble size and color indicate the relative abundance of unique taxa within the sample types.*

## 3.6 Decapoda

The stacked bar plot on the COI dataset for *Decapoda* illustrated that five different unique taxa were found (Figure 15): *Chaceon sp.*, *Geryon sp.*, *Liocarcinus sp.*, *Liocarcinus navigator*, and *Liocarcinus pusillus*. All five unique taxa were crabs and were exclusively found in fecal samples. The two most abundant unique taxa were *Geryon sp*. and *L. pusillus.*

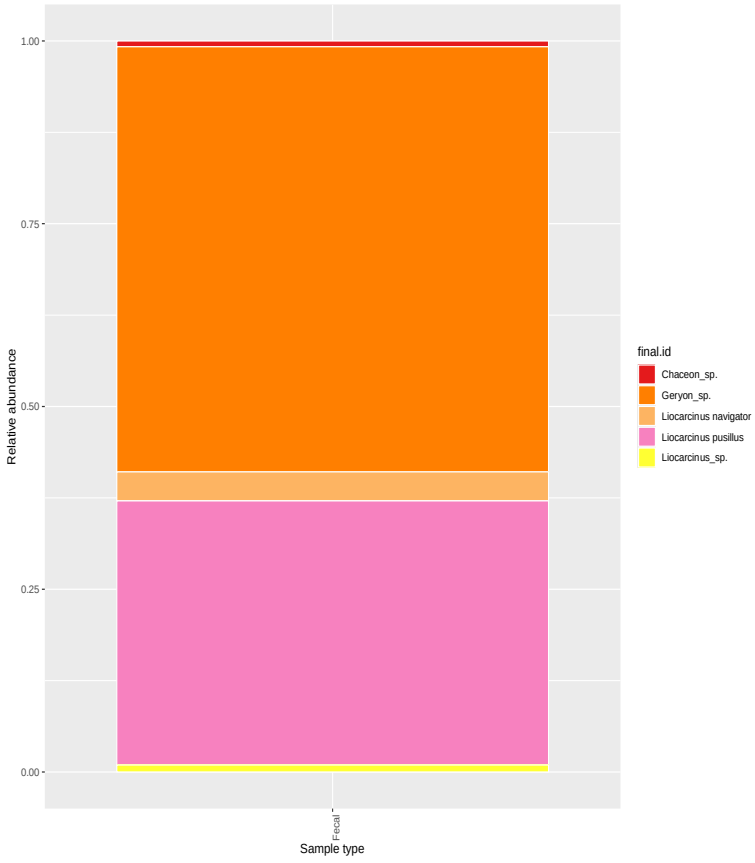



*Figure 15 - Stacked barplot for the COI-Decapoda dataset showing the relative abundance of unique* Decapoda *taxa in 16 fecal samples combined with their corresponding ethanol samples. The y-axis indicates the relative abundance of each unique taxa.*

## 3.7 Possible diet

To identify potential dietary unique taxa, a table was made of the unique fish taxa most likely to be part of the sperm whale diet (Table 8). This table was made with the criteria of exclusivity to fecal samples for either 12S, COI, or both datasets, or appearance in greater relative abundance in the fecal samples for one or both datasets. These unique taxa will be the focus of the discussion when discussing the sperm whale diet.

*Table 8 - Overview of unique fish taxa likely to be part of the sperm whale diet. The presence of unique taxa in each sample type is marked with X if the unique taxa were found on a species level or with ± if the unique taxa were found on a higher taxonomic level: genus or family level, in COI or 12S. Absence in a sample type is marked with –. The number of sample groups (SG1–SG5) the unique taxa appeared in is shown in the group count columns.*

| Fish species appearance in sample groups | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | COI | | | | 12S | | | |
| **Unique taxa** | Fecal samples | Group count | Water samples | Group count | Fecal samples | Group count | Water samples | Group count |
| *Clupea harengus* | X | 3 | – | 0 | X | 5 | X | 5 |
| *Pollachius virens* | X | 4 | X | 1 | X | 5 | X | 4 |
| *Melanogrammus aeglefinus* | X | 2 | X | 0 | ± | 4 | ± | 1 |
| *Trisopterus esmarkii* | X | 2 | – | 1 | ± | 3 | ± | 1 |
| *Gadus morhua* | X | 4 | – | 0 | ± | 5 | ± | 5 |
| *Reinhardtius hippoglossoides* | – | 0 | – | 0 | X | 1 | – | 0 |
| *Zeugopterus norvegicus* | – | 0 | – | 0 | X | 1 | – | 0 |
| *Sprattus sprattus* | X | 2 | – | 0 | – | 0 | – | 0 |
| *Scomber scombrus* | – | 0 | – | 0 | X | 3 | X | 1 |

# 4 Discussion

DNA metabarcoding of fecal samples is a promising new method of analyzing diet in various species (Serite et al., 2023; Cabodevilla et al., 2023). It provides a non-invasive method of investigating diet, which is useful for studying endangered species. It is especially useful for species with behaviors such as the sperm whale, which feeds at great depths (Watwood et al., 2006), making observation of foraging difficult. The overall objective of this study was to investigate the diet of sperm whales using DNA metabarcoding of fecal samples using the COI and 12S metabarcoding markers. The results revealed that 24 unique fish taxa were found for fecal samples in total: 20 from the 12S dataset and nine from the COI dataset. Both the 12S and the COI datasets contained unique taxa that were exclusively found in fecal samples. Furthermore, it was revealed that the water samples contained a greater abundance of unique benthic fish taxa than the fecal samples. This study was the first attempt to investigate the diet of sperm whales with metabarcoding of DNA from fecal samples. This will shed light on the diet of sperm whales in the area of Bleik Canyon in the Norwegian Sea and on the effectiveness of fecal sample metabarcoding. I will discuss the results in relation to what has previously been observed in analyses of sperm whale stomach content.

## 4.1 Comparing species richness of sample types

The individual ethanol and fecal samples generally exhibited lower unique taxa counts than surface water samples for both 12S and COI. This indicated a lower species richness for the two types of fecal samples compared to water samples. The species accumulation curves revealed that the cumulative number of unique taxa counts in the COI dataset for ethanol and fecal samples was higher compared to those from water samples. However, it is important to note that the number of water samples was almost half that of both fecal sample types. Furthermore, the 12S data revealed that water samples had more unique taxa than fecal samples.

There are many possible explanations for why the read counts were generally lower for fecal samples compared to water samples, even though the opposite was expected. It was observed that water samples had a far greater proportion of *Viridiplantae* and *Protista,* which is adding to the diversity of surface water samples. One possible explanation for this could be found in the sampling methods. Since fecal samples were taken with a plankton net, many species of microplankton that were caught for the water samples might have been missed for fecal

samples. Conversely, the plankton net might also catch more unique taxa of zooplankton since it covered a larger area during the sampling than the surface water sampling did. However, that does not explain why the number of registered unique taxa was lower for fecal samples compared to water samples in the 12S dataset which amplified the 12S ribosomal RNA region in vertebrate species. It was expected that the fecal samples contained more unique vertebrate taxa, as the fecal matter was thought to increase the number of species in the surface water. One reason could be that it was due to randomness since the sample size is not very large (Erickson et al., 2019). A small sample size increases the chance of large variations, which might result in the surface water containing a lot of fish DNA due to random variation in the specific places they were sampled. Furthermore, the fecal samples quickly got diluted because sperm whales do not produce solid fecal matter but rather buoyant fecal plumes (Roman & McCarthy, 2010), which may be dispersed by wave action, currents, and the propulsion from the tail of the sperm whale. This means that you only get a very small fraction of the total fecal matter from the sperm whale. Therefore, you are more likely to get reads from the species that are most abundant when taking the sample. Another possibility is that the DNA degradation during the transport and digestion of food items through the digestion system of the sperm whale made the DNA traces harder to amplify. Contrasting the undigested DNA traces from fish in the surface water, coming from scales and tissue (Wang, Yan et al., 2021). The explanation can possibly not be found in one factor alone; many different factors might have affected this unexpected outcome of the unique taxa counts. However, these findings contradicted the initial hypothesis, which suggested that fecal samples would display a greater amount of diversity than surface water samples due to the addition of fecal matter to the water. Therefore, the hypothesis is rejected.

## 4.2 Ethanol samples compared with fecal samples

This study also aimed to compare the species diversity of ethanol samples with regular fecal samples to understand whether the ethanol samples were as effective at yielding a high diversity of unique taxa as regular fecal samples. It is interesting to examine since ethanol samples are easier to make because they do not require homogenization of the fecal matter, and it would therefore make extraction of DNA from fecal samples easier. For the COI data, no significant difference was found between ethanol and fecal samples when comparing them with nMDS and ANOSIM. However, it was discovered that ethanol fecal samples consistently detected fewer unique taxa than regular fecal samples for both COI and 12S. Additionally, five ethanol samples for the 12S data and three from the COI data contained no

reads. The ethanol samples were not without use since they contained unique taxa that had not been detected in the fecal samples. These findings suggest that ethanol samples would not be suitable for standalone use in diet analysis. However, the results found herein suggested that a combination of data from ethanol samples and fecal samples would be valuable for future studies, as they combined detected a higher diversity compared to using the fecal samples alone. One shortcoming of the ethanol samples was the sampling procedure for the fecal samples. The ethanol samples were derived from ethanol that was secondarily added to the fecal matter. Initially, ethanol was used to transfer the fecal matter from the plankton net into the sample jar, where it became mixed with the seawater from the fecal matter. Subsequently, the ethanol was poured out, and new ethanol was added to the fecal matter. The initial ethanol, which had been mixed with seawater, likely contained a greater diversity of unique taxa compared to the ethanol that was added afterwards. The results were not aligned with the hypothesis, stating that ethanol samples would display the same diversity in unique taxa as the fecal samples. Therefore, the hypothesis is rejected.

## 4.3 Dissimilarities between fecal and water samples

There was a difference in the relative abundance of the communities between fecal and water samples for SGC when looking at all the taxonomic kingdoms together, and the ANOSIM analysis indicated that the distinction was moderate. However, there was not significant dissimilarity for all sample groups in the different datasets. If only including the two classes of fish, *Actinopteri* and *Chondrichthyes*, there was a significant dissimilarity for SGC in the COI dataset, indicated by the R-value of 0.409. However, fecal and water samples were very similar for 12S when looking at SGC, but for SG1, which had the most reliable comparison, there was a clear and significant distinction between the two sample types.

It was expected that the distinction between fecal samples and water samples would have been clearer for COI and especially for 12S since it has been documented that the fish species that sperm whales, in the Norwegian Sea, primarily prey on are benthic (Roe, 1969; Similä et al., 2022). It was therefore hypothesized that the fecal samples would have the greatest relative abundance of benthic fish, while water samples would exhibit a higher relative abundance of pelagic fish species. Contrary to this expectation, water samples exhibited the highest relative abundance of benthic fish such as *Pleuronectidae sp.*, *C. lumpus*, and *Sebastes sp*., compared to fecal samples. An explanation for finding more benthic species in the water samples could be that the samples were sometimes taken near fishing vessels.

*Pleuronectidae sp.*, *C. lumpus*, and *Sebastes sp*. are target species for fishing vessels in the Norwegian Sea (Similä et al., 2022), and DNA will naturally shred from the fish when they are brought up, potentially leading to detection of the species in the water samples. Furthermore, there are other toothed whales in the area where the sampling took place, such as *Orcinus orca* (killer whale) (Cosentino, 2015) and *Globicephala* (pilot whale) (Moors-Murphy, 2014), and feces from those whales might include benthic fish species, which could then be detected by the water samples. Lastly, many benthic fish species have pelagic larvae that drift in the water column in their early life stages (Leis, 2006), and these might have been caught in the surface water samples. However, since the fecal samples were sampled with a plankton net that covered a larger area than the water sampling method, it should have caught more fish larvae than the water sampling. In contrast, the species with the highest relative abundance in the fecal samples for the 12S dataset was the pelagic species *C. harengus* (Olsen et al., 2007). It is therefore possible that sperm whales feed on pelagic fish to a greater extent in the Norwegian Sea than previously believed. Alternatively, it may also be that *C. harengus* serves as a secondary prey for sperm whales. It is possible that other species that sperm whales prey on, such as cephalopods, may rely on *C. harengus* as a food source, contributing to its presence in the fecal samples. In any case, since the results revealed a higher proportion of benthic fish species in the water samples compared to the fecal samples, the hypothesis that fecal samples contained a higher number of unique taxa from benthic species compared to water samples was rejected.

## 4.4 Contamination

Since the path from sampling to results involved many steps, various stages in the process are susceptible to contamination. Four types of blanks — air, ethanol, extraction, and PCR blanks — were utilized throughout the sampling and laboratory processes to monitor contamination levels. They generally showed no signs of significant contamination, but the steps most prone to contamination seemed to be the DNA extraction for COI and filtering of the water samples for 12S; however, none of the contamination remained after curation of the datasets. Therefore, there has not been any notable contamination compromising the results. However, contamination might have happened in steps that were not tested. For instance, the same plankton net was used for all the fecal samples, and although it was rinsed between each sample, there might still be a risk of carryover contamination. No blanks were made for checking the plankton net after rinsing, which might have proven valuable for discovering if it was a source of contamination.

## 4.5 Sampling bias

The variance in the sampling methods for surface water and fecal samples presents a sampling bias. Fecal samples were taken with a plankton net that covered a greater area than the water sampler during the sampling, complicating the comparison between the fecal and water samples. Furthermore, the mesh size of the plankton net is not small enough to catch all microplankton, thereby making *Chlorophyta* underrepresented in the fecal samples. One approach to solving this complication could be standardization of the sampling by having the same sampling method for surface water samples and fecal samples. Fecal samples could be sampled with a water sampler, just like the water samples. This would eliminate the bias, but it would possibly result in fewer reads from the fecal samples.

## 4.6 Sperm whale diet

After analyzing the datasets, each of the species that was possibly a prey item for sperm whales was examined, and the different analyses and plots were compared to get a better understanding of the likelihood of the specific species being part of the sperm whale diet. It was of great importance to determine which sample groups the different species occurred in. SG1 was the only group that had directly comparable fecal and water samples and was therefore the group that weighed the most in the interpretation of the results. The other groups had their value as well, as they could be used to determine if the different unique taxa were occurring consistently in the samples.

### 4.6.1 *Gadus morhua*

One possible prey species was *G. morhua*, which is a widespread species in the Norwegian Sea (Nordeide et al., 2011) and has been found in the diet of sperm whales in several stomach content analyses in the Norwegian Sea and Icelandic waters (Similä et al., 2022; Roe, 1969). In the COI dataset, it was registered exclusively for fecal samples and was found in four sample groups, including SG1 and SG2, which are the most reliable. It was not detected in the 12S dataset at the species level; however, its genus, *Gadus*, and family, *Gadidae*, were registered in all five sample groups for both water and fecal samples, and some of those unique taxa were likely *G. morhua*, though it was not possible to say for certain. But due to the fact that *G. morhua* was found exclusively in fecal samples in the two most reliable sample groups in the COI dataset and the fact that it has been observed in the diet of sperm whales in the Norwegian Sea previously, it seemed highly likely that it could be a part of the diet of sperm whales in the Bleik Canyon area in the Norwegian Sea.

### 4.6.2 *Sprattus sprattus*

The small pelagic fish species *S. sprattus* (sprat) is a widespread species in Norway (Glover et al., 2011). There was no registration of *S. sprattus* in the 12S dataset, but it was found in the two first sample groups for the COI dataset, which are the most reliable, and was not found in any of the surface water samples. Therefore, *S. sprattus* could likely be part of the diet of sperm whales in the Norwegian Sea. There has not been any registration of *S. sprattus* in sperm whale diet before. However, *S. sprattus* is a rather small fish species and is an important food source for *G. morhua* (Pachur & Horbowy, 2013), meaning that *S. sprattus* could have been found in the diet of the sperm whales because it had been consumed by *G. morhua*. Furthermore, *G. morhua* was found in the same two sample groups as *S. sprattus*, which would be expected if *G. morhua* had been feeding on *S. sprattus*. Since sperm whales are known to feed on *G. morhua* (Similä et al., 2022), it is possible that *S. sprattus* is a secondary prey for sperm whales. However, it cannot be ruled out that sperm whales themselves feed on *S. sprattus* as well.

### 4.6.3 *Clupea harengus*

Another likely species to be a part of the diet of sperm whales in the Norwegian Sea was *C. harengus,* which is a pelagic species that is common in the Norwegian Sea (Olsen et al., 2007). There was found presence of *C. harengus* in the fecal samples in both the 12S and COI datasets, particularly in SG1 where it was exclusively found for fecal samples in COI and had a much higher relative abundance compared to water samples for 12S. These results suggested that it was likely a part of the sperm whale diet. This was supported by the results from the SIMPER analyses, which indicated that it played a large role in the dissimilarity between fecal and water samples. Previously, *C. harengus* has been found in the stomach content of a stranded sperm whale close to the area where the fecal samples were obtained (Similä et al., 2022), which further strengthens the likelihood of *C. harengus* being a prey for sperm whales in the Norwegian Sea.

### 4.6.4 *Trisopterus esmarkii*

*T. esmarkii,* a member of the family *Gadidae,* was the species with the second-highest abundance for fecal samples in the COI dataset, where *T. esmarkii* was found in three sample groups for fecal samples and one sample group for water samples. These results, and specifically the fact that *T. esmarkii* was found exclusively for fecal samples in SG1, indicated that *T. esmarkii* was potentially a prey for sperm whales in the Norwegian Sea. In

the 12S dataset, *Trisopterus* was only found at the genus level (*Trisopterus sp.)*, but it is likely that it was also *T. esmarkii* that was registered, since *Trisopterus sp*. was found in the same sample groups where *T. esmarkii* was found. The relative abundance was not as high in the 12S dataset, but it was also found exclusively in fecal samples for 12S in SG1, SG2, and SG3. Therefore, *T. esmarkii* is likely a part of the sperm whale diet in the Norwegian Sea. This is interesting, since *T. esmarkii* has not been found as a prey item for sperm whales in previous studies.

### 4.6.5 *Scomber scombrus*

The fish species *S. scombrus* was found in three sample groups in the 12S data for fecal samples, and one for water samples. Specifically, it was exclusively detected in fecal samples in SG1, which directly relates to water samples. Therefore, it seems possible that *S. scombrus* was consumed by sperm whales and could therefore have been a likely prey for the sperm whales. Further strengthening the results, the genus *Scomber* has been found in sperm whale stomachs previously (Silas et al., 1988). And since *S. scombrus* is a common species in the Norwegian Sea (Bjørdal et al., 2022), it is not unlikely that sperm whales consume *S. scombrus*. However, no evidence of *S. scombrus* was found in the COI dataset, and it was not among the top contributors to dissimilarity in any of the SIMPER analyses.

### *4.6.6 Pollachius virens*

Another fish species that could be part of the sperm whale diet in the Norwegian Sea is *P. virens*. This species is common in the Norwegian Sea and in the sampling area. It was found in the fecal samples in four sample groups for the COI dataset and exclusively for fecal samples in the first three sample groups (SG1–SG3), but only occurred in one group for the water samples. These results indicated that *P. virens* is likely a prey item for sperm whales in the Norwegian Sea. The results are further strengthened by *P. virens* previously having been identified in the stomach of a stranded sperm whale in the Netherlands and in the Norwegian Sea (Santos et al., 2002; Roe, 1969). However, *P. virens* had the highest relative abundance in the water samples compared to the fecal samples in SGC for both COI and 12S. Considering the higher relative abundance in the 12S and COI datasets, it seemed uncertain as a prey item for sperm whales in the Norwegian Sea, and it would require more research to reach a decisive conclusion.

### 4.6.7 *Melanogrammus aeglefinus*

The species with the highest relative abundance for fecal samples in the COI-fish dataset was *M. aeglefinus*, a member of the family *Gadidae*, and it was found in two fecal sample groups and was not registered in the water samples. However, since *M. aeglefinus* was found in SG3 and SG5, which do not include directly comparable water samples, it was not possible to conclude that *M. aeglefinus* was consumed by the sperm whales. There has previously been one uncertain observation of *M. aeglefinus* from a sperm whale stomach analysis (Roe, 1969), where it could have been a misidentified *G. morhua*, but has otherwise not been observed as a prey item for sperm whales. Furthermore, *M. aeglefinus* was only identified at the family level in the 12S dataset and was found in four fecal sample groups and two water sample groups. The results indicate that *M. aeglefinus* potentially could be a prey item for sperm whales in the Norwegian Sea, but the results are not clear, and it would require further research to reach a definitive conclusion.

### 4.6.8 *Decapoda*

Five unique taxa of *Decapoda* were found in the COI dataset: *Chaceon sp.*, *Geryon sp.*, *Liocarcinus sp.*, *Liocarcinus navigator*, and *Liocarcinus pusillus*, which were all crab species. They were found in SG3 and SG4 and were only found in fecal samples. Crabs have been found in the diet of sperm whales before (Evans & Hindell, 2004), but whether they are ingested incidentally while hunting other prey types, are part of the diet of the fish or cephalopods that the sperm whales consume, or are a primary prey item for sperm whales is still not clear (Evans & Hindell, 2004).

### 4.6.9 *Cephalopoda*

There were no signs of *Cephalopoda* in the COI results. It was expected that *Cephalopoda* would be registered in the COI data since they have been found in the form of squid beaks in the stomachs of stranded sperm whales close to the area where the fecal sampling took place (Similä et al., 2022). It can mean that sperm whales do not feed on *Cephalopoda* to a great extent in the Norwegian Sea, but since they have been found in the diet in the Norwegian Sea previously, it was also likely to be a problem with the COI sequencing. COI is a great genetic marker for identifying species among all taxa, but it is not perfect. Some species of *Cephalopoda* have proven to be more difficult to amplify with COI than others (Xu et al., 2017), and that might have been the case for the *Cephalopoda* species in the Norwegian Sea. Therefore, it would be valuable to include a *Cephalopoda*-specific primer set such as

Ceph18S (de Jonge et al., 2021) in future research on the sperm whale diet to be able to check with greater certainty if they are feeding on *Cephalopoda* in the Bleik Canyon area of the Norwegian Sea. Another possibility is that sperm whales could be changing their diet throughout the year. Since the fecal samples only covered five months, there are more than half of the year that has not been samples in. Some of the months that were not sampled in were potentially the months where they were feeding on *Cephalopoda*.

Generally, *Cephalopoda* are believed to be a great part of the diet of sperm whales (Cherel, 2021), which has mainly been based on findings of *Cephalopoda* beaks in stomach content analyses. However, there might be a bias in stomach content analyses since *Cephalopoda* beaks are hard chitinous structures that are hard to digest. Therefore, they probably have a higher retention time than other parts of prey (Gibbs, 2007), meaning they will accumulate in the stomach. It has been documented that a rescued *Neophoca cinera* (Australian sea lion) that was fed only fish for 14 days, exclusively had *Cephalopoda* beaks in its stomach upon its death (Gibbs, 2007). This could likely be the case for other species as well. This means that previous studies on the sperm whale diet with stomach content analysis might have been biased towards a greater proportion of *Cephalopoda*. This would make it seem like *Cephalopoda* was a bigger part of the diet of sperm whales than it actually was. However, metabarcoding on fecal samples could be a good way to work around that bias in the future since it doesn't take accumulated beaks in the stomach into account. But it would require that the correct genetic markers were used.

## 4.7 Limitations and strengths of metabarcoding on fecal sample DNA

DNA metabarcoding analysis for fecal samples is a great, non-invasive method for examining diet. One of the strengths of metabarcoding is that it works well with animals that are difficult to observe in their foraging behaviors, such as sperm whales. Furthermore, metabarcoding can detect species in fecal samples that would not be possible to identify with traditional methods because they have been digested. This study discovered species that had not been associated with the sperm whale diet before, such as *T. esmarkii* and *S. sprattus*. However, metabarcoding also has limitations. For instance, fecal samples represent a very small part of the whole fecal matter from a sperm whale, since fecal matter from a sperm whale is liquid in structure and therefore gets diluted by water movements (Roman & McCarthy, 2010). To get the most concentrated fecal sample, it is important to be as close to the defecation point as

possible and sample it as fast as possible. However, that is no easy task. Therefore, it is likely the most abundant species within the fecal matter that gets registered with metabarcoding, and some of the rarer species might not be detected.

Furthermore, the choice of genetic markers is not easy either. COI is a good marker for getting the widest possible diversity. However, some of the most abundant ASVs in this study, such as ASVs three and four, were not identified. This skewed the overall data, possibly making it seem like some taxonomic groups were relatively more abundant than they actually were. In addition, COI is not good at detecting fish (Grey et al., 2018) in comparison to the 12S marker, as was demonstrated in this study. The 12S marker is great at detecting fish, but it also has its limitations, such as difficulties with differentiating some species, such as the species within the family *Gadidae* (Wang et al., 2017), which was demonstrated in this study as well. Furthermore, this study demonstrated that 12S had problems discerning species within the taxonomic groups *Sebastes, Trisopterus, Salmonidae, Anarhichas*, and *Pleuronectidae*. This was especially a problem when several of the species within a family that cannot be differentiated live within the same habitat range, which made it impossible to determine the unique taxa at the species level. Furthermore, neither COI nor 12S are good at registering cephalopods, for which you need another marker. When compared with traditional diet analysis methods, the metabarcoding method misses out on a lot of valuable information, such as the size, age, and number of each species (Franklin et al., 2019). Furthermore, there is the problem of not knowing whether the species found in the fecal samples are primary or secondary prey. Therefore, traditional methods of diet analysis are still valuable, especially in combination with metabarcoding approaches. The traditional methods can be used to determine which species the sperm whale eats, and the metabarcoding approach can then be used to determine how often those species occur in the fecal matter, determining which species are the most important prey.

## 4.8 Further research

The results of this study can only be used to get insight into the diet of the sperm whales in the Bleik Canyon area of the Norwegian Sea since other potential prey species, which the sperm whales may prefer, inhabit other habitats. For future research on the sperm whale diet, it would be interesting to investigate the diet with DNA metabarcoding on fecal samples throughout its entire habitat range to gain broader insights into its diet. Additional primers could be utilized, such as the Ceph18S primer for amplifying *Cephalopoda* sequences.

Furthermore, other analyses could be utilized, such as quantitative fatty acid analysis (QFASA) and stable isotope analysis (SIA) on biopsies. It would be especially interesting to analyze biopsies from whales that there are also fecal samples from, to be able to compare the results from metabarcoding with the results from QFASA and SIA. Lastly, it would be interesting to standardize the sampling methods of surface water samples and fecal samples to allow for a more direct comparison between the sample types.

# 5 Conclusion

It can be concluded that ethanol samples were not as good as fecal samples for diet analysis, but when combined, they registered more unique taxa than fecal samples did alone. It can also be concluded that fecal samples did not detect more unique taxa than water samples, therefore rejecting the hypothesis that fecal matter would yield a higher number of unique taxa than water samples. In this study, it was discovered that benthic fish species seemed to play a smaller role in the diet of sperm whales than hypothesized. However, it cannot be concluded whether it is due to sperm whales relying more on pelagic species as prey than previously believed, or if it is due to other factors affecting the results. It was also discovered that *Gadus morhua* (Atlantic cod), *Sprattus sprattus* (sprat), *Clupea harengus* (Atlantic herring), *Trisopterus esmarkii* (Norway pout), *and Scomber scombrus* (Atlantic mackerel) seemed likely to have been a part of the diet of the sperm whales in the Norwegian Sea in the area of Bleik Canyon, either as primary or secondary prey items. Notably, *T. esmarkii* and *S. sprattus*, had not been registered in the diet of sperm whales before, and this study therefore adds two new potential prey species to the diet of sperm whales. *However*, more research and potentially a fresh approach to sampling are needed before definitive conclusions on the diet of sperm whales in the Bleik Canyon area of the Norwegian Sea can be drawn.

# Works cited


Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215(3), 403-410.

Bjørdal, V. R., Mørk, H. L., Utne, K. R., Fernö, A., & Nøttestad, L. (2022). The diet of juvenile Atlantic mackerel (Scomber scombrus) feeding in new northern nursery areas along the Norwegian coast. *Marine Biology Research*, *18*(7-8), 415-425.

Bjørke, H. (2001). Predators of the squid Gonatus fabricii (Lichtenstein) in the Norwegian Sea. *Fisheries Research*, *52*(1-2), 113-120.

Brito, C., Jordão, V. L., & Pierce, G. J. (2016). Ambergris as an overlooked historical marine resource: its biology and role as a global economic commodity. *Journal of the Marine Biological Association of the United Kingdom*, *96*(3), 585-596.

Cabodevilla, X., Gómez-moliner B.J., Abad N., Madeira M.J. (2023). Simultaneous analysis of the intestinal parasites and diet through eDNA metabarcoding. *Integrative Zoology*, 18(3), 399-413.

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature methods*, 13(7), 581-583.

Cawthorn, D. M., Steinman, H. A., & Witthuhn, R. C. (2012). Evaluation of the 16S and 12S rRNA genes as universal markers for the identification of commercial fish species in South Africa. *Gene*, 491(1), 40-48.

Cherel, Y. (2021). Revisiting taxonomy of cephalopod prey of sperm whales caught commercially in subtropical and Southern Ocean waters. Deep Sea Research Part I: *Oceanographic Research Papers*, 169, 103490.

Clare, E. L., Economou, C. K., Faulkes, C. G., Gilbert, J. D., Bennett, F., Drinkwater, R., & Littlefair, J. E. (2021). eDNAir: proof of concept that animal DNA can be collected from air sampling. *PeerJ*, *9*, e11030.

Cordain, L., Watkins, B. A., Florant, G. L., Kelher, M., Rogers, L., & Li, Y. (2002). Fatty acid analysis of wild ruminant tissues: evolutionary implications for reducing diet-related chronic disease. *European Journal of Clinical Nutrition*, *56*(3), 181-191.

Cosentino, A. M. (2015). First record of Norwegian killer whales attacking and feeding on a harbour porpoise. *Marine Biodiversity Records*, *8*, e108.

De Jonge, D. S., Merten, V., Bayer, T., Puebla, O., Reusch, T. B., & Hoving, H. J. T. (2021).

A novel metabarcoding primer pair for environmental DNA analysis of Cephalopoda (Mollusca) targeting the nuclear 18S rRNA region. *Royal Society Open Science*, *8*(2), 201388.

Díaz-Abad, L., Bacco-Mannina, N., Madeira, F. M., Neiva, J., Aires, T., Serrao, E. A., ... & Frade, P. R. (2022). eDNA metabarcoding for diet analyses of green sea turtles (Chelonia mydas). *Marine Biology*, *169*(1), 18.

Eguiguren, A., Konrad Clarke, C. M., & Cantor, M. (2023). Sperm whale reproductive strategies: current knowledge and future directions. *Sex in Cetaceans: Morphology, Behavior, and the Evolution of Sexual Strategies*, 443-467.

Erickson, R. A., Merkes, C. M., & Mize, E. L. (2019). Sampling designs for landscape-level eDNA monitoring programs. *Integrated Environmental Assessment and Management*, *15*(5), 760-771.

Evans, K., & Hindell, M. A. (2004). The diet of sperm whales (Physeter macrocephalus) in southern Australian waters. *ICES Journal of Marine Science*, *61*(8), 1313-1329.

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., De Barba, M., ... & Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular ecology resources*, *15*(3), 543-556.

Foskolos, I., Koutouzi, N., Polychronidis, L., Alexiadou, P., & Frantzis, A. (2020). A taste for squid: the diet of sperm whales stranded in Greece, Eastern Mediterranean. *Deep Sea Research Part I: Oceanographic Research Papers*, *155*, 103164

Franklin, T. W., McKelvey, K. S., Golding, J. D., Mason, D. H., Dysthe, J. C., Pilgrim, K. L., ... & Schwartz, M. K. (2019). Using environmental DNA methods to improve winter surveys for rare carnivores: DNA from snow and improved noninvasive techniques. *Biological Conservation*, *229*, 50-58.

Geller, J., Meyer, C., Parker, M., & Hawk, H. (2013). Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Molecular ecology resources*, *13*(5), 851-861.

Gibbs, S. E. (2007). Retention and condition of cephalopod beaks in the stomach of an Australian sea lion (Neophoca cinerea). *Australian Mammalogy*, *29*(2), 241-244.

Gibert, J. P. (2019). Temperature directly and indirectly influences food web structure. *Scientific reports*, *9*(1), 5312.

Gissi, E., Manea, E., Mazaris, A. D., Fraschetti, S., Almpanidou, V., Bevilacqua, S., ... &

Katsanevakis, S. (2021). A review of the combined effects of climate change and other local human stressors on the marine environment. *Science of the Total Environment*, *755*, 142564.

Glover, K. A., Skaala, Ø., Limborg, M., Kvamme, C., & Torstensen, E. (2011). Microsatellite DNA reveals population genetic differentiation among sprat (Sprattus sprattus) sampled throughout the Northeast Atlantic, including Norwegian fjords. *ICES Journal of Marine Science*, *68*(10), 2145-2151.

Grey, E. K., Bernatchez, L., Cassey, P., Deiner, K., Deveney, M., Howland, K. L., ... & Lodge, D. M. (2018). Effects of sampling effort on biodiversity patterns estimated from environmental DNA metabarcoding surveys. *Scientific Reports*, *8*(1), 8843.

Kobayashi, H., Whitehead, H., & Amano, M. (2020). Long-term associations among male sperm whales (Physeter macrocephalus). *PLoS One*, *15*(12), e0244204.

Leis, J. M. (2006). Are larvae of demersal fishes plankton or nekton? *Advances in marine biology*, *51*, 57-141.

Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., ... & Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. *Frontiers in zoology*, *10*, 1-14.

Maio, N., Fioravanti, T., Latini, L., Petraccioli, A., Mezzasalma, M., Cozzi, B., ... & Caputo Barucchi, V. (2022). Life History Traits of Sperm Whales Physeter macrocephalus Linnaeus, 1758 Stranded along Italian Coasts (Cetartiodactyla: Physeteridae). *Animals*, *13*(1), 79.

Matthews, H. D., & Wynes, S. (2022). Current global efforts are insufficient to limit warming to 1.5 C. *Science*, *376*(6600), 1404-1409.

Matthews, J. K., Ridley, A., Kaplin, B. A., & Grueter, C. C. (2020). A comparison of fecal sampling and direct feeding observations for quantifying the diet of a frugivorous primate. *Current Zoology*, *66*(4), 333-343.

Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., ... & Iwasaki, W. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Society open science*, *2*(7), 150088.

Moors-Murphy, H. B. (2014). Submarine canyons as important habitat for cetaceans, with special reference to the Gully: a review. *Deep Sea Research Part II: Topical Studies in Oceanography*, *104*, 6-19.

Murray, D. C., Bunce, M., Cannell, B. L., Oliver, R., Houston, J., White, N. E., ... & Haile, J. (2011). DNA-based faecal dietary analysis: a comparison of qPCR and high throughput sequencing approaches. *PloS one*, *6*(10), e25776.

Nordeide, J. T., Johansen, S. D., Jørgensen, T. E., Karlsen, B. O., & Moum, T. (2011). Population connectivity among migratory and stationary cod Gadus morhua in the Northeast Atlantic—a review of 80 years of study. *Marine Ecology Progress Series*, *435*, 269-283.

Olsen, E. M., Melle, W., Kaartvedt, S., Holst, J. C., & Mork, K. A. (2007). Spatially structured interactions between a migratory pelagic predator, the Norwegian spring-spawning herring Clupea harengus L., and its zooplankton prey. *Journal of Fish Biology*, *70*(3), 799-815.

Oyarbide, U., Feyrer, L. J., & Gordon, J. (2023). Sperm and northern bottlenose whale interactions with deep-water trawlers in the western North Atlantic. *Plos one*, *18*(8), e0289626.

Pachur, M. E., & Horbowy, J. (2013). Food composition and prey selection of cod, Gadus morhua (Actinopterygii: Gadiformes: Gadidae), in the southern Baltic Sea. *Acta Ichthyologica et Piscatoria*, *43*(2), 109-118.

Pierce, G. J., Santos, M. B., Learmonth, J. A., Mente, E., & Stowasser, G. (2004). Methods for dietary studies on marine mammals. In *Investigating the roles of cetaceans in marine ecosystems. The Mediterranean Science Commission, CIESM Workshop Monographs* (Vol. 25, pp. 29-36).

Pearson, H. C., Savoca, M. S., Costa, D. P., Lomas, M. W., Molina, R., Pershing, A. J., ... & Roman, J. (2023). Whales in the carbon cycle: can recovery remove carbon dioxide?. *Trends in Ecology & Evolution*, *38*(3), 238-249.

Posdaljian, N., Soderstjerna, C., Jones, J. M., Solsona-Berga, A., Hildebrand, J. A., Westdal, K., ... & Baumann-Pickering, S. (2022). Changes in sea ice and range expansion of sperm whales in the eclipse sound region of Baffin Bay, Canada. *Global change biology*, *28*(12), 3860-3870.

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., ... & Laaksonen, A. (2022). The Arctic has warmed nearly four times faster than the globe since 1979. *Communications earth & environment*, *3*(1), 168.

Rishan, S. T., Kline, R. J., & Rahman, M. S. (2023). Applications of environmental DNA

(eDNA) to detect subterranean and aquatic invasive species: a critical review on the challenges and limitations of eDNA metabarcoding. *Environmental Advances*, 100370.

Roe, H. S. J. (1969). The food and feeding habits of the sperm whales (Physeter catodon L.) taken off the west coast of Iceland. *ICES Journal of Marine Science*, *33*(1), 93-102.

Roman, J., & McCarthy, J. J. (2010). The whale pump: marine mammals enhance primary productivity in a coastal basin. *PloS one*, *5*(10), e13255.

Roussel, J. M., Paillisson, J. M., Treguier, A., & Petit, E. (2015). The downside of eDNA as a survey tool in water bodies. *Journal of Applied Ecology*, 823-826.

Rødland, E. S., & Bjørge, A. (2015). Residency and abundance of sperm whales (Physeter macrocephalus) in the Bleik Canyon, Norway. *Marine Biology Research*, *11*(9), 974-982.

Santos, M. B., Pierce, G. J., Hartmann, M. G., Smeenk, C., Addink, M. J., Kuiken, T., ... & Mente, E. (2002). Additional notes on stomach contents of sperm whales Physeter macrocephalus stranded in the north-east Atlantic. *Journal of the Marine Biological Association of the United Kingdom*, *82*(3), 501-507.

Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., ... & Sherry, S. T. (2021). Database resources of the national center for biotechnology information. *Nucleic acids research*, *49*(D1), D10.

Schnell, I. B., Bohmann, K., Schultze, S. E., Richter, S. R., Murray, D. C., Sinding, M. H. S., ... & Gilbert, M. T. P. (2018). Debugging diversity–a pan-continental exploration of the potential of terrestrial blood-feeding leeches as a vertebrate monitoring tool. *Molecular ecology resources*, *18*(6), 1282-1298.

Serite, C. P., Emami-Khoyi, A., Ntshudisane, O. K., James, N. C., Jansen van Vuuren, B., Bodill, T., ... & Teske, P. R. (2023). eDNA metabarcoding vs metagenomics: an assessment of dietary competition in two estuarine pipefishes. *Frontiers in Marine Science*, *10*, 1116741.

Siegenthaler, A., Wangensteen, O. S., Soto, A. Z., Benvenuto, C., Corrigan, L., & Mariani, S. (2019). Metabarcoding of shrimp stomach content: Harnessing a natural sampler for fish biodiversity monitoring. *Molecular Ecology Resources*, *19*(1), 206-220.

Sigsgaard, E. E., Soraggi, S., Jensen, M. R., Repollés, A. G., Thomassen, E. E., & Thomsen, P. F. (2022). MetaBarFlow (Version 0.1.1) [Computer software].

Silas, E. G., Nammalwar, P., & Sarvesan, R. (1988). On the food of stranded sperm whale

Physeter macrocephalus Linnaeus at Tranquebar with a note on food habits of sperm whales, 65-71.

Similä, T., Haug, T., Lindblom, L., Lockyer, C., & O'Callaghan, S. A. (2022). Stomach Contents of Three Sperm Whales (Physeter macrocephalus) Stranded on Andøya, Northern Norway. *Aquatic Mammals*, *48*(5).

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA–An emerging tool in conservation for monitoring past and present biodiversity. *Biological conservation*, *183*, 4-18.

Towers, J. R., Tixier, P., Ross, K. A., Bennett, J., Arnould, J. P., Pitman, R. L., & Durban, J. W. (2019). Movements and dive behaviour of a toothfish-depredating killer and sperm whale. *ICES Journal of Marine Science*, *76*(1), 298-311.

Töbe, K., Meyer, B., & Fuentes, V. (2010). Detection of zooplankton items in the stomach and gut content of larval krill, Euphausia superba, using a molecular approach. *Polar Biology*, *33*, 407-414.

Trites, A. W., & Joy, R. (2005). Dietary analysis from fecal samples: how many scats are enough?. *Journal of Mammalogy*, *86*(4), 704-712.

Tykot, R. H. (2004). Stable isotopes and diet: you are what you eat. In *Physics methods in archaeometry* (pp. 433-444). IOS press.

Urban, P., Præbel, K., Bhat, S., Dierking, J., & Wangensteen, O. S. (2022). DNA metabarcoding reveals the importance of gelatinous zooplankton in the diet of Pandalus borealis, a keystone species in the Arctic. *Molecular Ecology*, *31*(5), 1562-1576.

van Weelden, C., Towers, J. R., & Bosker, T. (2021). Impacts of climate change on cetacean distribution, habitat and migration. *Climate Change Ecology*, *1*, 100009.

Vihtakari, M. (2024). *ggOceanMaps: Plot Data on Oceanographic Maps using 'ggplot2'*. R package version 2.2.0.

Wang, Y., Chen, K., Gao, J., Wang, M., Dong, J., Xie, Y., ... & Wang, B. (2021). Environmental DNA of preservative ethanol performed better than water samples in detecting macroinvertebrate diversity using metabarcoding. *Diversity and Distributions*, *27*(10), 1989-2002.

Wang, Y., Duan, R., & Zhang, J. (2017). Differentiating collagens based on mitochondrion 12SrRNA gene. *Food chemistry*, *234*, 139-143.

Wang, S., Yan, Z., Hänfling, B., Zheng, X., Wang, P., Fan, J., & Li, J. (2021). Methodology

of fish eDNA and its applications in ecology and environment. *Science of the Total Environment, 755*, 142622.

Wangensteen, O. S., Palacín, C., Guardiola, M., & Turon, X. (2018). DNA metabarcoding of littoral hard-bottom communities: high diversity and database gaps revealed by two molecular markers. *PeerJ, 6*, e4705.

Watwood, S. L., Miller, P. J., Johnson, M., Madsen, P. T., & Tyack, P. L. (2006). Deep-diving foraging behaviour of sperm whales (Physeter macrocephalus). *Journal of Animal Ecology, 75*(3), 814-825.

Weber, S., Junk, I., Brink, L., Wörner, M., Künzel, S., Veith, M., ... & Krehenwinkel, H. (2023). Molecular diet analysis in mussels and other metazoan filter feeders and an assessment of their utility as natural eDNA samplers. *Molecular Ecology Resources, 23*(2), 471-485.

Whitehead, H. (2018). Sperm whale: Physeter macrocephalus. In *Encyclopedia of marine mammals* (pp. 919-925). Academic Press.

Whitehead, H., & Shin, M. (2022). Current global population size, post-whaling trend and historical trajectory of sperm whales. *Scientific Reports*, *12*(1), 19468.

Wild, L. A., Mueter, F., Witteveen, B., & Straley, J. M. (2020). Exploring variability in the diet of depredating sperm whales in the Gulf of Alaska through stable isotope analysis. *Royal Society Open Science, 7*(3), 191110.

Xu, L., Huang, Q., Xu, S., Wang, X., Zhang, P., Xu, L., & Du, F. (2017). A new set of primers for COI amplification from purpleback flying squid (*Sthenoteuthis oualaniensis*). *Mitochondrial DNA Part B, 2*(2), 439-443.

Yoshida, T., Kawato, M., Fujiwara, Y., Nagano, Y., Tsuchida, S., & Yabuki, A. (2023). Optimization of environmental DNA analysis using pumped deep-sea water for the monitoring of fish biodiversity. *Frontiers in Marine Science, 9*, 965800.

York, R. (2017). Why petroleum did not save the whales. *Socius, 3*, 2378023117739217.

Young, J. K., Mast, A. M., Walton, J. A., Rodgers, T., Piaggio, A. J., Taylor, D. R., & Mock, K. E. (2023). Straight from the coyote's mouth: genetic identification of prey through oral swabs of predators. *Wildlife Biology, 2023*(6), e01155.

# Supplementary plots and materials

*Suppl. Table 1 - Fecal sample information overview with sampling notes.*

| Fecal samples in the UiT freezer - 16 samples (5 biopsies and 5 tags) | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Date** | **Sample ID** | **Latitude** | **Longitude** | **Biopsy** | **Tag** | **eDNA** | **Notes** |
| 05.06.2021 | PmaF21001 | 69.430483 | 15.229883 | No | No | No | No |
| 05.06.2021 | PmaF21002 | 69.392867 | 15.094333 | #5 | No | No | No |
| 05.06.2021 | PmaF21003 | 69.379602 | 15.046562 | No | No | No | No |
| 30.05.2022 | PmaF22001 | 69.612778 | 15.909444 | #2 | No | No | 17:42 - 69 29 486; 15 42 836 - BIOPSY #2 / POO SAMPLE / Away from FB |
| 27.03.2023 | PmaF23001 | 69.713056 | 16.403611 | No | #235820 | No | Taken with a plankton net - 13:25 Tag on (splash) Steinar TAG235820 distance 7m and 20 bars / Poo sample#1 |
| 28.03.2023 | PmaF23002 | 69.536389 | 15.872222 | #7 | No | No | Bleak Canyon - Biopsy#7 / Poo sample#2 |
| 29.03.2023 | PmaF23003 | 69.476389 | 15.888056 | No | No | No | Poo sample#3, really diluted |
| 29.03.2023 | PmaF23004 | 69.619444 | 15.869444 | #16 | No | #2 | Biopsy#16 / Poo sample#4 / eDNA#2 |
| 29.03.2023 | PmaF23005 | 69.635000 | 15.819722 | No | No | No | Missed biopsy 15:27 / Poo sample#5 |
| 29.06.2023 | PmaF23006 | 69.569444 | 15.877778 | Maybe #16 | No | No | Possibly sperm whale as biopsy #16 - Could have been same individual as biopsy#16, sight nbr 10 / Poo sample#6 |
| 10.06.2023 | PmaF23007 | 69.569167 | 15.491389 | #1 | #245465 | #1 | Behind Borehella <500m / Heading to Andhella still hauling / 12:54 SPLASH tag on / Biopsy#1 tube 1 / Poo sample#1 / eDNA#1 / Tag spef: Anker system 3.7.3, right side, angle 25<b0>, pressure 20 bars, petales-ankers not taped |
| 10.06.2023 | PmaF23008 | 69.460833 | 15.548889 | #2 | #235824 | #2 | <500m from Andhella hauling / 13:35 Tag on / Biopsy#2 tube 10 / Poo sample#2 / eDNA#2 |
| 13.08.2023 | PmaF23009 | 69.437530 | 15.422510 | No | CatCam | No | Cat cat Sperm whale. 69.44580; 15.42251. Possibly depredating or after depredation - 13:19 CAT CAM ON / Drone up / Poo Sample#1 / Done depredating |
| 14.08.2023 | PmaF23010 | 69.438820 | 15.228200 | No | CatCam | No | Besides fishing boat - <20m back of Triton / Drone up / 13:46 CAT CAM ON / Poo sample#2 / 13:43 1 SW up <200m from Triton, slow |
| 17.08.2023 | PmaF23011 | 69.564444 | 15.491944 | #5 | No | No | Very thin sample! - Blow sample#9 / Biopsy#5 tube 1 / Only picture of flank / Poo sample#3, quite light, maybe mostly plankton, not sure if any poo in it |
| 17.08.2023 | PmFa23012 | 69.427778 | 15.195278 | No | No | No | 16:21 Drone up / 16:22 failed biopsy (broke) + blow sample#11 + shallow diving / Poo sample#4 |

*Suppl. Table 2 - Surface water sample information overview, with sampling notes.*

| eDNA Surface Water Samples | | | | | | | |
|---|---|---|---|---|---|---|---|
| Date | Sample ID | Latitude | Longitude | Biopsy | Tag | Poo sample | Notes |
| 09.06.2021 | MOBY21008A-C | 69.470133 | 15.399350 | No | No | No | 3 replicates + 1 blank. 69.28.908 N, 15.23,961 E. No cetaceans seen this day. But whale ID 0274 (tag# 220658 was in area during night). |
| 22.04.2022 | MOBY22007A-C | 69.392700 | 16.308550 | No | No | No | No notes. |
| 28.03.2023 | MOBY23002A-C | 69.391388 | 15.849722 | No | No | No | No notes. |
| 29.03.2023 | MOBY23003A-C | 69.619444 | 15.869444 | #16 | No | #4 | No notes. |
| 10.06.2023 | MOBY23007A-C | 69.569167 | 15.491389 | #1 | 245 465 | #1 | No notes. |
| 10.06.2023 | MOBY23008A-C | 69.460833 | 15.548889 | #2 | 235 829 | #2 | No notes. |
| 13.08.2023 | MOBY23013A-B | 69.430000 | 15.250000 | No | No | #1 | Tag cruise. Only 2 filters (sample volume was too small). |

*Suppl. Table 3. Overview of blank reads from ethanol, extraction, air, and PCR blanks. Reads from both 12S and COI is shown.*

| Blank reads | | | | | | |
|---|---|---|---|---|---|---|
| Date | Sample ID | Blank type | COI before curation | COI after curation | 12S before curation | 12S after curation |
| 03.11.2023 | EtOH_blank_051123 | Ethanol | 12 | 0 | 0 | 0 |
| 05.11.2023 | EtOH_blank_031123 | Ethanol | 152 | 0 | 18 | 0 |
| 13.10.2023 | Blank_Ext_131023 | Extraction | 636 | 0 | 0 | 0 |
| 18.10.2023 | Blank_Ext_181023 | Extraction | 39 | 0 | 0 | 0 |
| 24.10.2023 | Blank_Ext_241023 | Extraction | 4091 | 0 | 0 | 0 |
| 27.10.2023 | Blank_Ext_271023 | Extraction | 1320 | 0 | 0 | 0 |
| 04.11.2023 | Blank_Ext_041123 | Extraction | 135 | 0 | 0 | 0 |
| 16.11.2023 | Blank_Ext_161123 | Extraction | 825 | 0 | 1 | 0 |
| 16.11.2023 | Blank_Ext2_161123 | Extraction | 242 | 0 | 0 | 0 |
| 26.04.2021 | MOBY21005Air | Air | 1540 | 0 | 157 | 0 |
| 02.05.2021 | MOBY21006Air | Air | 625 | 0 | 0 | 0 |
| 08.05.2021 | MOBY21007Air | Air | 137 | 0 | 0 | 0 |
| 09.06.2021 | MOBY21008Air | Air | 845 | 0 | 755 | 0 |
| 31.07.2021 | MOBY21009Air | Air | 514 | 0 | 1 | 0 |
| 11.08.2021 | MOBY21010Air | Air | 355 | 0 | 185 | 0 |
| 17.11.2023 | PCR_blank1 | PCR | 11 | 0 | 3 | 0 |
| 17.11.2023 | PCR_blank2 | PCR | 53 | 0 | 433 | 0 |
| 17.11.2023 | PCR_blank3 | PCR | 1 | 0 | 43 | 0 |
| 17.11.2023 | PCR_blank4 | PCR | 1 | 0 | 0 | 0 |

*Suppl. Table 4 - SIMPER analysis on SGC (SG1 – SG5) for the COI dataset including all kingdoms. Cumulative sum (cumsum) values are presented for the species contributing to 70.1% of the dissimilarity between surface water samples and fecal samples. Cumsum contribution values, indicating the amount of dissimilarity contributed by each species, are presented in percentage. Species contributing to dissimilarity from the fecal samples are highlighted in bold.*

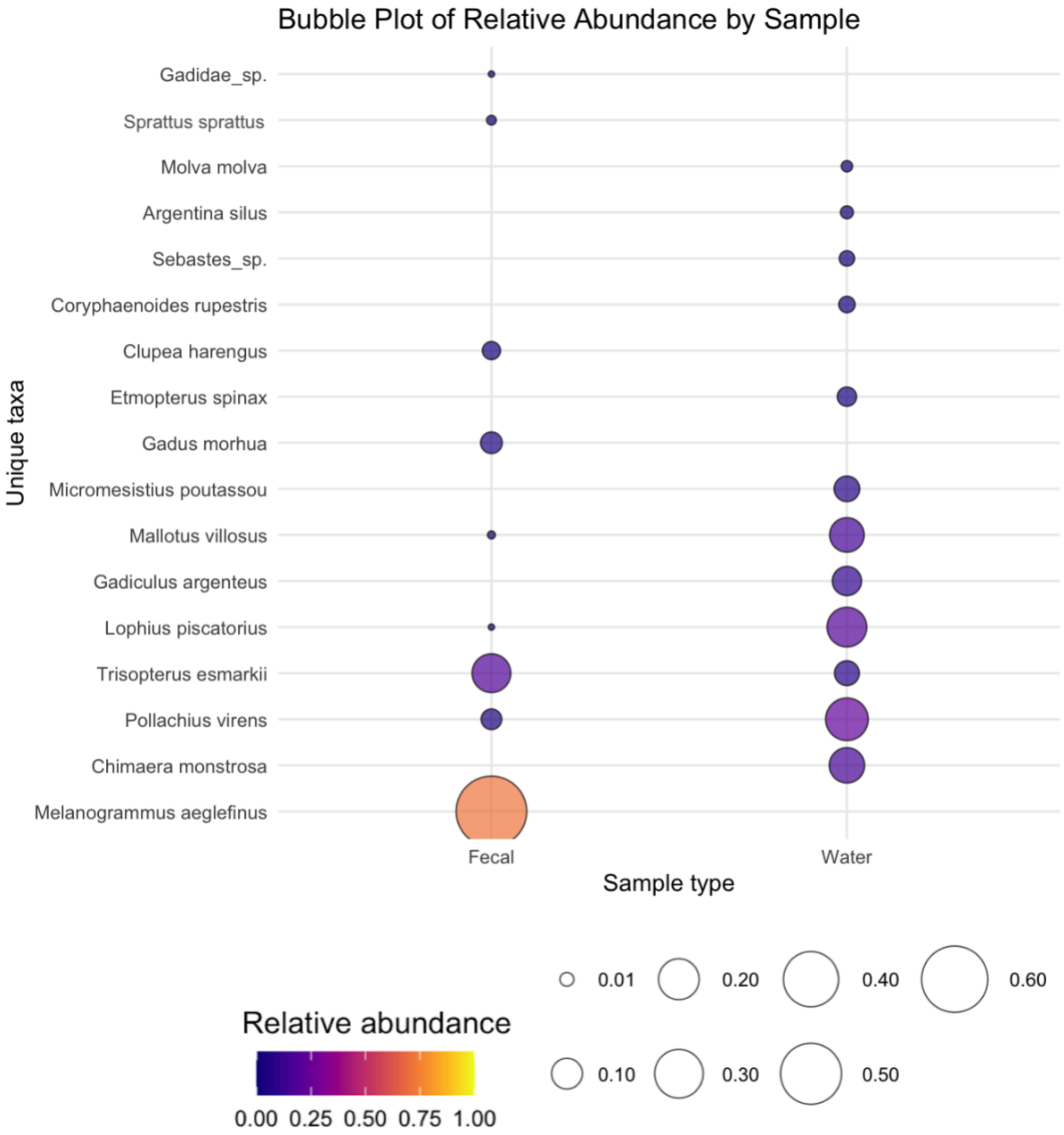
| **SIMPER analysis - COI all kingdoms** | | | |
|---|---|---|---|
| **Sequence** | **Unique taxa** | **Cumsum** | **Cumsum contribution** |
| seq1 | *Phaeocystis sp.* | 0.101 | 10.14% |
| seq6 | *Bathycoccus prasinos* | 0.190 | 8.84% |
| seq13 | ***Nanomia cara*** | 0.257 | 6.71% |
| seq179 | *Dinophyceae sp.* | 0.309 | 5.18% |
| seq15 | ***Physeter catodon*** | 0.359 | 5.01% |
| seq36 | *Heterocapsa rotundata* | 0.403 | 4.48% |
| seq8 | ***Calanus finmarchicus*** | 0.446 | 4.25% |
| seq77 | ***Leptothecata sp.*** | 0.482 | 3.64% |
| seq598 | *Pelagibacter sp.* | 0.507 | 2.42% |
| seq181 | *Pseudo-nitzschia sp.* | 0.528 | 2.19% |
| seq223 | *Pseudochattonella farcimen* | 0.549 | 2.10% |
| seq154 | *Oithona similis* | 0.569 | 1.97% |
| seq27 | ***Euphausiacea sp.*** | 0.587 | 1.80% |
| seq202 | *Polaribacter sp.* | 0.604 | 1.64% |
| seq84 | *Chloroparvula pacifica* | 0.620 | 1.64% |
| seq25 | ***Apolemia sp.*** | 0.636 | 1.56% |
| seq64 | ***Paracalanus parvus*** | 0.650 | 1.46% |
| seq193 | *Aureococcus anophagefferens* | 0.664 | 1.41% |
| seq2009 | *Kareniaceae sp.* | 0.678 | 1.36% |
| seq62 | ***Lizzia blondina*** | 0.690 | 1.21% |
| seq691 | *Teleaulax amphioxeia* | 0.701 | 1.12% |

*Suppl. Table 5 - SIMPER analysis on SGC (SG1 – SG5) for the COI-Metazoa dataset. Cumulative sum (cumsum) values are presented for the species contributing to 71.5% of the dissimilarity between surface water samples and fecal samples. Cumsum contribution values, indicating the amount of dissimilarity contributed by each species, are presented in percentage. Species contributing to dissimilarity from the fecal samples are highlighted in bold.*

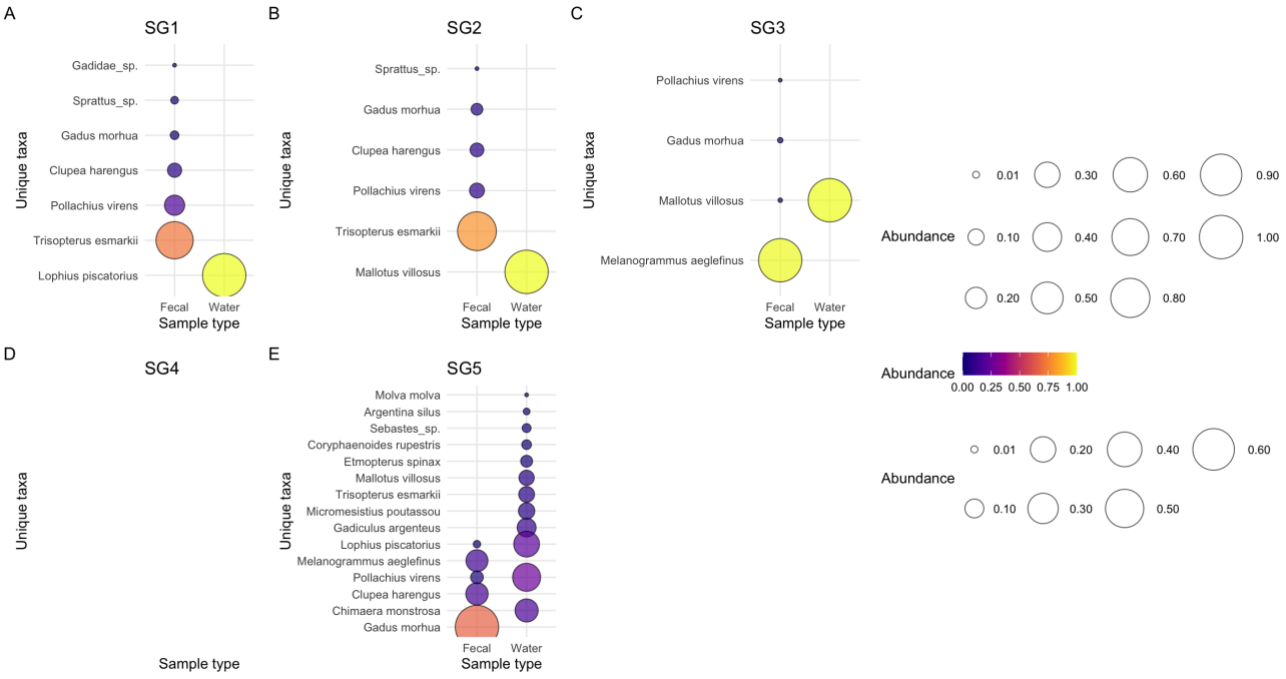
| **SIMPER analysis - COI-metazoa - SGC** | | | |
|---|---|---|---|
| **Sequence number** | **Unique taxa** | **Cumsum** | **Cumsum contribution** |
| seq13 | *Nanomia cara* | 0.133 | 13.31% |
| seq8 | ***Calanus finmarchicus*** | 0.215 | 8.20% |
| seq27 | ***Euphausiacea sp.*** | 0.288 | 7.24% |
| seq77 | ***Leptothecata sp.*** | 0.352 | 6.48% |
| seq154 | *Oithona similis* | 0.413 | 6.06% |
| seq91 | ***Arachnanthus sarsi*** | 0.465 | 5.24% |
| seq7010 | *Bolinopsis sp.* | 0.505 | 3.96% |
| seq25 | ***Apolemia sp.*** | 0.541 | 3.58% |
| seq6752 | *Mallotus villosus* | 0.576 | 3.54% |
| seq4312 | *Lophius piscatorius* | 0.607 | 3.09% |
| seq5183 | *Metazoa sp.* | 0.638 | 3.08% |
| seq9130 | *Globicephala sp.* | 0.668 | 3.06% |
| seq64 | ***Paracalanus parvus*** | 0.692 | 2.31% |
| seq3067 | *Semibalanus balanoides* | 0.715 | 2.31% |

*Suppl. Table 6 - SIMPER analysis on SGC for the COI-fish dataset. Overview of all SIMPER data.*

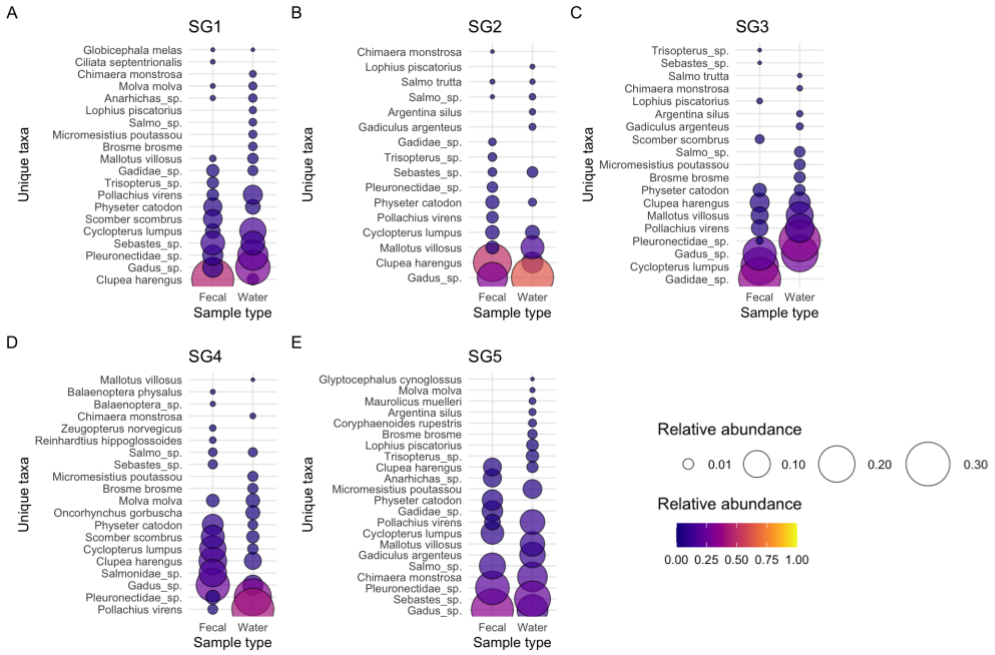
| SIMPER analysis - all data - SGC | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Water _Fecal average** | **Water _Fecal sd** | **Water _Fecal ratio** | **Water _Fecal ava** | **Water _Fecal avb** | **Water _Fecal cumsum** | **Water _Fecal p** |
| **Seq 4312** | 0.178 | 0.188 | 0.945 | 0.485 | 0.012 | 0.194 | 0.041 |
| **Seq 6752** | 0.168 | 0.197 | 0.851 | 0.415 | 0.111 | 0.378 | 0.114 |
| **Seq 1051** | 0.113 | 0.129 | 0.876 | 0.167 | 0.299 | 0.501 | 0.813 |
| **Seq 10799** | 0.112 | 0.136 | 0.819 | 0.000 | 0.302 | 0.623 | 0.549 |
| **Seq 1955** | 0.084 | 0.135 | 0.619 | 0.086 | 0.213 | 0.715 | 0.711 |
| **Seq 4536** | 0.071 | 0.111 | 0.643 | 0.000 | 0.216 | 0.792 | 0.817 |
| **Seq 771** | 0.057 | 0.134 | 0.424 | 0.000 | 0.153 | 0.855 | 0.794 |
| **Seq 5252** | 0.030 | 0.044 | 0.692 | 0.134 | 0.000 | 0.888 | 0.236 |
| **Seq 6817** | 0.024 | 0.035 | 0.692 | 0.106 | 0.000 | 0.914 | 0.236 |
| **Seq 5112** | 0.020 | 0.029 | 0.692 | 0.089 | 0.000 | 0.936 | 0.236 |
| **Seq 4919** | 0.014 | 0.020 | 0.692 | 0.060 | 0.000 | 0.951 | 0.236 |
| **Seq 10568** | 0.011 | 0.016 | 0.692 | 0.047 | 0.000 | 0.963 | 0.236 |
| **Seq 20051** | 0.010 | 0.014 | 0.692 | 0.043 | 0.000 | 0.974 | 0.236 |
| **Seq 16855** | 0.009 | 0.017 | 0.498 | 0.000 | 0.025 | 0.983 | 0.804 |
| **Seq 9103** | 0.007 | 0.010 | 0.692 | 0.032 | 0.000 | 0.991 | 0.236 |
| **Seq 25645** | 0.006 | 0.008 | 0.692 | 0.025 | 0.000 | 0.997 | 0.236 |
| **Seq 29998** | 0.003 | 0.008 | 0.336 | 0.000 | 0.008 | 1.000 | 0.744 |




*Suppl. Figure 1 Bubble plot of COI fish data for SGC, with bubble size and color indicating the relative abundance within each sample type.*




*Suppl. Figure 2 - Bubble plot of COI fish data for SG1–SG5 (SG4 has no fish taxa), with bubble size and color indicating the relative abundance within each sample type.*




*Suppl. Figure 3 - Bubble plot of 12S data for SG1–SG5, with bubble size and color indicating the relative abundance within each sample type.*

# Appendix A - Fecal sampling protocol

Ove Mikal Pedersen

## Fecal Sampling Protocol

**Info**

If the whale releases feces at the beginning of the dive, it might be possible to retrieve a sample. The fecal matter will quickly/immediately dissolve into a cloud. The longer this cloud remains concentrated and undisturbed by fluke movement (or boat propeller) the higher the chance of a good sample.

**Materials**

- A plankton net mounted on a 2,5 m long pole (approximately) with a small open/close valve (approximately 150 cl)
- Dive mask.
- Snorkel.
- Fins.
- Thermal protection.
- Spray bottle.
- Ethanol.
- A jar per sample.

**Fecal sampling**

1. When feces are visible at the surface, the center of the concentration is kept in sight while the boat approaches.
2. The diver enters the water with the net while trying to avoid further disturbing the concentration.
3. The net is repeatedly pushed through the concentration of the cloud while kept in a continuous movement to prevent pulses of water from expelling particles from the net.
4. Once the sample is achieved the net is kept straight up in a horizontal position over the surface, allowing water and fecal particles to end up in the closed valve at the end of the plankton net.
5. The net is retrieved on the boat.
6. On site processing of the sample is then done by repeatedly spraying the net with ethanol, allowing any remnants in the net to move down into the valve.
7. The valve is then opened over a sample jar and the rinsing of the net with ethanol continues until all signs of visible traces in the net and valve are gone.
8. Ethanol is then added to the sample and the jar is left to settle for further processing.

# Appendix B - Surface water sampling protocol

Kim Præbel

# Surface Water Sampling Protocol

**Short sampling instruction**

For the "transect" sampling and sampling from the vessel in the whale segregation (currents do not matter as we do not know where the whales have been within the last 0-48 hours).

**Materials**

- Nitril gloves
- 10-20% Bleach.
- 50% ethanol.
- Milli-Q.
- Niskin water sampler.
- Sterivex filters.
- Three falcon tubes per sample site.
- 2l beaker.
- Three 50ml syringes per sample site.
- One ziplock bag per sample.
- Cooling bag if no freezer.

**Instructions**

1. Put on gloves and spray them with 10-20% bleach solution, Milli-Q, then 50% EtOH. Let the gloves dry.
2. Start by spraying the Niskin thoroughly on the in and outside with a 10-20% bleach solution, let it work for 5-10 minutes.
3. Make sure that the white outlet vent on the side is closed before lowering it into the water.
4. When arriving at the sample site, mark three 50ml centrifuge tubes with location, station nr., date, and A, B, C for the three replicates. Apply clear tape over your writing to preserve the label.
5. Find two Sterivex filters and two syringes.
6. Then bleach the 2l beaker in a sink.
7. When at the sampling site, flush the Niskin in the surface water to remove residual bleach.
8. Lower the Niskin to 1-2m below surface and drop the lead to close the Niskin sampler.
9. Retrieve the Niskin.
10. Poor out the bleach from the beaker and flush it three times with a small amount of the collected seawater to make sure the bleach is gone.
11. Then fill the beaker with water and fill the syringe.
12. Place the filter on the syringe and press the water through it. Make sure that the inlet/outlet of the filter do not touch any surface.
13. Repeat 10 times (500 ml).
14. Transfer the filter to the labeled sterile centrifuge tube, then put it into a ziplock bag (labeled with data, cruise, name) and transfer it to the freezer.

# Appendix C - Preparing fecal samples for extraction

Rasmus Buhl Søiland

## Fecal Sample Protocol- Preparing for Extraction

**Description:**

All the laboratory work takes place in B-310.

Two fractions are obtained through this protocol:

1. The first fraction you will get is the ethanol which the fecal samples have been stored in.
2. The second fraction is the fecal matter.

After these preparation steps, both fractions are further treated using the "Qiaqen PowerSoil Pro kit". For instructions see the "Qiaqen PowerSoil Pro" protocol.

This provides two different sampling methods from one fecal sample, which makes comparison of methods possible.

This whole process takes at least three days. If short on time on day two, it is also possible to stop the work after step 10 and proceed with the homogenization another day. However, the fecal jars must be transferred back to the freezer if not proceeding with homogenization on day two.

**Materials:**

For all parts:

- 10% bleach.
- Milli-Q.
- 70% ethanol.
- 96% ethanol.
- One pack of nitril gloves.
- One roll of paper towel.

For ethanol filtering:

- One 50ml syringe per sample.
- One 0,22µm Sterivex filter per sample + one for ethanol blank.
- One forceps.
- Zip-lock bags.

- Empty jar.

For homogenization:

- Spray bottle.
- Centrifuge for falcon tubes.
- One cell spreader for each sample.
- Aluminum foil.
- One sterile 16 G needle for each sample.
- Three mini choppers (POINT POCH7037 Mini chopper).
    - One blender can suffice, but more blenders will speed up the process.
- Two 50ml falcon tubes for each sample.
- One funnel.

For opening of Sterivex filters:

- PVC pipe cutter
- One clean scalpel for each ethanol sample.
- One forceps.
- One petri dish.

**Cleanness**

It is important that the workspace is clean to avoid contamination. Therefore, the laminar flow-hood must be disinfected with UV-light for at least 30 minutes before starting the work. Afterwards it must be cleaned thoroughly with 10% bleach, Milli Q and 70% ethanol, in that order. If something spills on the table, it must be cleaned again. When done using the flow-hood, it must be cleaned with bleach, Milli-Q, and ethanol, followed by UV-light for at least 30 minutes.

When using bleach, allow the bleach to work for at least two minutes before removing it with milli-Q

Rasmus Buhl Søiland

Gloves must be worn throughout all steps. Gloves must be changes if anything, that could contaminate (e.g., liquid from the samples) comes into contact with the gloves, or anything outside the flow-hood is touched.

Never use a tool that has been used for one sample on another, without cleaning it thoroughly first. It is preferred to use disposable tools, which makes it possible to use a new one for each sample.

**Ethanol filtering**

Day one:

1. Turn on the UV-light in the laminar flow-hood, and let it be on for 30 minutes.
2. Clean everything in the laminar flow-hood with 10% bleach, milli-Q and 70% ethanol.
3. Clean the fecal sample jars with 10% bleach, milli-Q and 70% ethanol.
    a. Make sure that the labels on the outside of the jar are covered with clear tape, or a label is inside the jar, otherwise the ethanol will remove the writing.
4. Transfer the fecal sample jars to the flow-hood and let them stand untouched for at least 24 hours to allow the particles to precipitate. Remember to be careful when moving the samples around, to avoid swirling up the particles, especially after the 24 hours of precipitation.

Day two:

5. Clean everything in the laminar flow-hood with 10% bleach, milli-Q and 70% ethanol.
6. Important! For step 7 and 8 note how many ml is filtered through the Sterivex filter for each sample (otherwise it is impossible to quantify later), and have an empty jar where the ethanol coming out of the Sterivex filters can go into.
7. An ethanol blank must be made with clean ethanol through a 0,22µm Sterivex filter.
8. Remove the lid from the sample jar and suck up the ethanol with a syringe down to about 2 centimeters above the fecal matter (be careful to avoid sucking in the fecal matter), and filter it through a 0,22µm Sterivex filter right away.

9. Place the Sterivex filters into marked (date and sample ID) zip-lock bags, with one sample per bag. The samples must be double bagged.
10. Transfer the bags with Sterivex filters to a - 20°C freezer.

**Homogenization of fecal matter**

11. Clean everything in the laminar flow-hood with 10% bleach, milli-Q and 70% ethanol.
12. Transfer the fecal matter from each individual jar into a separate 50ml falcon tube (marked with sample ID) for each sample. A spray bottle with 96% ethanol is used to flush any remaining fecal matter down, to get all the fecal matter into the falcon tube.
13. The falcon tubes are centrifuged at 4000xg for 4 minutes, to create a pellet at the bottom of the falcon tube.
14. Afterwards, there should be a clear line between ethanol and particular matter. If there is not, it must be centrifuged again.
15. Remove some of the excess ethanol by carefully pouring it out, there must be about 4 cm of ethanol left above the fecal matter.
16. Transfer the fecal matter from one sample to the mini chopper. A spray bottle with 96% ethanol is used to flush any remaining fecal matter down, to get all the fecal matter into the mini chopper.
17. The sample is now blended for 1 minute to homogenize the fecal matter. The blender must be cleaned thoroughly with 10% bleach, milli Q and 70% ethanol between each sample.
18. The homogenized fecal matter for each individual sample is then poured into a new falcon tube (use a funnel to avoid spilling), marked with sample ID. To make sure all the fecal matter gets into the tube, the mini chopper is flushed with a spray bottle containing 96% ethanol, and afterwards the funnel is flushed as well.
19. The falcon tubes are centrifuged at 4000xg for 4 minutes, to create a pellet at the bottom of the falcon tube.
20. Gently pour out as much ethanol as possible and be careful that no fecal matter leaves with the ethanol.
21. Use cell spreaders (a new for each sample) to mix the fecal matter.

22. Aluminum foil is tightened around the top of each falcon tube.
23. Ten small holes are poked in the aluminum foil with a sterile 16 G needle (a new one for each sample).
24. Leave the samples for 20 hours, to allow the ethanol to evaporate.

Day three:

25. Clean everything in the laminar flow-hood with 10% bleach, milli-Q and 70% ethanol.
26. Check if the fecal matter is dry. If not, leave them for a longer time, or if you are in a hurry put them in a 56°C oven until they are dry (if using the oven, clean it with 10% bleach, Milli-Q, and 70% ethanol before use).
27. Put the falcon tubes into zip lock bags marked with sample ID, with one sample per bag. They must be double bagged to avoid contamination.
28. Transfer the samples to the freezer (-20°C).

**Opening of ethanol Sterivex filters**

Before proceeding with Powersoil Pro Kit, the filters must be opened, and cut into pieces.

29. Clean the laminar flow-hood with 10% bleach, milli-Q, and ethanol.
30. Take the ethanol Sterivex filters out of the freezer.

The rest of the steps must be done one sample at a time, to avoid contamination.

31. For visualization of step 28-32 see Figure 1.
32. Cut the filter Sterivex filter open with a pipe cutter at the outlet-end and separate the filter from the casing.
33. With a clean scalpel, make a vertical and a horizontal cut in the filter.
34. Remove the filter for the inner casing with a clean forceps.
35. Place the filter into a clean petri dish and cut it into small pieces.
36. Insert the filter pieces into a microcentrifuge tube provided by the PowerSoil Pro Kit, and mark it with sample ID.

Now both sample types are ready to proceed with the Qiaqen PowerSoil Pro Kit protocol.

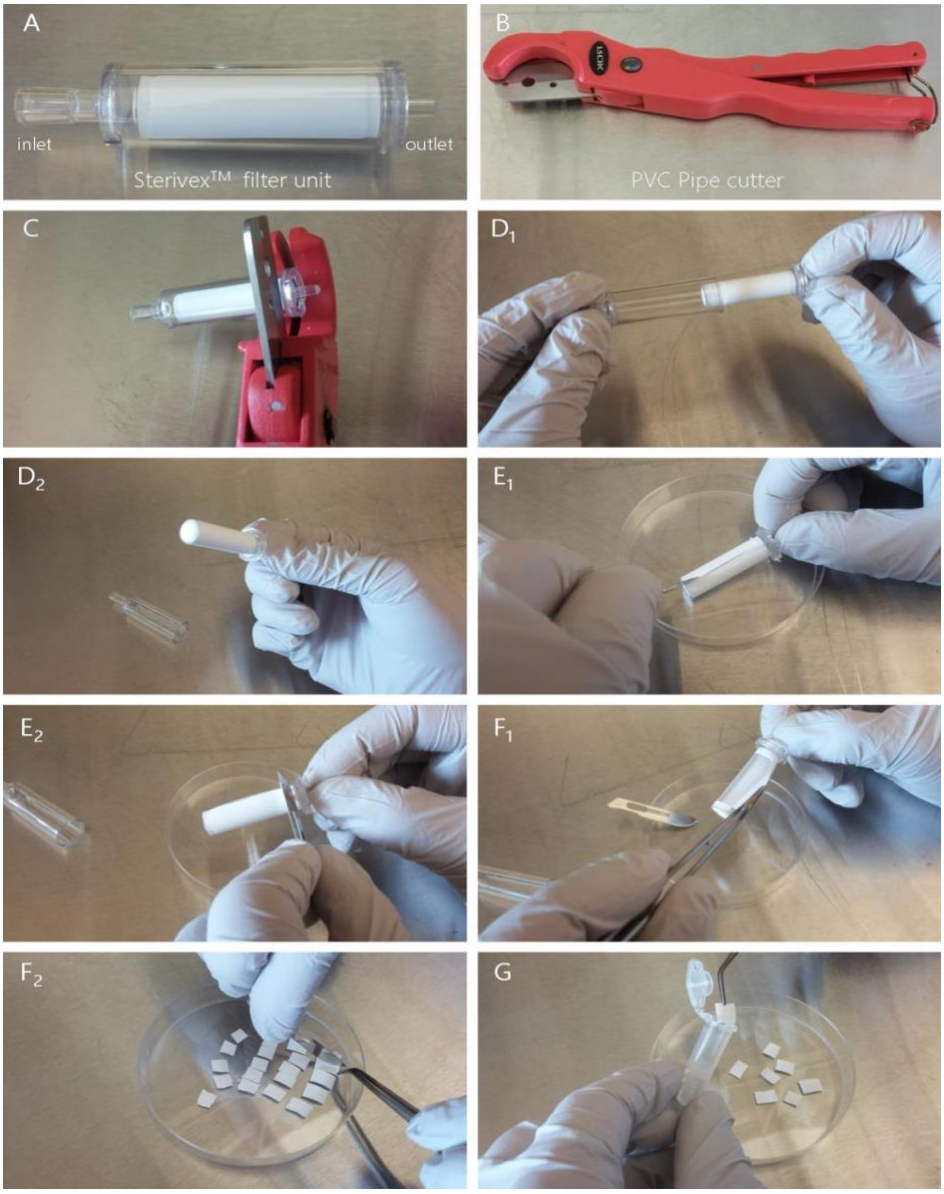



*Figure 1. Demonstration of filter separation from case (Cruaud et al., 2017). (A) Sterivex filter. (B) PVC pipe cutter. (C) Opening the Sterivex filter with the pipe cutter at the outlet-end. ($D_1$, $D_2$) Removing the case from the filter. ($E_1$, $E_2$) Separating the filter from the inner casing with a scalpel with a vertical cut (down the filter), followed by a horizontal cut (round the filter). ($F_1$) Removing the filter from inner casing with a forceps and placing it into a petri dish. ($F_2$) Cutting the filter into small pieces with a clean scalpel. (G) Putting the filter pieces into the microcentrifuge tube provided by the PowerSoil Pro Kit.*

Cruaud, P., Vigneron, A., Fradette, M.-S., Charette, S.J., Rodriguez, M.J., Dorea, C.C. and Culley, A.I. (2017), Open the Sterivex™ casing: An easy and effective way to improve DNA extraction yields. Limnol. Oceanogr. Methods, 15: 1015-1020. https://doi.org/10.1002/lom3.10221

# Appendix D - PowerSoil Pro protocol

## Protocol: Experienced User

### Important notes before starting

- Ensure that the PowerBead Pro Tubes rotate freely in the centrifuge without rubbing.
- If Solution CD3 has precipitated, heat at 60°C until precipitate dissolves.
- Perform all centrifugation steps at room temperature (15–25°C).

### Procedure

1. Spin the PowerBead Pro Tube briefly to ensure that the beads have settled at the bottom. Add up to 250 mg of soil and 800 µl of Solution CD1. Vortex briefly to mix.
2. Secure the PowerBead Pro Tube horizontally on a Vortex Adapter for 1.5–2 ml tubes (cat. no. 13000-V1-24). Vortex at maximum speed for 10 min.

   **Note**: If using the Vortex Adapter for more than 12 preps simultaneously, increase the vortexing time by 5–10 min.

   **Note**: For alternative ways to homogenize samples, see the detailed protocol on page 13–14.
3. Centrifuge the PowerBead Pro Tube at 15,000 x *g* for 1 min.
4. Transfer the supernatant to a clean 2 ml Microcentrifuge Tube (provided).

   **Note**: Expect 500–600 µl. The supernatant may still contain some soil particles.
5. Add 200 µl of Solution CD2 and vortex for 5 s.
6. Centrifuge at 15,000 x *g* for 1 min. Avoiding the pellet, transfer up to 700 µl of supernatant to a clean 2 ml Microcentrifuge Tube (provided).

   **Note**: Expect 500–600 µl.
7. Add 600 µl of Solution CD3 and vortex for 5 s.
8. Load 650 µl of lysate to an MB Spin Column. Centrifuge at 15,000 x *g* for 1 min.

9. Discard the flow-through and repeat step 8 to ensure that all of the lysate has passed through the MB Spin Column.
10. Carefully place the MB Spin Column into a clean 2 ml Collection Tube (provided). Avoid splashing any flow-through onto the MB Spin Column.
11. Add 500 µl of Solution EA to the MB Spin Column. Centrifuge at 15,000 x *g* for 1 min.
12. Discard the flow-through and place the MB Spin Column back into the same 2 ml Collection Tube.
13. Add 500 µl of Solution C5 to the MB Spin Column. Centrifuge at 15,000 x *g* for 1 min.
14. Discard the flow-through and place the MB Spin Column into a new 2 ml Collection Tube (provided).
15. Centrifuge at up to 16,000 x *g* for 2 min. Carefully place the MB Spin Column into a new 1.5 ml Elution Tube (provided).
16. Add 50–100 µl of Solution C6 to the center of the white filter membrane.
17. Centrifuge at 15,000 x *g* for 1 min. Discard the MB Spin Column. The DNA is now ready for downstream applications.

    **Note**: We recommend storing the DNA frozen (–30 to –15°C or –90 to –65°C) as Solution C6 does not contain EDTA. To concentrate DNA, refer to the Troubleshooting Guide.

# Appendix E - Clean lab routines – C358

**Norwegian College for Fishery Science**

**Genetics Group**

Last updated: April 2021

## CLEAN LAB ROUTINES – C358

Firstly, before you enter the lab, do you have everything you need to perform the lab task at hand? We do not want you to leave in the middle of the process. Remember to visit the restroom as well.

### GENERAL CLEAN LAB ROUTINES

1. Take your shoes off in the corridor and bring with you the plastic box (if there is any stuff in it) to bring into the labs. Enter the lock.
2. Leave personal belonging in the small box (i.e., phone, access card, rings etc.) Wash your hands, put on clean suit and blue shoe covers, and finally a pair of gloves.
3. Make fresh 10% bleach in the bucket, and more 50% EtOH solution in the spray bottles if needed. Take a new dish cloth.
4. Clean what was in the plastic box from the corridor and place them where it belongs.
5. If you have samples to bring into a clean lab:
    a. Remove the outer bag and clean the next bag with bleach before entering room. If your samples are not double-bagged (which means only one bag around you triplicates samples), then you clean the bag very thoroughly with bleach, but not ethanol since it will wipe off the label.
    b. Enter the room with only the bags of samples. Leave the sample box in the lock. Put samples in another box inside the extraction room box and put in the fridge.
    c. Exit back to the lock. Clean box used for samples and also the box from the corridor.
    d. Change gloves. Re-enter the extraction room with bleach, ethanol and lab equipment you may wish to bring in.
6. When inside the extraction room put on lab shoes. Put your lab equipment in assigned box with project name. Keep all your things in this box always and keep it clean.
7. Clean you gloves and put a second pair on.
8. Clean the flow hood and all equipment that are to be used for the protocol, i.e., vortex, centrifuges, pipettes, racks on so forth. Clean the orange 50ml tube adapters from the centrifuge and place inside the flowhood. Always clean everything.
9. Place everything in the flow hood and UV-treat it for 10 mins. At this step you can even place the tubes, tips, bags, markers etc. needed for the lab protocol in the hood to be UV'ed too. NEVER UV the chemicals from the kits.
10. While waiting for the UV to finish, first clean the heating cabinet and switch it on, the proceed to clean the large falcon centrifuge, sample wheel (rotator), chair and any other surfaces. Put bags in the two trash cans – one for regular waste and one for lab waste.
11. Change the second pair of gloves and clean them.
12. Proceed to the appropriate lab protocol.

When extractions are done, your samples should be stored in the stock and aliquot freezers respectively. Remember, do not freeze your aliquots if they are to be used within 4 weeks for downstream protocol (i.e., lib.prep., PCR etc.). Store in fridge if that is the case. Store your aliquots in PCR plates or strips. Put two bags around and make sure to label it properly. The rest of your extracted DNA (stock) is stored in 1.5 ml tubes in a cryobox that you have purchased at the store and brought with you. Make sure to label the box properly. Put the cryobox in two bags before storing it in the stock freezer.

# Appendix F - Extraction Protocol for Sterivex Filters

UiT
THE ARCTIC
UNIVERSITY
OF NORWAY

Norwegian College for Fishery Science
Research Group for Genetics, K. Præbel
Last updated: November 2019, edit. J. Bitz

## EXTRACTION PROTOCOL FOR STERIVEX FILTERS

DAY 1:

1. Follow the descriptions in the 'Clean Lab Routines' of how to enter the labs.
2. Clean the outside of the 50ml falcon tubes containing the filters with bleach. Alternatively, if your filters are in ziplock bags, clean the outside of the bag. Do not use ethanol if there is any labelling on the tube/bag.
3. To remove excess water inside the filters, place the inlet of the filter (narrow end) in a 1.5 ml Eppendorf tube and gently slide filter and tube into the 50 ml falcon tube that contained the filter (or in a new 50 ml tube if samples were stored in ziplock bags). If more than one filter is in the tube, label a new tube for the second filter. When done with filters from one species/station, clean everything again (forceps, gloves, working surface) with bleach, MilliQ water and ethanol, before proceeding to the next **species/station.**
4. **Centrifuge the tubes at 1500 x g for 3 minutes to remove the remaining seawater from the filters.**
5. **Make extraction buffer solution for adding 2.5X the recommended volume = 500µl per filter.**
   - **Recommended volume is 20µl Proteinase K + 180µl Buffer ATL per sample:**
     - **2.5 * 20µl ProK = 50µl**
     - **2.5 * 180µl ATL = 450 ul**
     - **Total amount of extraction buffer per sample = 500µl**
     - **E.g. for 20 samples: 1000µl ProK, 9000µl ATL. First, pipette the 9ml with a sterile glass pipette into a clean 50ml falcon tube. Then pipette 1 ml of ProK into the same tube. Close with lid and invert solution, avoiding foaming.**
6. **Add 500µl of the extraction solution to each filter, starting with blanks, by pushing the 1000µl tip tight into the outlet end of the filter and gently aspirating the solution into the filter. Take care that all the solution goes into the filter. If the filter is clogged, then aspirate from the inlet end of the filter.**
7. **Cap the filters with sterile caps. Make sure that its completely sealed.**
8. **MAKE SURE YOU LABEL ALL THE FILTERS CORRESPONDING TO THE TUBES, by writing the label and the replicate letter (A, B, C etc.) on the filter and cover with tape.**
9. **Place the filters in rotator and fasten them with the elastic band.**
10. **When done with all filters, move the rotator to the incubator oven (56°C). Make sure that the rotator is moving at 6 rpm and not hitting the oven. Check the filters after a couple of hours and leave them overnight for the 2nd day of extractions. Minimum 8-12 hours incubation.**
11. **Note: Always use similar incubation time for all filters within a project. Note the time for when incubation in the incubator oven started.**

DAY 2

12. **Enter lab and clean according to the Clean Lab Routines.**
13. **Label all tubes needed for the process: 2ml Eppendorf tubes, spin columns and the final 1.5ml Eppendorf tubes that will hold the eluted DNA (sample ID on top, and more details on the side including replication (A,B,C), depth, date of collection, date of extraction and your initials).**
14. **Note the time when the filters are removed from the incubator oven.**
15. **Reopen the sealed filters and transfer them to a marked 2ml tube inside a new 50ml falcon tube with the inlet facing down into the 2ml Eppendorf tube.**
16. **Centrifuge the 50ml tubes containing the 2ml tubes and the filters at 1700 x g for 3 minutes.**
17. **Remove the filter from the 50ml tube and discard it. Then carefully remove the 2ml tube from the bottom of 50ml tubes with a tweezer holding the root of the cap, without touching the cap itself or the edge of**

the tube opening. Close the 2ml tube and place it in a rack. Again, start with the lowest concentration (e.g. air-> blank -> real samples).

18. "Measure" the approximate volume of 2-3 samples using a pipette with NEW tips for each sample. Round the mean volume to nearest 50µl.
19. Add an equal volume of the Buffer AL as the one determined above (7.) and ensure to mix it with the pipette immediately using new tips for each sample.
20. Add an equal volume of 100% EtOH as the one determined above (7.) and ensure to mix it with the pipette immediately using new tips for each sample.
21. Vortex and spin down the samples to make sure it is mixed and liquid from the cap is removed.
22. Place the spin columns in front of the samples in the rack.
23. Transfer 630µl of the sample into corresponding spin column. Be careful not to make any bubbles but at the same time try not to leave liquid in the tip because it is precious DNA.
24. Centrifuge the columns at 15.000 x g for 2 mins.
25. Discard the collection tube with the flow-through and transfer the spin column to a new collection tube. Make sure the flow-through has not spilled back to the column when you removed it from the centrifuge.
26. Transfer the rest of the sample to the corresponding spin column. If more than 630µl, three rounds of spinning are required.
27. Centrifuge the columns at 15.000 x g for 2 mins.
28. Discard the collection tube with the flow-through and transfer the spin column to a new collection tube. Make sure the flow-through has not spilled back to the column when you removed it from the centrifuge.
29. Add 500µl Buffer AW1 (check EtOH has been added to buffer) using new tips for each tube.
30. Centrifuge at 15000 x g for 2 mins.
31. Discard the collection tube with the flow-through and transfer the spin column to a new collection tube. Make sure the flow-through has not spilled back to the column when you removed it from the centrifuge.
32. Add 500µl Buffer AW2 and centrifuge for 4 mins at 20.000 x g.
33. While centrifuging, clean flowhood, pipettes, and pens with bleach, MilliQ and ethanol.
34. TAKE GREAT CARE that no flow-through is present on the sides of the spin columns. If so, spin the columns again in a new collection tube at 20.000 x g for 2 mins. Note what samples that have been centrifuged twice.
35. Transfer the spin-columns to the corresponding Eppendorf tubes. Make sure that the lid/tap of the spin column does not touch the cap of the Eppendorf tube to avoid contamination.
36. Add 75µl of Buffer AE to each spin columns. Make sure to add the buffer at the center of the membrane without touching the membrane. Incubate for 1 min, then spin the samples at 20.000 x g for 2 mins.
37. Discard the spin columns and transfer a 20µl aliquot of the extracted DNA from each sample to a PCR plate or PCR strips. It is very important the plate/strip is labeled properly with all necessary information (if using strips, use empty pipette tip boxes as racks). Wrap aliquots in two bags before temporary storage. Place the aliquot in the fridge at 4°C if you are certain it will be processed within the next 2-3 weeks or in the aliquot freezer if longer.
38. Store the rest of the DNA as stock in the freezer located in the extraction lab. Store the 1.5ml tubes in a cryobox that you have purchased at the store and brought with you. Make sure to label the box properly. Put the cryobox in two bags before storing it and **ONLY** thaw the stock if absolutely necessary.
39. Clean flowhood and all equipment according to the guidelines.

# Appendix G - Library preparation protocol

## Procedure

1. Program a thermocycler according to Table 2 using the predetermined FX fragmentation time for step 2. Be sure to use the instrument's heated lid, and if possible, set the temperature of the heated lid to 70°C.

**Table 2. Input DNA (20 pg –1000 ng) free of EDTA, Buffer EB, or in 0.1x TE**

| Step | Incubation temperature | Incubation time |
|---|---|---|
| 1 | 4°C | 1 min |
| 2 | 32°C | 1–30 min* |
| 3 | 65°C | 30 min |
| 4 | 4°C | Hold |

* To determine the reaction time for step 2, refer to Figure 1 and Table 1.

2. Start the program. When the thermocycler block reaches 4°C, pause the program.
3. Prepare the FX reaction mix in a PCR plate or tube on ice according to Table 3 for >10 ng input DNA or Table 4 for <10 ng input DNA. Mix well by gently pipetting (do not vortex to mix).

**Table 3. FX reaction mix setup (per sample) for >10 ng input DNA**

| Component | Volume (µL) |
|---|---|
| FX Buffer, 10x | 5 |
| Purified DNA | Variable |
| Nuclease-free Water | Variable |
| **Total without FX Enzyme Mix** | **40** |

**Table 4. FX reaction mix setup (per sample) for <10 ng input DNA**

| Component | Volume (µL) |
|---|---|
| FX Buffer, 10x | 5 |
| Purified DNA | Variable |
| FX Enhancer | 2.5 |
| Nuclease-free water | Variable |
| **Total without FX Enzyme Mix** | **40** |

4. Add 10 µL FX Enzyme Mix to each reaction and mix well by pipetting up and down 20 times. It is critical to keep the reactions on ice for the entire time during reaction setup.
5. Briefly spin down the PCR plate/tubes and immediately transfer to the prechilled thermocycler (4°C). Resume the cycling program.
6. When the thermocycler program is complete and the sample block has returned to 4°C, remove samples and place them on ice.
7. Immediately proceed with adapter ligation.
8. Pierce the foil seal for each adapter well to be used, and transfer 5 µL from one DNA adapter well to each 50 µL sample from the previous protocol. Track the barcodes from each adapter well used for each sample.

   **Note**: If your DNA input is <10 ng, dilute the adapters according to Table 5.

**Table 5. Adapter dilution factors**

| Sample DNA amount | Adapter dilution |
|---|---|
| 20–99 pg | 1:1000 |
| 100–999 pg | 1:100 |
| 1–9 ng | 1:10 |

9. Freeze the adapter plate containing unused adapters. QIAseq adapters are stable for a minimum of 10 freeze-thaw cycles.

   **Important**: Only 1 single adapter should be used per ligation reaction. If adapters from another supplier are used, follow the manufacturer's instructions. Do not reuse adapter wells once the foil seal has been pierced.

10. Prepare the ligation Master Mix (per DNA sample, Table 6) in a separate PCR plate or tube on ice, and mix well by pipetting.

**Table 6. Ligation master mix setup (per sample)**

| Component | Volume (µL) |
|---|---|
| Ligation buffer, 5x | 20 |
| DNA ligase | 10 |
| Nuclease-free water | 15 |
| **Total** | **45** |

11. Add 45 µL of the ligation Master Mix to each sample, for a total of 100 µL, and mix well by pipetting. Incubate the ligation reaction at 20°C for 15 min.

    **Important**: Do not use a thermocycler with a heated lid.

12. Proceed immediately to adapter ligation cleanup using 0.8x (80 µL) Agencourt AMPure XP beads or QIAseq Beads.
13. Add 80 µL of resuspended Agencourt AMPure XP beads or QIAseq Beads to each ligated sample and mix well by pipetting.
14. Incubate the mixture for 5 min at room temperature. Pellet the beads on a magnetic stand (e.g., DynaMag) for 2 min, then carefully discard the supernatant.
15. Wash the beads by adding 200 µL of 80% ethanol. Pellet the beads on the magnetic stand and discard the supernatant. Repeat the wash once, for a total of 2 ethanol washes. Remove as much excess ethanol as possible.
16. Incubate the beads on the magnetic stand for 5–10 min or until the beads are dry.

    16a. Overdrying of Ampure XP beads may result in lower DNA recovery.

    16b. If using QIAseq Beads, ensure that the pellet is completely dry by visual inspection. Over drying QIAseq Beads will not affect the DNA elution.

17. Remove from the magnetic stand. Elute by resuspending in 52.5 µL of Buffer EB or 10 mM Tris·Cl, pH 8.0. Pellet the beads on the magnetic stand. Carefully transfer 50 µL of supernatant into a new plate or tube.

18. Perform a second purification using 1x (50 µL) Agencourt AMPure XP beads or 1.1x (55 µL) of QIAseq Beads following steps 14–16 for DNA binding and washing. Elute DNA by adding 26 µL Buffer EB or 10 mM Tris·Cl, pH 8.0. Pellet the beads and carefully collect 23.5 µL of purified DNA sample in a DNA LoBind tube for library amplification. If not proceeding immediately, the sample can be stored at –30°C to –15°C.

## Amplification of Library DNA

PCR-based library amplification is normally required if the input DNA amount is below 100 ng or if large amounts of libraries are required for downstream hybrid capture. This protocol is for high-fidelity amplification of the DNA library using the amplification reagents provided in the QIAseq FX DNA Library Kit.

### Things to do before starting

- Thaw QIAseq HiFi PCR Master Mix and Primer Mix on ice. Once reagents are thawed, mix them thoroughly by quick vortexing to avoid any localized concentrations.
- Always start with the cycling conditions specified in this protocol. The cycling has been optimized for use with QIAseq HiFi PCR Master Mix for even and high-fidelity amplification of sequencing libraries.
- Equilibrate Agencourt AMPure XP or QIAseq Beads to room temperature (15–25°C) for 20–30 min before use.

# Appendix H - RStudio code used for all plots and analyses

```
################## All code used for this project ##############

################ Bathymetric map ##########
#### packages used for making the map
library(ggplot2)
library(ggOceanMaps)
library(ggrepel)
library(tidyverse)
library(phyloseq)
library(stringi)
library(vegan)

# Read datasets
surface_sample_coordinates <- read_delim(file = "Surface_coord.csv")
fecal_sample_coordinates <- read_delim(file = "fecal_cord.csv")

# Transformation of coordinates
transformed_surface_sample_coordinates<-
transform_coord(surface_sample_coordinates)
transformed_fecal_sample_coordinates <-
transform_coord(fecal_sample_coordinates)

# Create map
bathymetric_map <- basemap(limits = c(14, 18, 69, 70), bathy.style =
"rcb") +
  ggspatial::annotation_scale(location = "br") +
  ggspatial::annotation_north_arrow(location = "tr", which_north =
"true")  +
  geom_point(data = transformed_fecal_sample_coordinates, aes(x =
Longitude, y = Latitude, color = "Fecal samples", shape = "Fecal
samples"), size = 2.5, position = position_jitter(width = 1500, height =
1500)) +
  geom_point(data = transformed_surface_sample_coordinates, aes(x =
Longitude, y = Latitude, color = "Surface samples", shape = "Surface
samples"), size = 2.5, position = position_jitter(width = 1500, height =
1500)) +  # Third dataset
  annotate(geom = "text", x = 628042.8, y = -2190246, label = "Andenes",
fontface = "italic", color = "grey22", size = 6, alpha=1) +
  annotate(geom = "text", x = 567976.9, y = -2109719, label = "Norwegian
Sea", fontface = "italic", color = "black", size = 6, alpha=1) +
  scale_color_manual(values = c("Fecal samples" = "brown", "Surface
samples" = "black"), name = "Data") +  # Specify colors for each dataset
  scale_shape_manual(values = c("Fecal samples" = 1, "Surface samples" =
2), name = "Data") +  # Specify shapes for each dataset
  guides(color = guide_legend(override.aes = list(fill = "white")),
shape = guide_legend(override.aes = list(fill = "white"))) +  # Specify
both color and shape legends with white background
```

```
  theme(legend.background = element_rect(fill = "white"),  # Set legend
background to white
        legend.key = element_rect(fill = "white")) +  # Set legend key
background to white
  ggtitle("Sampling Points of Sperm Whale Biopsies, Fecal Samples and
Surface samples in the Norwegian Sea")


############### General code for all plots and analyses ##############
# Read datafiles
table_of_counts <- read_delim(file = "name of DADA2 nochim.table")
metadata <- read_delim(file = "name of metadata file")
taxonomy <- read_delim("name of taxonomy dataset")

#### Multiple metadata options ####
# Only Pma No Blanks (fecal samples)
Pma_metadata <- metadata %>% filter(str_detect(Sample_name, pattern =
"Pma"))
# Only MOBY No Blanks (surface water samples)
Moby_metadata <- metadata %>% filter(str_detect(Sample_name, pattern =
"MOBY"))
# All metadata No Blanks
All_metadata_no_blanks <- metadata %>% filter(!str_detect(Sample_name,
pattern = "blank|air"))


##### Choose what metadata information you will use ####
chosen_metadata <- All_metadata_no_blanks

# Filter out empty cells and samples with less than 200 counts
otucounts_without_empty_cells <- table_of_counts %>%
  select(all_of(chosen_metadata$Sample_name)) %>%
  as.data.frame()

otucounts <- otucounts_without_empty_cells %>%
  select(which(colSums(otucounts_without_empty_cells) >= 400))

rownames(otucounts) <- table_of_counts$...1
as_tibble(otucounts)

taxonomy_for_phyloseq <- taxonomy %>%
  select(qseqid, kingdom, phylum, class, order, family, genus, species,
final.id)

#generating phyloseq objects
OTU = otu_table(otucounts, taxa_are_rows = TRUE)
```

```
matrix_taxonomy <- taxonomy_for_phyloseq %>%
  select(-qseqid) %>%
  as.matrix()
row.names(matrix_taxonomy) <- taxonomy_for_phyloseq$qseqid
TAX = tax_table(matrix_taxonomy)

sample_metada_data.frame <- as.data.frame(chosen_metadata)
row.names(sample_metada_data.frame) <- chosen_metadata$Sample_name
METADATA <- sample_data(sample_metada_data.frame)

SpermWhale_phyloseq = phyloseq(OTU, TAX, sample_data(METADATA))

### Sum technical replicates
SpermWhale_phyloseq_summed <- merge_samples(SpermWhale_phyloseq,
"Replicate", fun = sum)

### Sum by final.id to get lowest level resolution
Taxonomical_level <- "final.id"
SpermWhale_phyloseq_summed_tax_glommed <-
tax_glom(SpermWhale_phyloseq_summed, Taxonomical_level)

### Make relative counts
## Choose the tax level phyloseq object
SpermWhale_phyloseq_summed_for_plotting <-
SpermWhale_phyloseq_summed_tax_glommed

Spermwhale_otucounts_summed_df <-
as.data.frame(otu_table(SpermWhale_phyloseq_summed_for_plotting))
rltv_summed = sweep(t(Spermwhale_otucounts_summed_df), 2,
colSums(t(Spermwhale_otucounts_summed_df)), "/")
rltv_summed_OTU = otu_table(rltv_summed, taxa_are_rows = TRUE)

#re-generating a phyloseq objects with the relative counts and relevant
metadata
metadata_samples_after_summed_replicates <-
match(colnames(rltv_summed_OTU), METADATA$Replicate)
rltv_summed_METADATA =
sample_data(METADATA[metadata_samples_after_summed_replicates,2:4])
rownames(rltv_summed_METADATA) <-
METADATA[metadata_samples_after_summed_replicates,]$Replicate
colnames(rltv_summed_METADATA)[1] <- "Sample_name"

############### Rarefaction curves ############

# load libraries
library(tidyverse)
library(phyloseq)
```

```
library(stringi)
library(vegan)
library(dplyr)


################
# Rarefaction per sample
# Define a color palette with three distinct colors
my_colors <- c("green4", "blue", "red")

# Convert sample types to factor to ensure categorical mapping
sample_types <- factor(rltv_summed_METADATA$Sample_type)

# Reorder levels of sample_types to match the order of colors in
my_colors
sample_types <- factor(sample_types, levels = c("EtOH", "Water",
"Fecal"))

# Generate rarefaction curve with specified colors
rarefaction_curve <- rarecurve(Spermwhale_otucounts_summed_df, col =
my_colors[sample_types], label = FALSE, ylab = "MOTU's", xlab = "Reads")


################ Species accumulation curves ###############

library(tidyverse)
library(phyloseq)
library(stringi)
library(vegan)
library(dplyr)

# Filter samples based on sample types (Fecal samples summed)
Water_samples <- Spermwhale_otucounts_summed_df %>%
  filter(str_detect(row.names(Spermwhale_otucounts_summed_df), "^MOBY"))

Fecal_samples <- Spermwhale_otucounts_summed_df %>%
  filter(str_detect(row.names(Spermwhale_otucounts_summed_df), "^Pma"))


# Filter samples based on sample types (all types)
Water_samples <- Spermwhale_otucounts_summed_df %>%
  filter(str_detect(row.names(Spermwhale_otucounts_summed_df), "^MOBY"))

Fecal_samples <- Spermwhale_otucounts_summed_df %>%
  filter(str_detect(row.names(Spermwhale_otucounts_summed_df), "F$"))

Ethanol_samples <- Spermwhale_otucounts_summed_df %>%
```

```
  filter(str_detect(row.names(Spermwhale_otucounts_summed_df), "E$"))

# Function to plot species accumulation curve with box plot and sample
type label
spec_accum_Spermwhale <- function(df, color, Sample_type) {
  specaccum_1 <- specaccum(comm = df, method = "exact", permutations =
100)
  specaccum_2 <- specaccum(comm = df, method = "random", permutations =
100)

  # Plot species accumulation curve without adding a title
  plot(specaccum_1, xlab = "Number of Samples", ylab = "MOTU's")

  # Add box plot
  boxplot(specaccum_2, col = color, add = TRUE, pch = "+")

  # Add sample type label
  mtext(text = Sample_type, side = 3, line = 1, adj = 0, cex = 1.2)

  # Add grid lines at every 50 units on y-axis
  abline(h = seq(0, max(specaccum_1$richness), by = 10), col = "gray",
lty = 3)
}

# Plot species accumulation curves for each sample type
par(mfrow = c(1, 3))
spec_accum_Spermwhale(Water_samples, "blue", "Water samples")
spec_accum_Spermwhale(Fecal_samples, "red", "Fecal samples")
spec_accum_Spermwhale(Ethanol_samples, "green4", "Ethanol samples")


############## nMDS plots and ANOSIM #############
library(tidyverse)
library(phyloseq)
library(stringi)
library(vegan)
library(dplyr)
library(ggplot2)
library(ggthemes)

# Vegan ordination testing and plotting
# Use rows as samples and columns as sequences for the NMDS
################
# transpose otu table for use in vegan, rows are samples, columns are
otus
# rlt_otu_table <- sweep(otu_table, 2, colSums(otu_table), "/")
rltv_summed_OTU = as.data.frame(t(rltv_summed))
```

```
#t_otutable = as.data.frame(t(otu_table))

#transform the otu table (square root) and determine the best method for
calculating a distance matrix from it
sqrt_t_otus = sqrt(rltv_summed_OTU)



rank.totus = rankindex(as.matrix(sqrt_t_otus), sqrt_t_otus, indices =
c("bray", "euclid", "manhattan", "horn"), method = "spearman")
print(paste("The highest rank was given by the", names(sort(rank.totus,
decreasing = TRUE) [1]), "method."))

# using the bray-curtis method to construct community distance matrix
otus_dist = as.matrix((vegdist(sqrt_t_otus, "bray")))

#perform NMDS, define coordinates
veganNMDS = metaMDS(sqrt_t_otus, distance = "bray", autotransform =
TRUE)
veganNMDS
MDS1 = veganNMDS$points[,1]
MDS2 = veganNMDS$points[,2]
#ordination_plot <- ordiplot(ord = veganNMDS, display = "sites", shrink
= shrink)
ordination_plot <- ordiplot(ord = veganNMDS, display = "sites")

# Ordering mapping file as MDS vectors
ordering_map <- match(x = names(MDS1), table =
rownames(rltv_summed_METADATA))


# build a data frame with NMDS coordinates and metadata =  mapping file
NMDSvegan = data.frame(MDS1 = MDS1, MDS2 = MDS2, Group =
rltv_summed_METADATA$Sample_type[ordering_map])
head(NMDSvegan)

# Plotting NMDS with ellipses and a title
ggplot(NMDSvegan, aes(x = MDS1, y = MDS2, colour = Group)) +
  geom_point(size = 2) +
  stat_ellipse() +  # Add ellipses
  theme_bw() +
  scale_color_manual(values = c("green", "red")) +  # Assigning colors
manually
  ggtitle("Title of nMDS plot")  # Adding a title


# Perform ANOSIM
```

```
anosim_result <- anosim(otus_dist, grouping =
rltv_summed_METADATA$Sample_type[ordering_map])


# Print ANOSIM results
print(anosim_result)

########## SIMPER analysis ##########
# Run simper on final.id level
# Learned from here:
# https://kaiw3.github.io/Multivariate_Data_Tutorial_Core/docs/#section4
# Filter out columns with zero counts
sqrt_t_otus_filtered <- sqrt_t_otus[, colSums(sqrt_t_otus) > 0]

# Continue with your analysis using the filtered dataset
sqrt_t_otus_filtered
simp_water_feces <- simper(comm = as.matrix(sqrt_t_otus_filtered), group
= rltv_summed_METADATA$Sample_sum)
simper_summary <- summary(simp_water_feces, ordered = TRUE, digits = 3)
view(simper_summary)


################ Bubble plots ############
# packages used
library(tidyverse)
library(phyloseq)
library(stringi)
library(phyloseq)
library(RColorBrewer)
library("ggplot2")


# Aggregate data if necessary
rltmds_aggregate <- aggregate(rltmds$Abundance, by = list(Sample =
rltmds$Sample, Taxa = as.factor(rltmds[, tax_level])), FUN = sum)
colnames(rltmds_aggregate) <- c("Sample", tax_level, "Abundance")

# Define custom palette
custom_palette <- brewer.pal(n = 5, name = "Dark2")

# Filter out phyla with zero abundance
rltmds_filtered <- rltmds_aggregate %>%
  filter(Abundance > 0)
```

```
Bubble_plot <- ggplot(data = rltmds_filtered, aes(x = Sample, y =
reorder(final.id, desc(Abundance)), size = Abundance, fill = Abundance))
+
  geom_point(alpha = 0.7, shape = 21) +
  scale_size_continuous(range = c(1, 15), breaks = c(0.01, 0.1, 0.2,
0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1)) +
  scale_fill_viridis(option = "plasma", limits = c(0, 1)) +  # Adjusting
color scale limits
  labs(title = "Same Place",
       x = "Sample type", y = "Taxa (Final.id)") +
  theme_minimal() +
  theme(legend.position = "bottom")


############## Stacked barplots #############
# Stacked barplots for chosen taxonomic level:
rltmds = psmelt(SpermWhale_phyloseq_summed_relative)

## Choose either: Kingdom, Phylum, Order, Family, Genus, or Final.ID
tax_level = "phylum"

rltmds_aggregate <- aggregate(rltmds$Abundance,
by=list(Sample=rltmds$Sample, Taxa = as.factor(rltmds[,tax_level])),
FUN=sum)
colnames(rltmds_aggregate) <-c("Sample", tax_level, "Abundance")

##Plotting

#defining if plot legend separated or together
tax_degree = nlevels(as.factor(rltmds[,tax_level]))

# tell the name of your plot
opt <- list()
opt$out <- "Chosen name"

# Choose a nice pallete

Small <- c("#FFFF33", "#7B5B3E", "#4682B4",
           "#BC80BD", "#2E8B57")

Medium <- c("#252525", "#999999", "#FDB462",
            "#2E8B57", "#FFFF33", "#F781BF", "#7B5B3E",
            "#BEBADA",
            "#FB8072", "#B3DE69", "#FCCDE5",
            "#CD853F", "#66C2A5", "#BC80BD",
            "#98FB98", "#FF7F00", "#FF1493", "#E41A1C",
            "#00CED1", "#4682B4")
```

```
Many <- c("#66C2A5", "#2E8B57", "#999999", "#984EA3", "#FF7F00",
          "#E41A1C", "#FFFF33", "#FDB462", "#F781BF", "#543005",
          "#8DA0CB", "#8DD3C7", "#4DAF4A", "#377EB8", "#BEBADA",
          "#FB8072", "#80B1D3", "#E5C494", "#B3DE69", "#FCCDE5",
          "#CD853F", "#BC80BD", "#8A2BE2", "#A65628", "#252525",
          "#98FB98", "#FFFFB3", "#FF1493", "#FFD700", "#F0F0F0",
          "#D9D9D9", "#FFD92F", "#00CED1", "#4682B4", "#467467",
"#304304")


custom_palette <- Medium

if(tax_degree <= 20) {
  pdf(paste(opt$out,".pdf",sep = ""), width = 10, height = 10)
  plot = ggplot(data = rltmds_aggregate, aes_string(x="Sample", y =
"Abundance", fill = tax_level)) +
    geom_bar(stat = "identity", colour = "white") +
    scale_fill_manual(values = custom_palette)
  print(plot + theme(axis.text.x = element_text(angle = 90, hjust = 1))
+ labs(y = "Abundance", x = "Sample type"))
  dev.off()
  print(paste("Plot was created in file ",getwd(),"/",opt$out,".pdf",
sep=""))
} else {
  pdf(paste(opt$out,".pdf",sep = ""), width = 10, height = 10)
  plot = ggplot(data = rltmds_aggregate, aes_string(x="Sample", y =
"Abundance", fill = tax_level)) +
    geom_bar(stat = "identity", colour = "white") +
    scale_fill_manual(values = custom_palette)
  print(plot + theme(legend.position = "none",axis.text.x =
element_text(angle = 90, hjust = 1)) + labs(y = "Relative abundance", x
= "Sample Type"))
  dev.off()
}
#plotting legend isolated
pdf(file = paste("legend_",opt$out,".pdf",sep = ""), width = 12, height
= 8)
plot_legend <- function(a.gplot){
  tmp <- ggplot_gtable(ggplot_build(a.gplot))
  leg <- which(sapply(tmp$grobs, function(x) x$name) == "guide-box")
  legend <- tmp$grobs[[leg]]
  return(legend)
}
#Plotting only legend. Important for species and genus factors. Too many
levels.
mylegend = plot_legend(plot)
```

```
library(grid)
grid.draw(mylegend)
dev.off()
print("Legend and Plot were created in separate files in reason of the
high number of taxonomical levels")

#### View read count from each phylum
phylum.sum = tapply(taxa_sums(SpermWhale_phyloseq_summed),
tax_table(SpermWhale_phyloseq_summed)[, "phylum"], sum, na.rm=TRUE)
view(phylum.sum)
```

# Appendix I - Metadata

Metadata can be handed out. The files were too large to include in appendix, but contact me on my e-mail (soeiland@hotmail.com) and I will send it if wanted.