Metabarcoding-driven discovery of copepod parasites

Even Sletteng Garvang

Thesis submitted for the degree of
Master in bioscience
(Marine biology and limnology)
60 credits

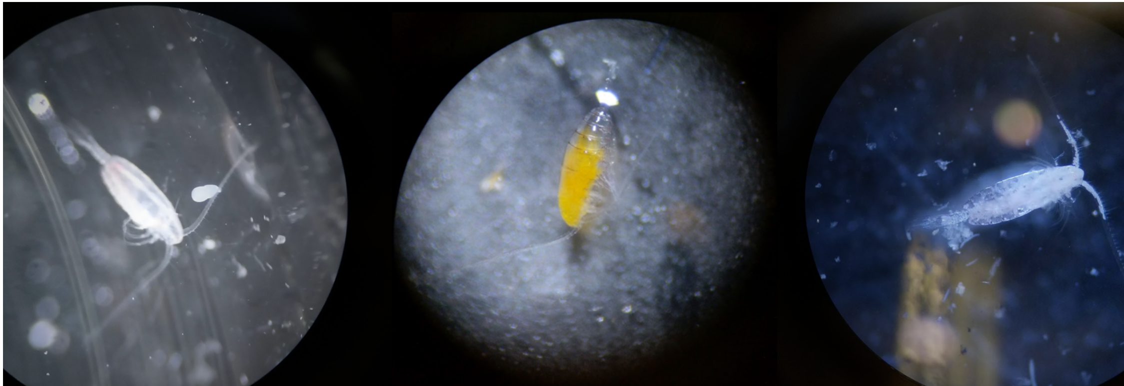
Section for aquatic biology and toxicology
Department of biosciences
Faculty of mathematics and natural sciences

UNIVERSITY OF OSLO

Spring 2021

Metabarcoding-driven discovery of copepod parasites

Even Sletteng Garvang



© 2021 Even Sletteng Garvang

Metabarcoding-driven discovery of copepod parasites

http://www.duo.uio.no/

Printed: Reprosentralen, University of Oslo

Abstract

Pelagic copepods are hosts to numerous protistan parasites, which can have devastating effects on host fitness. Effects of parasitism include increased mortality, behavioral modifications, sterility and death. Although both pelagic copepods and parasitism are considered important, the topic of parasitism in copepods is vastly understudied. Traditionally, the parasites have been difficult to detect in regular plankton samples due to their small size, low visibility, and often low prevalences. However, modern molecular methods like metabarcoding can increase our ability to reliably detect parasites.

In this thesis, I used metabarcoding to find parasites in zooplankton samples from Oslofjorden, Norway. I also re-analyzed metabarcoding data from the BioMarKs project, which sampled water and sediments in several locations across Europe. I used this data to investigate two fundamental questions about parasites in copepods: *Where* are the parasites, and *When* are they there? In addition, I identified new DNA sequences for 5 different copepod parasites, which aided in the search for parasites in the data. I also evaluated metabarcoding as a method for studying parasitism in copepods.

I used two different primer sets for the metabarcoding in this thesis, both amplifying DNA from the the V4 region of the 18S gene. One primer set was general, made to amplify all taxa equally. The other was specifically made to block metazoan sequences (anti-metazoan), in an attempt to make parasites easier to detect by overcoming the biomass differences between hosts and parasite. The use of two primer sets in this study had no obvious benefit, as most sequences were still metazoan. In addition, the 18SV4 region could not distinguish between important metazoan groups. For future studies, I recommend using the anti-metazoan primers in conjunction with primers from a different genomic region, for example COI or 28S.

My main conclusions are that parasites of copepods are found everywhere you look, and that they are present year-round in Oslofjorden. Many of the parasites seem to have a seasonal variation that follows the variation of hosts, as predicted by theory. In Oslofjorden, seasonal differences in parasite occurrence are larger than spatial differences. Still, there are some differences between the station outside the Drøbak sill and those inside of the sill that I attribute to host availability. In the BioMarKs data, which spans a larger geographic area, there are large differences in parasites detected between sites. Metabarcoding has its limitations, but is a promising tool for researching parasites of copepods. With good study design and more reference sequences becoming available in the future, many of those limitations can be overcome.

The 5 newly identified parasite sequences were very important for parasite detection in this thesis, and more than doubled the detected parasite genera in the Oslofjord data. In addition, the sequencing of the previously reported parasite *Ichthyophonus sp.* gave new taxonomic insight. Another of the sequenced parasites, *Chromidina sp.*, represents a previously undiscovered parasite of copepods. I argue that obtaining more DNA sequences of copepod parasites will be very important for future research.

Acknowledgements

Thanks to my supervisors, Tom Andersen and Josefin Titelman for always being willing to help with their range of expertise, from the purely theoretical (how can mathematics describe biology?), to the very practical aspects (how *do* you get a zooplankton sample out of the cup at the end of the net?) of this thesis. In particular, Josefin's advice of "start reading right away!" turned out to be some of the best advice I could have gotten. Thanks for being supportive throughout the whole process, for the weekly "parasite-friday" meetings, and for always replying quickly and giving feedback whenever I needed to discuss anything. You are both inspiring with your vast amount of experience and knowledge, and I have learned a lot from you!

A special thanks to Lasse Eliassen for countless hours in field and lab, and for our adventures in Drøbak. This thesis is a result of our long days together, with struggles, successes and lots fun. Many people, at UiO or elsewhere, has helped with various matters during the thesis work: Thanks to Alexander Eiler, Bente Edvardsen and Jing Wei for help with lab work and more; to Lasse Riemann, Micah Durnthon, Dominik Forster and Frédéric Mahé for various help with metabarcoding and data analysis; to Gregor Gilfillan and Teodora Ribarska at the Norwegian Sequencing Centre for patiently answering our questions; to Per Færøvig and Berit Kaasa for help with a lot of practical matters; to André Staalstrøm and Jon Albretsen from NIVA for providing CTD data; and to Olav Brevik for assisting with picking colorblind-friendly colors. Thanks to Rita for help with field work and all things copepod, and to captain Sindre and the rest of the crew at R/V Trygve Braarud for long days and genuine interest in what we are doing.

Thanks to my family and friends for always being interested in my studies and asking about this obscure topic that I write about. Thanks to my fellow students for making 5 years at UiO so great, to the fungal master students for taking care of me the last week before handing in, and to my band, Overgump, for constantly reminding me how fun biology is! Thanks to my brother, Kristian, for drawing me as a copepod, and to my mother, Marianne, for a lot of support and even more waffles. Thanks to my wife Karoline for giving me a reality check whenever I was stressed, for getting excited about dogs and ducks with me, and for believing in me ever since I decided overnight to become a biologist.

Thank you Arne, for everything you were, and everything you taught me.

*This thesis was written using the **bookdown** package (Xie, 2020) in R, which was built on top of R Markdown and **knitr** (Xie, 2015). All code was run using R version 4.0.3 (2020-10-10) (R Core Team, 2020) in RStudio version 1.4.1103 (RStudio Team, 2021)*

Table of contents

Introduction **1**

Parasitism 1

Pelagic copepods 3

Who are the parasites? Known parasites and their impacts 4

Two main questions 11

Methods **17**

The study area 17

Field sampling on RV Trygve Braarud 18

Homogenization and DNA extraction 20

Preparing DNA samples for sequencing 22

Bioinformatics pipeline 27

Data analysis 29

Analyzing existing data sets 31

Results **32**

Overview of the metabarcoding data 32

Primer assessment 35

Multivariate analysis 36

Patterns of abundant taxonomic groups 38

Identification of individual parasites 40

Patterns of parasite occurrence 41

Co-occurrence analysis 43

Searching in BioMarKs data 43

Discussion **46**

Who are the parasites? — Revisited 46

When are the parasites there? 48

Where are the parasites? 49

Metabarcoding as a tool 50

Going further 52

Conclusion 53

References **54**

Appendix A: Supplementary figures **64**

Appendix B: Scripts **68**

Bioinformatic processing 68

Data analysis 69

Introduction

Parasites are everywhere. The act of living off other organisms to survive has proven so successful that, by some estimates, over half of the world's species are parasites (Windsor, 1998). In pelagic systems, planktonic copepods—possibly the most abundant animal group on earth (Verity & Smetacek, 1996)—are hosts to a variety of different parasites (e.g., Shields, 1994). Despite the importance of both parasitism and pelagic copepods, the intersection of the two—parasites on copepods—is severely understudied (Skovgaard, 2014). In this thesis, I have explored the seasonal and spatial dynamics of copepod parasites in the Oslofjord, Norway, and other locations in Europe, using metabarcoding data. First, I present parasitism and copepods separately before looking at the parasites on copepods specifically. I review the field of parasitism in copepods, discussing what we know, what we do not know, and some challenges in studying these systems. Then, I outline how I have investigated this.

Parasitism

At an individual level parasites have, by definition, a negative impact on their host (see Box 1), impairing fecundity, survival or growth of their hosts (Combes, 1996). At a larger scale, parasites can impact whole populations and communities (Anderson & May, 1981). Parasites are essential for energy flow through trophic levels, and including them in food webs can aid our understanding of the ecosystem they are a part of (Lafferty et al., 2006, 2008). Parasitism also has a crucial role in host evolutionary history, even in humans, where selection driven by parasites has increased genetic diversity (Fumagalli et al., 2011).

Box 1: Symbiosis

In this text, I use the original symbiosis definitions of de de Bary (1879; translation by Oulhen et al., 2016), namely that symbiosis is different organisms living together. De Bary names three main modes of symbiosis:

- **Mutualism** where both parts benefit from the relationship
- **Commensalism** where one party benefits and the other is not affected
- **Parasitism** where one party benefits and the other is negatively affected

The boundaries between these modes are blurred, and some relationships can even change from one mode to another (e.g., Baker et al., 2018).

Parasites live off the energy their hosts generate that would otherwise have been used by the host for growth and reproduction, thereby directly impacting host fitness (Lafferty & Kuris, 2009). Interestingly, if a parasite is too virulent—i.e., reduces host survival too much—the parasite will not have time to complete its development within the host, and parasite fitness will be reduced (Lafferty & Kuris, 2009). A parasite that increases mortality too much will also risk that the host population they depend upon will go extinct (Sorensen & Minchella, 2001). There is a trade-off between maximizing parasite reproduction while still keeping the host alive as long as possible (the virulence trade-off, Sorensen & Minchella, 2001). As a result of this, many (but not all, see the conclusions of May, 1983) parasites evolve towards lower virulence.

One way for a parasite to maximize energy gain while minimizing host mortality is to draw energy from the reproductive organs of the host (Lafferty & Kuris, 2009). This will negatively affect the host's fecundity, with effects comparable to increased mortality in the host (Anderson & May, 1981). Consequently, the effects of a parasitoid are comparable with that of a parasitic castrator (Box 2), i.e., reproductive death has a similar impact to actual death (Kuris, 1974). However, one important difference is that a castrated individual will still compete for resources with healthy individuals, which may further impact the population (Lafferty & Kuris, 2009; Skovgaard, 2005). Naturally, hosts will adapt to coping with parasitic castration. Apart from avoiding infection altogether, an interesting adaption to castration is *fecundity compensation*, where hosts increase their reproductive effort in the window between becoming infected and being castrated (Gleichsner et al., 2016; Minchella, 1985).

Box 2: Types of parasites

Parasites can be divided into different categories based on the nature of their relationship with their host. Except for ecto- and endoparasites, the categories are not mutually exclusive.

- **Ectoparasites** attach to the outside of the host body.
- **Endoparasites** live inside the body of their host, either in the gut or in body cavities and tissues.
- **Parasitoids** obligately kill their host, i.e., killing the host is part of their life cycle.
- **Parasitic castrators** castrate their host, stopping host reproduction altogether.

Another significant effect of parasitism on an individual level is the manipulation of host behavior. Although this manipulation sometimes happens as a side-effect of a parasite infection, it is more often directly connected to many parasites' (complex) life cycles (Poulin, 1994, see also Box 3). Parasites manipulate their host to maximize transmission: some may manipulate their host behavior to get eaten and transmitted to the next host (e.g., Thomas & Poulin, 1998). Others may position its host for optimal spreading of spores (like the infamous *Ophiocordyceps* fungi in ants, Andersen et al., 2009). Thus, the behavioral manipulation may in itself affect host fitness (higher chance of getting eaten by predators) or be negligible compared to the other effects of the parasite.

On a larger scale, parasites can control their host populations. Regulation can happen when the parasite-induced mortality is higher than net host reproduction rates, which naturally can be achieved by high parasite-induced mortality, but also by a parasite negatively impacting fecundity of (or even castrating) the host (Anderson & May, 1981). Because of this relationship, a more pathogenic parasite will be more likely to regulate its host population. Furthermore, pathogenicity is inversely related to the prevalence of infection, meaning both that parasites can regulate a host population even at low prevalence and that a parasite with high prevalence actually is less likely to affect host population size (Anderson & May, 1981). Therefore, the degree of regulation of a parasite can not be inferred from prevalence alone but needs to be investigated with a combination of experiments and observational studies (Tompkins et al., 2002).

Box 3: Parasite life cycles

There are two distinct kinds of parasite life cycles (reviewed in Auld & Tinsley, 2015):

- **Simple life cycles**, where a parasite completes its life cycle in a single host
- **Complex life cycles**, where a parasite needs to subsequently infect multiple hosts to complete its life cycle.

Parasites with complex life cycles often undergo growth and asexual reproduction in **intermediate hosts**, while the host where sexual reproduction occurs is called the **final host**. Some parasites have free-living stages between different hosts, while others transmit from prey to predator—i.e., is dependent on one host being eaten by the next.

Pelagic copepods

Pelagic copepods are the most abundant animals in the ocean (Loeng & Drinkwater, 2007) and dominate the zooplankton biomass (Verity & Smetacek, 1996). Many copepods are important consumers of phytoplankton, but can also be carnivores or detritivores or use a combination of feeding modes (Mauchline, 1998). They have a well-developed sensory system, efficient feeding strategies and an almost unmatched escape ability, all contributing to making them some of the most successful animals on the planet (Kiørboe, 2011).

Most pelagic copepods are similar in shape, but they have a considerable variation in size and life-history traits (Verity & Smetacek, 1996). Small copepods (and early stages of larger copepods) can be smaller than 200 µm (Gallienne & Robins, 2001) while, for example, the large carnivorous *Paraeuchaeta norvegica* can be more than 6 mm long (Leinaas et al., 2016), and they naturally have very different challenges regarding growth, reproduction and survival. Copepods have different spawning strategies, ranging from small copepods that frequently reproduce to large, long-lived species reproducing once a year or less often, and variation in this exists even within the same species, e.g., species of *Calanus* (Mauchline, 1998). In seasonal seas, copepods have adopted different strategies to survive harsh winters: some have diapausing (resting) copepodid stages that overwinter at deep waters, while smaller coastal copepods have diapausing eggs in the sediments (Mauchline, 1998).

Copepods are ecologically important, acting as a trophic link between primary producers and larger carnivores (Loeng & Drinkwater, 2007). They graze on phytoplankton and ciliates (Gifford, 1991) and are preyed upon by many commercially and ecologically important fish species, including herring, capelin and cod larvae (Loeng & Drinkwater, 2007). Other animals, like chaetognaths and jellyfish, prey on copepods as well (Verity & Smetacek, 1996).

Copepods also play an important role in the global carbon cycle (Sanders et al., 2014). For one, copepods produce compact fecal pellets that quickly sink to the deep layers of the ocean. They also consume a substantial amount of plankton in the upper layers, often respired in deeper waters due to vertical migration. Both of these processes contribute to the removal of carbon from surface waters and the atmosphere.

Given the importance of parasitism and pelagic copepods, one would think that the topic of parasites on pelagic copepods is well studied. However, the ecological impact of parasitism in copepods (and the zooplankton in general) is poorly understood (Skovgaard, 2014). The diversity of parasites is underestimated, and it is likely that many parasites have yet to be discovered (Skovgaard, 2014). Below, I outline what we *do* know and briefly introduce the known parasites.

Who are the parasites? Known parasites and their impacts

Most of the species of parasites on copepods known today were described by French researcher Édouard Chatton in the early 1900s (Chatton, 1920, some of his work shown in Figure 1). Since then, some species have been discovered and some redefined, but the groups created by Chatton mostly stand today. The parasites are taxonomically diverse and have various effects on their hosts (Table 1).

The parasites of copepods have been subject to limited research since their initial discovery. Most studies focus on parasite morphology and taxonomy (Skovgaard et al., 2007, e.g., 2005; Skovgaard & Daugbjerg, 2008) or prevalence and species diversity (e.g., Alves-de-Souza et al., 2011; Coats et al., 2008). Little is known about host-specificity, but most parasites have been reported from several hosts (Table 2) Only a handful of studies have quantified the effects on individual hosts (Albaina & Irigoien, 2006; Fields et al., 2015), and a few have investigated the impact of parasitism on the population and community levels (Ianora et al., 1987; Skovgaard & Saiz, 2006).

Table 1: Known protistian parasite genera, and the effect on their copepod hosts. See references in text and Table 2.

Higher taxonomy	Genus	Type of parasite	Effect on host
Alveolata, ”MAGI”	”RP parasite”	endoparasite	unknown
Alveolata, Apicomplexa, Gregarinasea	Cephaloidophora	endoparasite	harmless?
Alveolata, Apicomplexa, Gregarinasea	Ganymedes	endoparasite	harmless?
Alveolata, Apicomplexa, Gregarinasea	Thiriotia	endoparasite	harmless?
Alveolata, Ciliophora, Apostomatida	Vampyrophrya	ectoparasite	increased mortality
Alveolata, Dinoflagellata, Dinophyceae	Blastodinium	endoparasite	fecundity, respiration
Alveolata, Dinoflagellata, Ellobiopsea	Ellobiopsis	ectoparasite	fecundity
Alveolata, Dinoflagellata, Gymnodiniales	Chythriodinium	egg parasite	lethal
Alveolata, Dinoflagellata, Gymnodiniales	Dissodinium	egg parasite	lethal
Alveolata, Dinoflagellata, Syndiniales	Syndinium	endoparasite/parasitoid	lethal
Ophistokonta, Ichthyosporea	Ichthyophonus	endoparasite	behavioral modification, color change
Rhizaria	Paradinium	endoparasite	fecundity?

One reason for the lack of studies may be that the parasites are difficult to find, in several ways. For one, the parasites are small, often microscopic, and identifying them requires special effort. Some are also endobionts and normally only visible through conspicuous coloration in the copepods (Skovgaard et al., 2005; Torgersen et al., 2002). Since transparency has to be actively maintained (Bagge, 2019), and any fixation method inevitably kills the copepods and makes them more opaque, endoparasites can be very hard to find in routine plankton counts. Furthermore, some parasites, like *Blastodinium*, *Syndinium* and *Paradinium* are invisible until they reach more mature stages of infection, and the visibility of the latter two is probably lost when samples are fixated (Skovgaard & Saiz, 2006).

The parasites can also be challenging to find due to low prevalences of infection in the studied populations (Table 3). However, the true prevalence may be difficult to assess due to limited visibility (Skovgaard & Saiz, 2006). This may have reduced the interest in and assumed impact of parasites on copepods. Low prevalence, however, does not necessarily mean that parasites are not important for the copepod populations, as parasites can still regulate their hosts at low prevalences (Anderson & May, 1981).

*Table 2: Copepod parasites and known hosts. Since some research papers were difficult to obtain, and Chatton wrote in French, some data is from secondary sources: *data from Skovgaard et al. (2012). †data from Horiguchi et al. (2006).*

Parasite	Hosts	References
"RP parasite"	*Calanoida, Calanus finmarchicus, Clausocalanus sp., Pseudocalanus elongatus*	Skovgaard & Daugbjerg (2008); Skovgaard & Daugbjerg (2008); Jepps (1937)
Blastodinium apsteini	*Clausocalanus arcuicornis, Clausocalanus furcatus, Paracalanus aculeatus*	Sewell (1951)*; Chatton (1920)*
Blastodinium chattoni	*C. furcatus, Clausocalanus arcuicornis, Cosmocalanus darwini, Eucheata indica, Nannocalanus minor, Paracalanus aculeatus, Paracalanus denudatus, Paracalanus parvus*	Sewell (1951)*
Blastodinium contortum	*Acartia clausi, Acrocalanus gracilis, Calocalanus styliremis, Clausocalanus arcuicornis, Clausocalanus furcatus, Cosmocalanus darwini, Eucheata indica, Nanocalanus minor, Paracalanus aculeatus, Paracalanus denudatus, Paracalanus parvus, Subeucalanus pileatus, Subeucalanus subtenuis*	Skovgaard & Saiz (2006); Pasternak et al. (1984)*; Sewell (1951)*; Coats et al. (2008); Skovgaard et al. (2007); Chatton (1920)*
Blastodinium crassum	*Calocalanus styliremis, Clausocalanus arcuicornis, Clausocalanus furcatus, Paracalanus aculeatus, Paracalanus nanus, Paracalanus parvus*	Sewell (1951)*; Coats et al. (2008); Chatton (1920)*; Chatton (1920)*
Blastodinium elongatum	*Centropages sp., Scolecithrix bradyi*	Chatton (1920)*
Blastodinium galatheanum	*Acartia negligens, Acartia sp.*	Skovgaard & Salomonsen (2009)
Blastodinium hyalinum	*Acartia clausi, Calanus finmarchicus, Centropages sp., Clausocalanus arcuicornis, Clausocalanus furcatus, Paracalanus aculeatus, Paracalanus denudatus, Paracalanus parvus, Paracalanus sp., Paraeuchaeta antarctica, Pseudocalanus elongatus, Pseudocalanus sp.*	Skovgaard & Saiz (2006); Jepps (1937); Vane (1952)*; Cattley (1948); Sewell (1951)*; Fields et al. (2015); Chatton (1920)*
Blastodinium mangini	*Farranula gibbula, Farranula rostrata, Oncaea cf. scottodicarloi, Oncaea media, Oncaea sp., Oncaea venustra, Triconia conifera*	Skovgaard & Saiz (2006); Sewell (1951)*; Skovgaard (2005); Chatton (1920)*
Blastodinium mangini var. oncaea	*Farranula Rostrata, Oncaea media, Triconia minuta*	Chatton (1920)*
Blastodinium navicula	*Corycaeus giesbrechti, Oncaea venustra*	Skovgaard & Saiz (2006); Sewell (1951)*; Skovgaard (2005); Skovgaard et al. (2007); Chatton (1920)*
Blastodinium oviforme	*Corycaeus crassiusculus, Corycaeus speciosus, Farranula gibbula, Oithona nana, Oithona plumifera, Oithona similis, Oncaea media, Oncaea venusta, Triconia conifera*	Skovgaard & Saiz (2006); Sewell (1951)*; Chatton (1920)*
Blastodinium pruvoti	*Clausocalanus arcuicornis, Clausocalanus furcatus, Nannocalanus minor, Paracalanus parvus*	Skovgaard & Saiz (2006); Sewell (1951)*; Chatton (1920)*
Blastodinium sp.	*Clausocalanus arcuicornis, Clausocalanus furcatus, Clausocalanus lividus, Clausocalanus sp., Corycaeus flaccus and Acartia clausi, Corycaeus typicus, Farranula carinata, Farranula sp., Paracalanus parvus, Paraeuchaeta antarctica*	Skovgaard & Saiz (2006); Ianora et al. (1990); Ianora et al. (1987); Øresland (1991); Drits & Semenova (1985)*; Horiguchi et al. (2006)
Blastodinium spinulosum	*Acrocalanus gracilis, Clausocalanus arcuicornis, Clausocalanus farrani, Clausocalanus spp., Paracalanus aculeatus, Paracalanus denudatus, Paracalanus parvus*	Skovgaard & Saiz (2006); Sewell (1951)*; Chatton (1920)*; Kofoid (1931)†
Blastodinium spp.	*Calanoida, Corycaeidae, Oithonidae, Oncaeidae*	Alves-de-Souza et al. (2011); Alves-de-Souza et al. (2011); Alves-de-Souza et al. (2011); Alves-de-Souza et al. (2011)
Ellobiopsis chattoni	*Acartia clausi, Acartia danae, Acrocalanus gibber, Calanus euxinus, Calanus finmarchicus, Centropages ponticus, Centropages typicus, Clausocalanus sp., Cosmocalanus darwinii, Euchaeta marina, Metridia longa, Paraclanus aculeatus, Paraclanus crassirostris, Undinula vulgaris*	Fahmi & Hussain (2003); Gómez, López-García, et al. (2009); Artüz (2016); Timofeev (2002); Jepps (1937); Marshall et al. (1934); Santhakumari & Saraswathy (1979); Wickstead (1963); Hoffman & Yancey (1966)
Ellobiopsis fagei	*Clausocalanus sp.*	Skovgaard & Saiz (2006)
Ellobiopsis sp.	*Calanus finmarchicus, Calanus glacialis, Calanus helgolandicus, Temora stylifera*	Albaina & Irigoien (2006); Walkusz & Rolbiecki (2007); Skovgaard & Saiz (2006)
Ichthyophonus sp.	*Calanus finmarchicus, Calanus sp.*	Jepps (1937); Torgersen et al. (2002)
Paradinium poucheti	*Calanus finmarchicus, Oithona similis*	Skovgaard & Daugbjerg (2008); Skovgaard & Daugbjerg (2008); Jepps (1937)
Syndinium sp.	*Calanus finmarchicus, Clausocalanus arcuicornis, Corycaeus giesbrechti, Corycaeus sp., Euterpina acutifronsa, Oithona similis, Paracalanus indicus, Paracalanus parvus*	Skovgaard & Saiz (2006); Jepps (1937); Ianora et al. (1990); Ianora et al. (1987); Kimmerer & McKinnon (1990); Marshall et al. (1934)
Syndinium turbo	*Corycaeus sp., Paracalanus parvus*	Skovgaard & Saiz (2006); Skovgaard et al. (2005); Skovgaard et al. (2005)
Vampyrophrya pelagica	*Acartia longiremis, Calanus sinicus, Centropages hamatus, Corycaeus affinis, Euterpina acutifrons, Paracalanus parvus*	Ohtsuka et al. (2004); Grimes & Bradbury (1992)

Figure 1: Various copepod parasites as drawn by Chatton (1920). 36: Blastodinium pruvoti *infecting* Clausocalanus arcuicornis. *37:* B. hyalinum *infecting* C. furcatus. *38:* B. spinulosum *infecting* C. furcatus. *39:* B. crassum *infecting* Paracalanus parvus. *40:* Schizodinium sparsum *(to my knowledge not reported in any later studies) infecting* Corycaeus rostratus. *41:* Syndinium turbo *infecting* P. parvus. *42:* Paradinium poucheti *infecting* Acartia clausi. *All are magnified 140x in the original format (ca. A4).*

Many of the parasites of copepods are dinoflagellates or related to these, while others have unclear taxonomic positions (Table 1). Some are endoparasites, occupying either the gut lumen or the body cavity of the copepods (e.g., Skovgaard, 2005; Skovgaard et al., 2005; Skovgaard & Daugbjerg, 2008). Other are ectoparasites, attaching to antennae, mouth appendages and other parts of the copepod body (e.g., Gómez, López-García, et al., 2009; Ohtsuka et al., 2004). Several species parasitize copepod eggs, with lethal outcome (Gómez, Moreira, et al., 2009). Although metazoan and fungal parasites on copepods have been recorded, most studies have been on protistan parasites, i.e., all parasites *except* metazoan and fungal. I emphasize the protistan parasites targeting copepodids and adult copepods (i.e., non-egg parasites) in this thesis. Below I describe the most common genera.

Blastodinium

Blastodinium is the most well-studied genus of the parasites of copepods and currently encompasses 13 species (Guiry & Guiry, 2020). They reside in the gut lumen of the copepod and are often so large that they fill the entire gut (Figures 1 (36-39) and 2; Chatton, 1920). Most species have chloroplasts and are thus pigmented and relatively easy to spot in a stereomicroscope or using epifluorescence (Skovgaard, 2005). However, *Blastodinium hyalinum* has degenerated chloroplasts and is colorless (Soyer, 1970). *B. hyalinum* is the largest species in the genus and is believed to be the only one found at higher latitudes (see Skovgaard et al., 2012, fig. 10).

Table 3: Reported prevalences from the literature. Where the authors reported multiple prevalences (e.g., the prevalence in each sample), the entire range of prevalences is shown. Since some research papers were difficult to obtain, and Chatton wrote in French, some data is from secondary sources. [*]*data from Skovgaard et al. (2012).*

Species	Prevalence	Reference
Blastodinium contortum	0-4 %	Skovgaard & Saiz (2006)
	6-12 %	Pasternak et al. (1984)[*]
Blastodinium crassum	0.6-2.0 %	Coats et al. (2008)
	1.5 %	Chatton (1920)[*]
Blastodinium hyalinum	3.7-66 %	Vane (1952)[*]
	0.3-20.5 %	Cattley (1948)
	0-58 %	Fields et al. (2015)
Blastodinium mangini	0-17 %	Skovgaard & Saiz (2006)
	2 %	Skovgaard (2005)
	10 %	Chatton (1920)[*]
Blastodinium navicula	9 %	Skovgaard (2005)
	20-30 %	Chatton (1920)[*]
Blastodinium oviforme	0-4 %	Skovgaard & Saiz (2006)
Blastodinium sp.	<1 %	Ianora et al. (1990)
	0.4 %	Ianora et al. (1987)
	6.6 %	Øresland (1991)
Blastodinium spinulosum	0.4-0.9 %	Chatton (1920)[*]
Blastodinium spp.	33 %	Alves-de-Souza et al. (2011)
	51 %	Alves-de-Souza et al. (2011)
	<2 %	Alves-de-Souza et al. (2011)
	<2 %	Alves-de-Souza et al. (2011)
Ellobiopsis chattoni	0.56-1 %	Fahmi & Hussain (2003)
	15 %	Gómez, López-García, et al. (2009)
	4.6-8.6 %	Artüz (2016)
	15 %	Timofeev (2002)
	0.3 %	Marshall et al. (1934)
	8.3 %	Santhakumari & Saraswathy (1979)
	26 %	Wickstead (1963)
	5-22.4 %	Hoffman & Yancey (1966)
Ellobiopsis sp.	6.8 %	Albaina & Irigoien (2006)
	0.06-0.09 %	Walkusz & Rolbiecki (2007)
Ichthyophonus sp.	0.1 %	Torgersen et al. (2002)
Syndinium sp.	0-13 %	Ianora et al. (1990)
	0-30 %	Ianora et al. (1987)
	0-28.5 %	Kimmerer & McKinnon (1990)
	12 %	Marshall et al. (1934)
Syndinium turbo	0-7 %	Skovgaard & Saiz (2006)
Vampyrophrya pelagica	0-100 %	Ohtsuka et al. (2004)
	0-100 %	Grimes & Bradbury (1992)

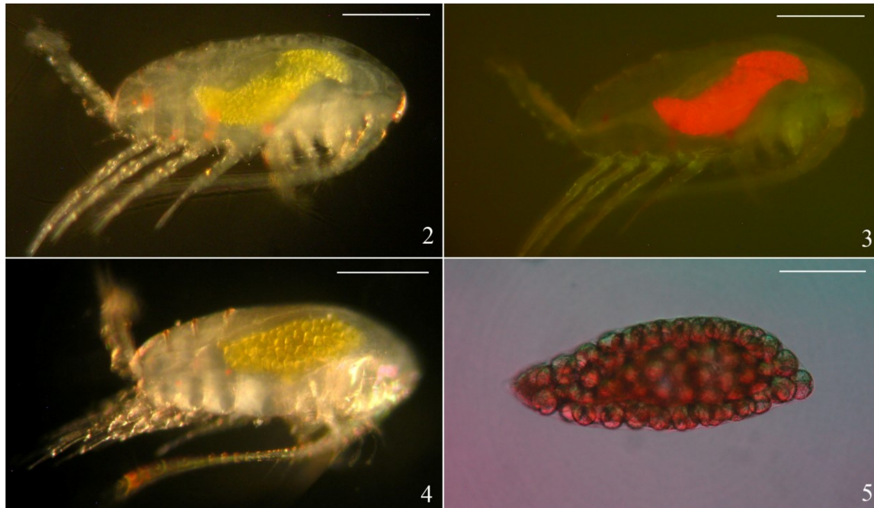


Figure 2: Paracalanus parvus *infected with* Blastodinium spp.. ***2:*** B. contortum *infecting the host.* ***3:*** *Same specimen as 2, with epifluorescence microscopy.* ***4:*** B. crassum *infecting the host.* ***5*** B. crassum *removed from the host. All pictures were taken from Coats et al. (2008).*

Despite being large (up to more than 1000 μm, Fields et al., 2015, fig. 2C) and multicellular, the *Blastodinium* belong to the dinoflagellates, albeit with a peculiar life cycle. The exact infection process has never been verified, but it has been hypothesized that infection starts when the dinospores of *Blastodinium* are ingested by a copepod in the nauplius or copepodid stage (Chatton, 1920; Skovgaard et al., 2012). The spore then develops into a large multicellular structure called a trophont, containing several specialized cells. The trophont produces dinospores, which presumably exit the copepod in fecal pellets (Fields et al., 2015). The dinospore stage is the only stage where *Blastodinium* morphologically resembles a free-living dinoflagellate, with two flagella and thecal plates (Chatton, 1920; Skovgaard et al., 2007). Most species have a functional chloroplast in the sporocyte (spore-producing) stage (Skovgaard et al., 2012; Soyer, 1970), which might contribute to acquiring energy for the parasite when inside the copepod gut (Shields, 1994). Little is known about the dinospore stage, including how long it can live, whether it actively seeks out its host, and how it infects new hosts.

Infection with *Blastodinium* sterilizes the host, possibly due to starvation (Fields et al., 2015; Skovgaard, 2005). Respiration rates of infected individuals were significantly reduced, and they had no measurable ingestion when grown in culture (Fields et al., 2015). Infection can also increase the mortality rates in host populations (Skovgaard, 2005). Chatton (1920) suggested that the reduced fecundity could be due to the sheer size of the parasite, filling and expanding the gut, and thus disrupting the egg-producing structures. He also noted that infection seemed to inhibit the development of males, resulting in only finding parasitized females.

Ellobiopsis

Ellobiopsis is a genus with three species belonging to the Ellobiopsea, which are alveolates that likely branches out as a basal group of dinoflagellates (Gómez, López-García, et al., 2009; Gómez, 2012). They are ectoparasites, attaching to the appendages—mostly antennae—of copepods (Figure 3), using a sucking device that penetrates the cuticle of the host (Albaina & Irigoien, 2006). A single copepod individual can be parasitized by many *Ellobiopsis* at once (Albaina & Irigoien, 2006). Since it is relatively large and attached to the outside of the copepod, it is one of the easier parasites to detect visually in zooplankton samples. Albaina & Irigoien (2006) found that infection with *Elllobiopsis sp.* effectively castrated female *Calanus helgolandicus*, meaning that population-level effects of *Ellobiopsis* could be profound.

Figure 3: Ellobiopsis sp. *infecting* Calanus sp.. *Specimen from live samples from the Oslofjord.*

Syndinium

Syndinium are parasitoids (Box 2) with a single confirmed species, *S. turbo*, and belong to the parasitic dinoflagellate order Syndiniales (Figures 1 (41) and 4; Skovgaard et al., 2005). *Syndinium* develops inside the prosome of the copepod, producing a large number of spores, before bursting out and killing the copepod when spores are mature (Ianora et al., 1990, misidentified as *Atelodinium sp.* in the study; Skovgaard et al., 2005). This way of developing and spreading inevitably kills the host, and the entire development takes less than 2 hours Jepps (1937). The pathogenicity and short window of infection and development mean that *Syndinium* may have devastating effects on copepod populations. It also means that detecting *Syndinium* visually in samples is challenging and that prevalence estimates from visual methods, representing a snapshot of the population, may not give the full picture of its impact.

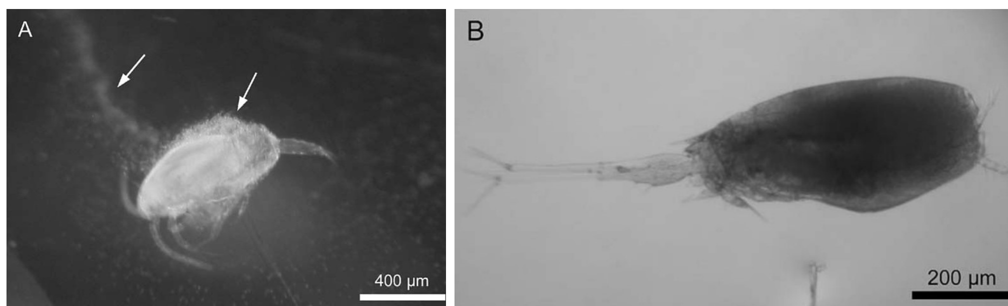


Figure 4: Syndinium sp. *from the Mediterranean. A:* Syndinium *spores leaving its dead host,* Paracalanus parvus. *B:* Syndinium *infecting* Corycaeus sp.. *Images from Skovgaard et al. (2005).*

Ichthyophonus sp.

Ichthyophonus sp. is an elusive copepod parasite investigated by Chatton (1920) with very few observations since. Members of genus *Ichthyophonus* (formerly *Ichthyosporidium*) are mainly fish parasites, and the *Ichthyophonus sp.* in copepods have been assigned to the genus based on morphological similarities (Jepps, 1937). The parasite is a yellow, hyphae-like growth filling the prosome of the copepod, giving it a conspicuous color (Figure 5; Jepps, 1937; Torgersen et al., 2002). To my knowledge, no DNA sequence exists to investigate the molecular taxonomy of the parasite.

Figure 5: Ichthyophonus sp. *infecting* Calanus sp.. *Specimen from live samples from Drøbak.*

Torgersen et al. (2002) frequently observed *Ichthyophonus sp.* parasitizing *Calanus spp.* in the Oslofjord in the summer. They noted that infected copepods aggregated in the surface and that this coupled with the conspicuous coloration increased risk of predation. We have consistently collected this parasite ourselves by blindly towing plankton nets in the surface in the Oslofjord, supporting the observations of Torgersen et al. (2002) (Eliassen et al., in prep.). Torgersen et al. (2002) discussed (as also proposed by Jepps, 1937) that due to the behavioral modification and increased predation risk, it is likely that *Ichthyophonus sp.* uses *Calanus spp.* as an intermediate host (Box 3), with a fish as the final host.

Vampyrophrya pelagica

Vampyrophrya pelagica is an apostome ciliate and an ectoparasite on copepods (Ohtsuka et al., 2004). *V. pelagica* attaches to the outside of the copepod and is relatively harmless until the copepod suffers mechanical damage by an external force (Grimes & Bradbury, 1992). When the host is damaged, the parasite can enter the copepod interior through cracks in the exoskeleton, where they consume the copepod and grow. *V. pelagica* can be present with incredibly high prevalence (Table 3), and thus it may contribute significantly to copepod mortality. However, the high prevalence may indicate that the parasite is harmless to most copepods, as it otherwise could wipe out the host population (see Anderson & May, 1981).

Paradinium

Paradinium is a genus of copepod parasites belonging to the clade Rhizaria and contains one confirmed species, *P. poucheti* (Skovgaard & Daugbjerg, 2008). *Paradinium* grows inside the host before exiting through the anus, making a structure on the outside of the copepod for spreading its spores (Figures 1 (42) and 6 A; Jepps, 1937). Thus it is primarily an endoparasite but looks like an ectoparasite in later stages. The effects of *Paradinium* infection are not well investigated, but Jepps (1937) states that it castrates its *Calanus* host by destroying the gonad tissue. Two recent Ph.D. dissertations support this, showing a negative correlation between parasitized and egg-carrying *Oithona similis* (Briseño-Avena, 2015; Orenstein, 2018).

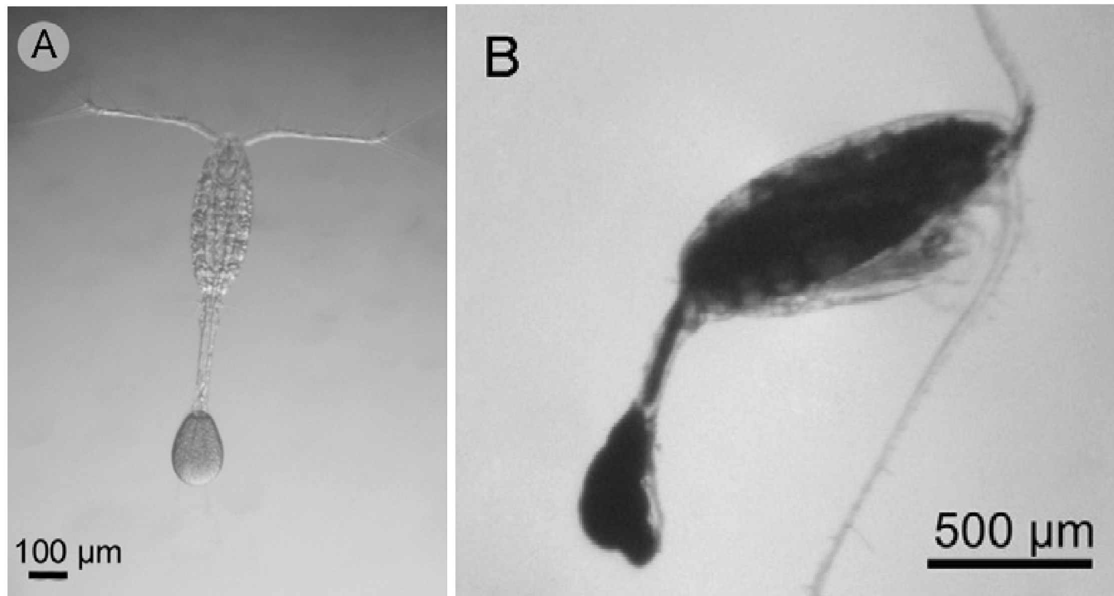


Figure 6: A: Paradinium poucheti *infecting* Oithona similis. *B: "RP parasite" infecting a calanoid copepod. Images from Skovgaard & Daugbjerg (2008)*

Other parasites

There are several other protistan parasites of copepods that I have not summarized above (Table 1), either because little is known about them or to limit the scope of this thesis. Some of the other parasites are briefly introduced here. *Chythriodinium* and *Dissodinium* parasitize copepod eggs, with lethal outcome (Gómez, Moreira, et al., 2009). A "Red plasmodial parasite" (RP parasite, Figure 6 B) resembling *Paradinium* with unknown effect was identified by Jepps (1937) and subsequently investigated by Skovgaard & Daugbjerg (2008). Several gregarines (Apicomplexa) are known to inhabit the gut of copepods, with no apparent effect (Jepps, 1937; Sano et al., 2016). Similarly, many suctorian ciliates are ectobionts of copepods and are generally assumed to be commensals (Fernandez-Leborans & Tato-Porto, 2000b; see Box 1). There are probably many parasites left to discover, and there is undoubtedly a lot of work to do before we understand the full impact of even the known parasites.

Two main questions

Since parasites of pelagic copepods—even the more well-studied taxa—have been so little studied since their discovery about 100 years ago, many fundamental aspects have yet to be covered. In this thesis, I have investigated two fundamental questions about the parasites and two more minor questions. The questions and goals, an outline of how I have investigated them, and my expectations are briefly summarized below.

1. **When** are parasites present?

   I have used metabarcoding data from a time series in the Oslofjord to investigate seasonal patterns of parasite occurrence. Copepod occurrence in the Oslofjord is seasonal, and I expect parasite occurrence to be seasonal as well.

2. **Where** are the parasites?

   I have used our metabarcoding data from the Oslofjord and metabarcoding data from the BioMarKs project in European waters to investigate spatial patterns of parasite occurrence. I expect that host distribution and abiotic factors influence parasite distribution.

In addition to the two main questions, I attempt to answer two more minor questions for this thesis.

3. **Who** are the parasites?

   We collected four different copepod parasites, which were sequenced together with the time series samples (see Methods). Obtaining sequences of these parasites aided in detecting them in our samples and in the BioMarKs data, as reference sequences for copepod parasites generally are scarce (see Discussion).

4. Is metabarcoding a good tool for studying parasitism in copepods?

   I discuss how good metabarcoding is for detecting parasites and patterns in occurrence in our data. I discuss advantages and limitations, challenges, and how study design and future research can solve some of the current issues (see Discussion). We used two different primer sets for the metabarcoding, and I evaluate this approach.

Below, I summarize what we already know about the two main questions, "Where" and "When." Then I outline how I have investigated the questions in this thesis.

When?

The occurrence and abundance of copepods is highly seasonal in temperate areas (Kiørboe & Nielsen, 1994). Copepod seasonal patterns are closely linked to the phytoplankton and have peak secondary production during the spring and autumn blooms when food availability is largest (Colebrook, 1979; Zervoudaki et al., 2009). In the Skagerrak area, the total copepod biomass has a unimodal distribution, with peak biomass around July (Kiørboe & Nielsen, 1994). In general, when host abundance is seasonal, the prevalence of the parasite also has seasonal cycles (Anderson & May, 1981). Based on this, I expect the parasites of Copepods to have seasonal patterns closely tied to those of their copepod hosts.

Only a handful of studies include time series with prevalence measures of parasites on copepods, and many of these are from areas without strong seasonality. The most comprehensive study of this kind was done by Skovgaard & Saiz (2006) in the Mediterranean. In the study, both several species of *Blastodinium* and *Syndinium turbo* had the highest prevalence in late summer to early winter, with the species of *Blastodinium* being more or less absent for the rest of the year. For many species, the peak prevalence occurred when the host's densities were highest (with a slight time lag), indicating that infection was density-dependent.

A handful of other studies have reported seasonal patterns in parasite occurrence. Marshall et al. (1934) noted that *Syndinium sp.* parasitizing *Calanus finmarchicus* had peak abundances in the fall and winter, while *Ellobiopsis chattoni* was most common in the summer. Ianora et al. (1990, 1987) found that *Syndinium sp.* had the highest prevalences in early autumn, following the peak densities of its host *Paracalanus parvus*. In the more seasonal seas of Scotland, Jepps (1937) reported that *Syndinium sp.* on *Calanus finmarchicus* was most abundant in January and not present after May. *Paradinium poucheti*, on the other hand, started appearing around March and continued throughout summer and autumn. Ohtsuka et al. (2004) reported that *Vampyrophrya pelagica* occurred most commonly (with prevalence up to 100%!) in autumn and winter in the North West Pacific Ocean, being completely absent in summer. Interestingly, Grimes & Bradbury (1992) showed no such pattern from the West Atlantic Ocean on the same species. For *Ichthyophonus sp.*, Torgersen et al. (2002) noted that the parasite had peak abundances in late summer and was completely absent between October and May. Additionally, the occurrence of the parasite in the surface was seemingly dependent on weather and tides.

To investigate the seasonality of the copepod parasites, we have taken zooplankton samples from the Oslofjord, Norway, in a time series of 1 year in 2020 (see Methods, p. 17). The original plan was to count the samples and visually detect parasitized individuals, but finding them in our samples proved difficult (see above, p. 4). Because of this, we turned to metabarcoding: sequencing the genetic material in the samples to infer what organisms were present. Metabarcoding allows for detecting diversity that is hard to see and a suitable candidate method for studying copepod parasites. To better detect parasites, we used one primer set that was made especially for detecting protists in samples with a lot of metazoan biomass (Bass & del Campo, 2020; see Methods).

Where?

Most of the samples that Chatton (1920) used in his work were from the Mediterranean Sea. Since then, many of the studies of the parasites has been in the same area (Alves-de-Souza et al., 2011; e.g., Skovgaard, 2005; Skovgaard & Saiz, 2006). However, parasites have been found all over the world's oceans (Figure 7, Table 4), and may very well live anywhere hosts are available.

For *Blastodinium* specifically, distribution seems to follow latitude. The species with chloroplasts are seemingly restricted to areas near the equator, while the colorless *B. hyalinum* is found at all latitudes (Skovgaard et al., 2012). Skovgaard et al. (2012) speculates that the species with chloroplasts use photosynthesis for part of their energy acquisition, using nutrients taken from the host to overcome nutrient limitations of the oligotrophic waters of, e.g., the Mediterranean.

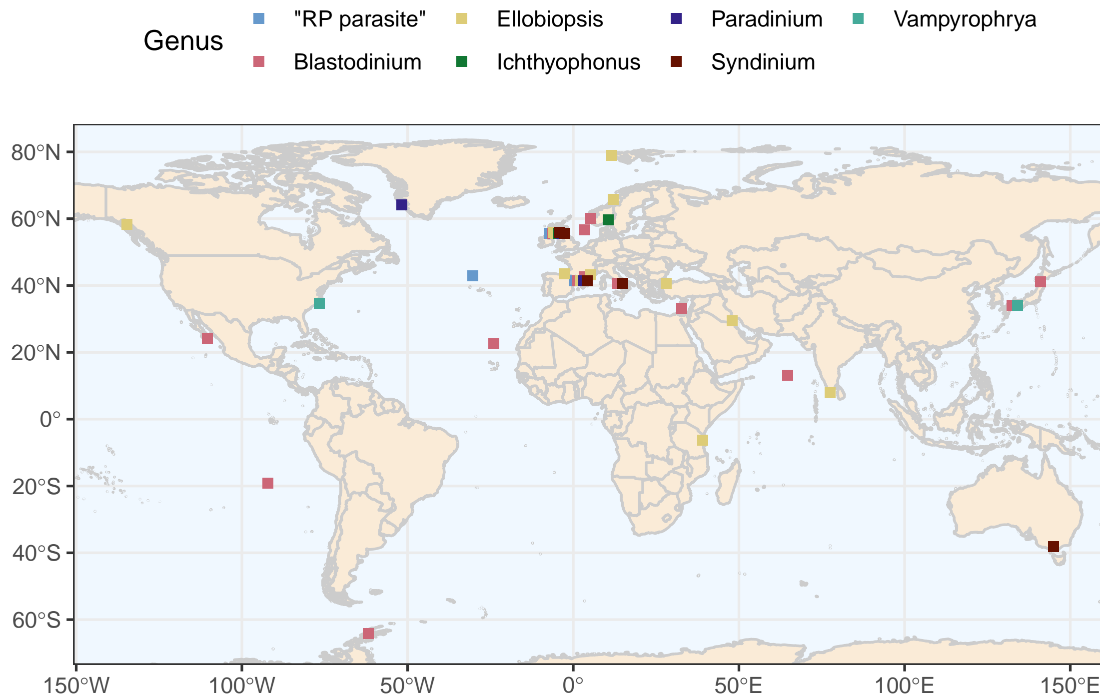


Figure 7: Worldwide distribution of parasites on copepods from the literature. References for each study is included in Table 4.

The other parasites have no clear pattern in distribution. From the map in Figure 7, there might seem to be a higher occurrence of parasites in the coastal regions in general and in the coasts of Europe in particular. This pattern, however, is probably due to the larger number of studies conducted in these regions, and no clear pattern exists in the data available so far.

Table 4: Location and hosts of parasites reported in the literature. Since some research papers were difficult to obtain, and Chatton wrote in French, some data is from secondary sources: [*]data from Skovgaard et al. (2012). [†]data from Horiguchi et al. (2006).

Parasite	Location	References	Parasite	Location	References
"RP parasite"	Clyde Sea, Scotland	Jepps (1937)	*Blastodinium pruvoti*	Western Mediterranean Sea	Chatton (1920)[*]
	North Atlantic Ocean	Skovgaard & Daugbjerg (2008)	*Blastodinium sp.*	Gerlache Strait	Øresland (1991)
	Port Olímpic, Barcelona	Skovgaard & Daugbjerg (2008)		Gulf of Naples	Ianora et al. (1990)
Blastodinium apsteini	Arabian Sea	Sewell (1951)[*]		Gulf of Naples	Ianora et al. (1987)
	Western Mediterranean Sea	Chatton (1920)[*]		Port Olímpic, Barcelona	Skovgaard & Saiz (2006)
Blastodinium chattoni	Arabian Sea	Sewell (1951)[*]		Seto Inland Sea, Japan	Horiguchi et al. (2006)
Blastodinium contortum	Arabian Sea	Sewell (1951)[*]	*Blastodinium spinulosum*	Arabian Sea	Sewell (1951)[*]
	Gulf of California	Coats et al. (2008)		Mutsu Bay, Japan	Kofoid (1931)[†]
	Port Olímpic, Barcelona	Skovgaard & Saiz (2006)		Port Olímpic, Barcelona	Skovgaard & Saiz (2006)
	Port Olímpic, Barcelona	Skovgaard et al. (2007)		Western Mediterranean Sea	Chatton (1920)[*]
	South East Pacific Ocean	Pasternak et al. (1984)[*]	*Blastodinium spp.*	Eastern Mediterranean Sea	Alves-de-Souza et al. (2011)
	Western Mediterranean Sea	Chatton (1920)[*]	*Ellobiopsis chattoni*	Auke Bay, Alaska	Hoffman & Yancey (1966)
Blastodinium crassum	Arabian Sea	Sewell (1951)[*]		Bay of Marseille	Gómez, López-García, et al. (2009)
	Gulf of California	Coats et al. (2008)		Cape Comorin, India	Santhakumari & Saraswathy (1979)
	Western Mediterranean Sea	Chatton (1920)[*]		Clyde Sea, Scotland	Jepps (1937)
Blastodinium elongatum	Western Mediterranean Sea	Chatton (1920)[*]		Kuwait Bay	Fahmi & Hussain (2003)
Blastodinium galatheanum	Central Atlantic Ocean	Skovgaard & Salomonsen (2009)		Loch Striven, Scotland	Marshall et al. (1934)
Blastodinium hyalinum	Arabian Sea	Sewell (1951)[*]		Norwegian Sea	Timofeev (2002)
	Austevoll, Norway	Fields et al. (2015)		Sea of Marmara	Artüz (2016)
	Clyde Sea, Scotland	Jepps (1937)		Zanzibar Channel	Wickstead (1963)
	North Sea	Vane (1952)[*]	*Ellobiopsis fagei*	Port Olímpic, Barcelona	Skovgaard & Saiz (2006)
	North Sea	Cattley (1948)	*Ellobiopsis sp.*	Bay of Biscay	Albaina & Irigoien (2006)
	Port Olímpic, Barcelona	Skovgaard & Saiz (2006)		Kongsfjorden, Spitsbergen	Walkusz & Rolbiecki (2007)
	Western Mediterranean Sea	Chatton (1920)[*]		Port Olímpic, Barcelona	Skovgaard & Saiz (2006)
Blastodinium mangini	Arabian Sea	Sewell (1951)[*]	*Ichthyophonus sp.*	Clyde Sea, Scotland	Jepps (1937)
	Port Olímpic, Barcelona	Skovgaard & Saiz (2006)		Oslofjorden, Norway	Torgersen et al. (2002)
	Port Olímpic, Barcelona	Skovgaard (2005)	*Paradinium poucheti*	Clyde Sea, Scotland	Jepps (1937)
	Western Mediterranean Sea	Chatton (1920)[*]		Godthåbsfjord, Greenland	Skovgaard & Daugbjerg (2008)
Blastodinium mangini var. oncaea	Western Mediterranean Sea	Chatton (1920)[*]		Port Olímpic, Barcelona	Skovgaard & Daugbjerg (2008)
Blastodinium navicula	Arabian Sea	Sewell (1951)[*]	*Syndinium sp.*	Clyde Sea, Scotland	Jepps (1937)
	Port Olímpic, Barcelona	Skovgaard & Saiz (2006)		Gulf of Naples	Ianora et al. (1990)
	Port Olímpic, Barcelona	Skovgaard (2005)		Gulf of Naples	Ianora et al. (1987)
	Port Olímpic, Barcelona	Skovgaard et al. (2007)		Loch Striven, Scotland	Marshall et al. (1934)
	Western Mediterranean Sea	Chatton (1920)[*]		Port Olímpic, Barcelona	Skovgaard & Saiz (2006)
Blastodinium oviforme	Arabian Sea	Sewell (1951)[*]		Port Philip Bay, Australia	Kimmerer & McKinnon (1990)
	Port Olímpic, Barcelona	Skovgaard & Saiz (2006)	*Syndinium turbo*	Port Olímpic, Barcelona	Skovgaard & Saiz (2006)
	Western Mediterranean Sea	Chatton (1920)[*]		Port Olímpic, Barcelona	Skovgaard et al. (2005)
Blastodinium pruvoti	Arabian Sea	Sewell (1951)[*]	*Vampyrophrya pelagica*	Beaufort, USA	Grimes & Bradbury (1992)
	Port Olímpic, Barcelona	Skovgaard & Saiz (2006)		Seto Inland Sea, Japan	Ohtsuka et al. (2004)

A growing amount of molecular data is available from pelagic studies, which have not been analyzed specifically with parasites of copepods in mind. The BioMarKs project (e.g., Logares et al., 2012) took samples from the surface, deep chlorophyll maximum (DCM) and sediments of six stations in European coastal waters (Figure 8). The Tara Oceans expedition (Karsenti et al., 2011) sampled plankton in a 3-year worldwide cruise from a total of 210 sampling stations (see Sunagawa et al., 2020 for a review of the project and its impact). DNA from the samples from both projects was sequenced, and the sequences are publicly available. Several studies have already analyzed this data (e.g., Logares et al., 2012; López-Escardó et al., 2018 (BioMarKs); and de Vargas et al., 2015; Lima-Mendez et al., 2015 (Tara)), but parasite-copepod relationships, among countless other things, have yet to be explored.

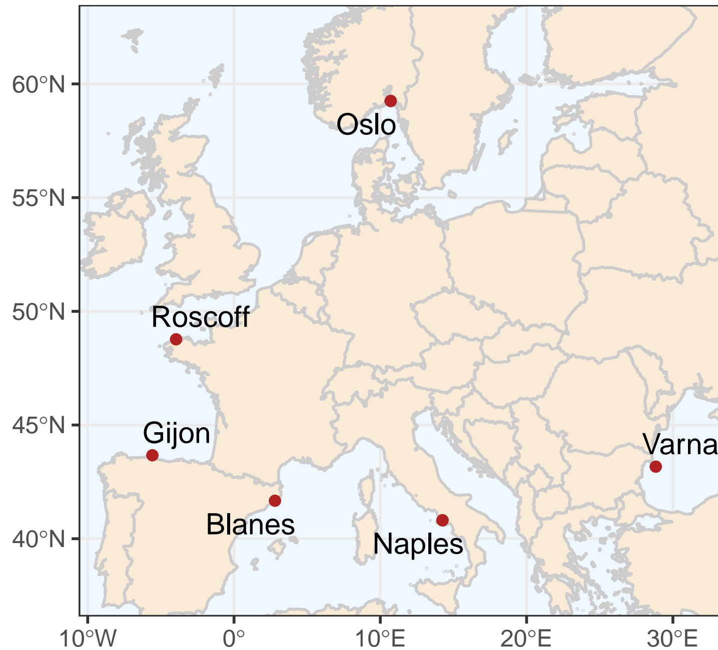


Figure 8: Locations of the sampling sites from the BioMarks project.

To answer the "where," I have looked for copepod parasites in the sequence data from the BioMarKs project, together with our own data from the Oslofjord. Combining both the available information in databases and our sequencing results, I investigated the spatial patterns in parasite occurrence across the 6 locations in European waters and along our transect in the Oslofjord. Parasite distribution is tied to host distribution (Hance et al., 2007), so I expect factors influencing copepod community composition (e.g., depth, see Methods, p. 17) to affect parasite composition as well. Furthermore, some parasites may have distributional patterns due to other factors, like photosynthetic vs. non-photosynthetic *Blastodinium* having a latitudinal pattern (Skovgaard et al., 2012).

The data from the Tara oceans project is so massive that it is beyond the scope of this thesis, but I have done some preliminary processing, searching for parasite genera. Comparing the results of that processing with the findings of copepod parasites in the literature (Figure 9 vs. Figure 7) serve as an excellent motivation for further investigation of existing metabarcoding data sets. The comparison of spatial data from 100 years of copepod parasite research against a single (albeit very impressive) modern project gives a positive outlook for the future of copepod parasite studies with the emergence of new methods.

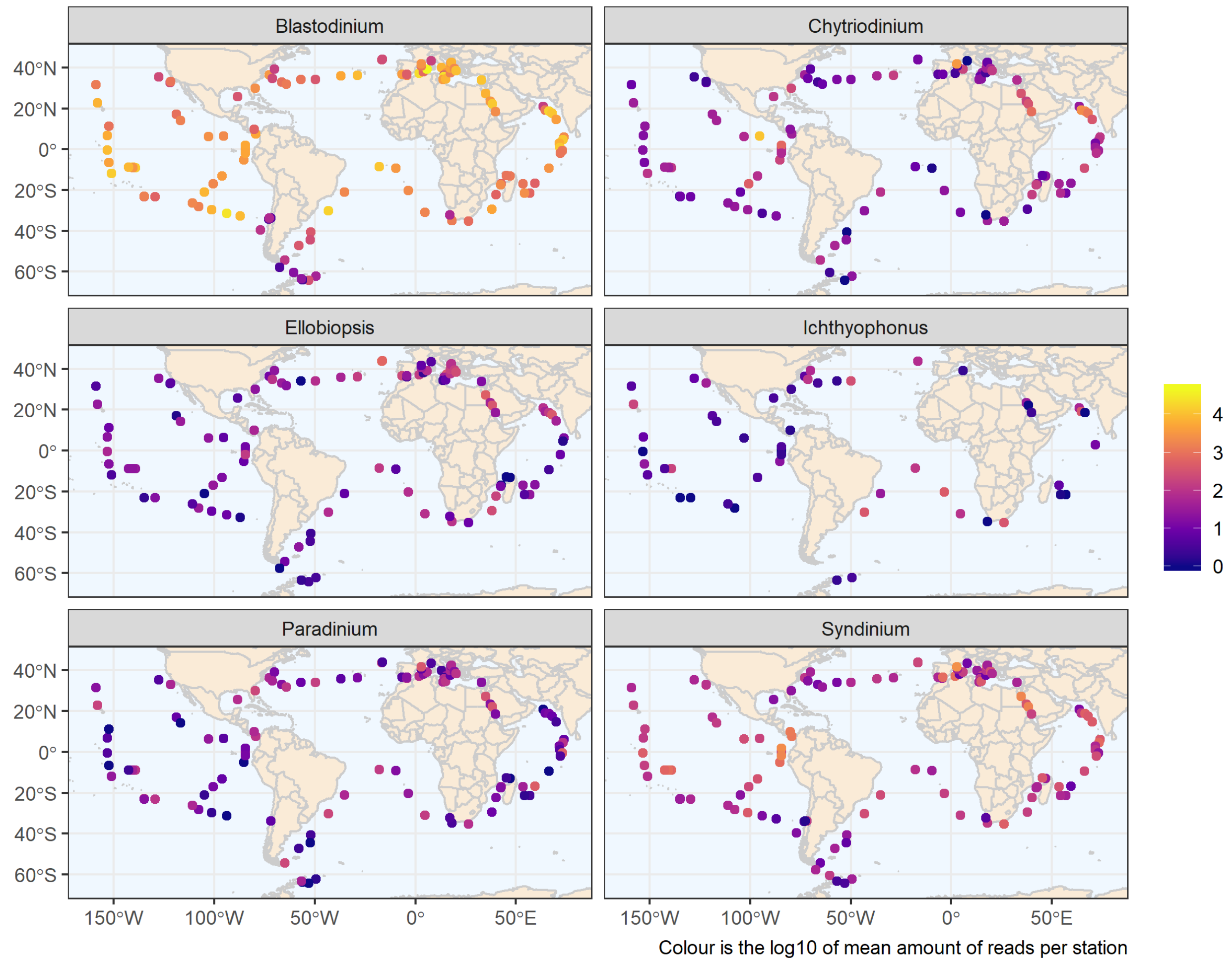


Figure 9: Distribution of known copepod parasites from the Tara Oceans data set. Parasites were present at all stations sampled, although not every parasite was present everywhere. Data from de Vargas et al. (2017), see Methods p. 31.

Methods

In this section, I describe the methods for field sampling, processing and metabarcoding of zooplankton samples. A graphic summary of the entire process is shown in Figure 10. Sampling and molecular lab work was a collective effort with Ph.D. candidate Lasse Eliassen, and I will describe those procedures using "we." Similarly, the data will be referred to as "our data." For the bioinformatics and data analysis, which were done mostly by myself, I will use "I."

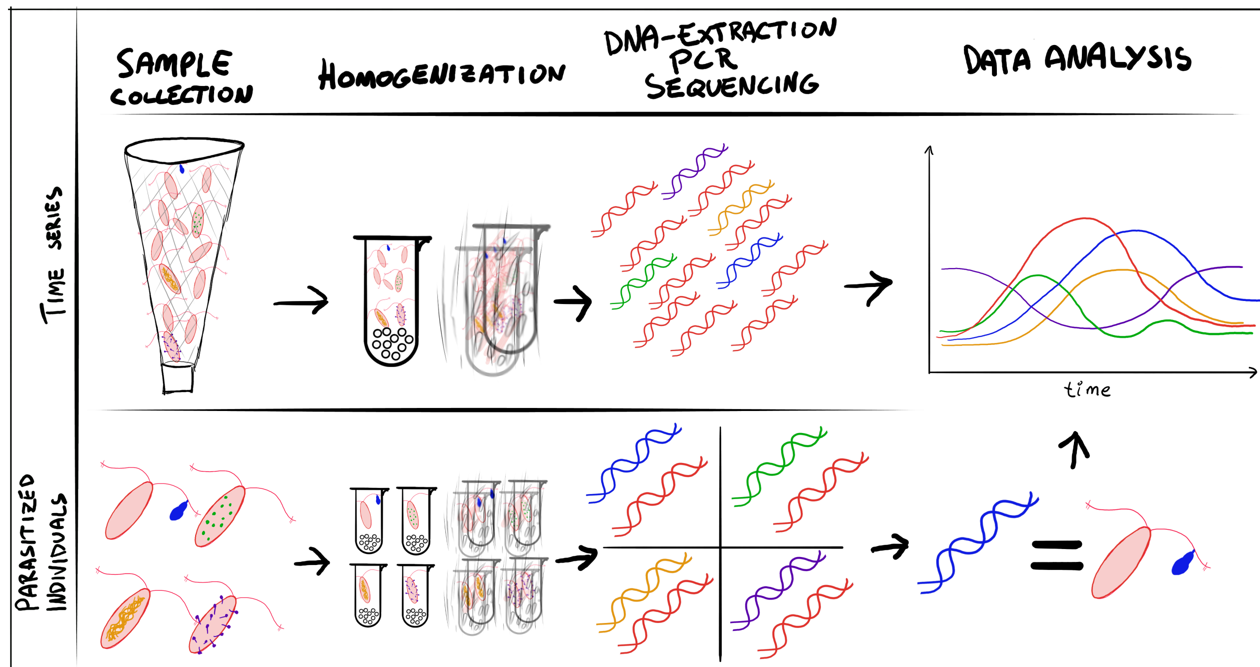


*Figure 10: Graphic summary of the methods. **Top:** Samples were collected in a time series from the Oslofjord with a WP2 net. Homogenization by bead beating, DNA-extraction, PCR and sequencing was performed on a bulk sample. These samples were analyzed to uncover spatial and seasonal patterns in the Oslofjord. **Bottom:** Parasitized individuals were isolated from live samples (individual parasite samples). Homogenization, DNA extraction, PCR and sequencing was performed on each parasite separately. The information from the sequencing was used to identify parasite sequences, which were used for subsequent analysis.*

The study area

The Oslofjord is a fjord in southern Norway with the capital Oslo located in the inner parts. The fjord is divided into the inner and outer Oslofjord by the shallow Drøbak sill, which has a maximum depth of 19.5 m and limits the supply of oceanic water to the inner fjord (Baalsrud & Magnusson, 2002). Water flowing from rivers into the fjord creates a strong stratification, with a variable mixed layer on top which is brackish in summer, and a deeper layer with relatively stable temperature and salinity (Baalsrud & Magnusson, 2002).

The Oslofjord is strongly seasonal, and especially in the mixed layer there are large yearly variations in temperature, salinity, light availability and nutrients (Lundsør et al., 2020). Life in the fjord is characterized by at least one phytoplankton bloom: Nutrients become available in the upper layer through mixing of the stratified layers during fall and winter, and when light availability increases in the spring—around March—the

phytoplankton grows rapidly until most of the available nutrients are used up and growth declines (Gran-Stadniczeñko et al., 2019). A second and third bloom may occur: in May-June due to nutrient supply from river run-offs, and in August-September due to nutrient supply from deep waters because of wind mixing (Gran-Stadniczeñko et al., 2019). The seasonality of copepods in the zooplankton is linked with that of the phytoplankton, with peak abundances corresponding to the phytoplankton abundances with a time-lag (Kiørboe, 1997; Kiørboe & Nielsen, 1994).

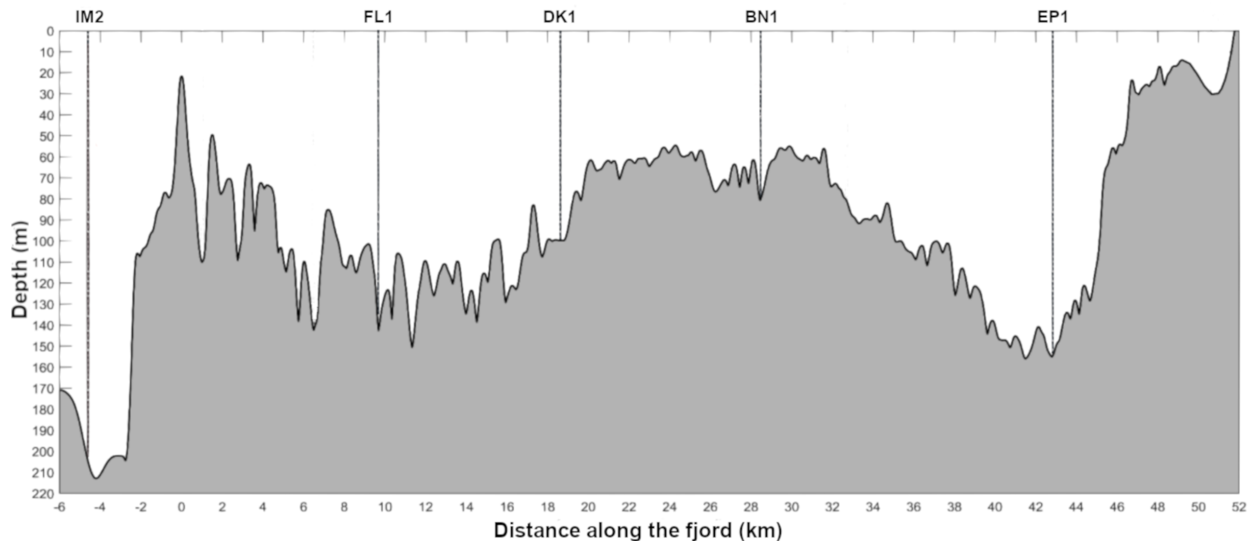


Figure 11: Depth profile of the Oslofjord with our sampling stations marked. Image modified from Fagrådet for vann og-avløpsteknisk samarbeid i indre Oslofjord, http://www.indre-oslofjord.no

Conditions in the Oslofjord vary considerably from the outer to inner parts. Outside of the Drøbak sill, the fjord is deeper, and there is more water exchange with the rest of the ocean (Baalsrud & Magnusson, 2002). These deep, oceanic-like waters host the large, carnivorous copepods *Chiridius armatus* and *Paraeuchaeta norvegica*, which reside in the deep during daytime to avoid predators, and migrates to shallower waters at night to feed (Schøyen & Kaartvedt, 2004; Skarra & Kaartvedt, 2003). Other, smaller copepods reside mostly in the shallower waters with limited vertical migration (Lagadeuc et al., 1997), and are expected to be present throughout the fjord. In the inner part of the inner fjord, there is a second sill of around 50 m separating the innermost Bunnefjorden from the rest of the inner fjord (Figure 11). Bunnefjorden is sheltered, and deep water is typically only exchanged every 2-3 years, resulting in long periods of hypoxia in the deep layers (Solberg et al., 2015). The contrasts between the inner and outer parts may be reflected in the parasite distribution in the fjord, as different hosts will be available in different places.

Field sampling on RV Trygve Braarud

The sampling was done in junction with the Ph.D. work of Lasse Eliassen. This thesis describes the sampling and metabarcoding analysis of zooplankton tows in 2020. For the project, we have also taken water samples on filters to investigate the free-living diversity of parasites. The time series will continue into 2021, and all samples will be analyzed both with metabarcoding and traditional plankton counts in the future.

We collected zooplankton samples and environmental data on RV Trygve Braarud from the Oslofjord each month from January to December 2020. There were no samples taken in April because the research vessel was inoperative due to the COVID-19 pandemic. We took samples from 5 different stations each month, in a transect from the innermost parts in Bunnefjorden to just outside of the Drøbak sill (Table 5, Figure 12). The Bunnefjorden locality was not sampled in January, so the total number of samples taken amounts to 54 (Table 5).

Table 5: Stations that were sampled in the time series. Station IDs correspond to stations used in environmental monitoring series.

ID	Name	Longitude	Latitude	Maximum depth (m)	Number of samples
IM2	Elle	10.6282	59.6220	200	11
FL1	Spro	10.5746	59.7540	160	11
DK1	Steilene	10.5800	59.8100	100	11
BN1	Lysakerfjorden	10.6468	59.8806	80	11
EP1	Bunnefjorden	10.7229	59.7869	150	10

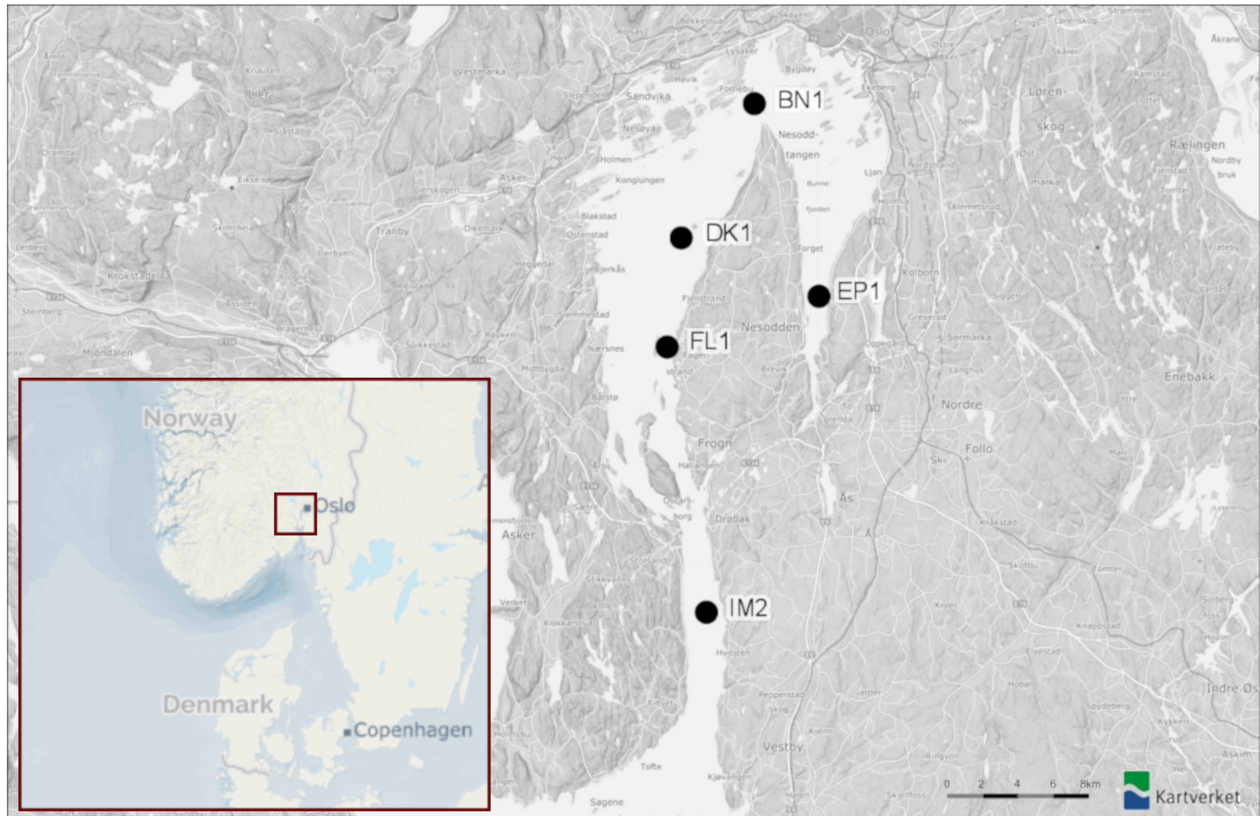


Figure 12: Location of the 5 sample sites. EP1-IM2 represent a transect from inner-outer fjord. Map from Kartverket via norgeskart.no

At each sampling site, we collected samples with a single vertical tow of the entire water column using a WP2 net with a diameter of 55cm and a 200µm mesh, i.e. excluding all organisms smaller than 200µm. We split the samples on the boat using a plankton splitter. Half of each sample was preserved in 96% ethanol for DNA analyses, while the rest of the samples were taken for plankton counts, preserved either in acidic Lugol's iodine or by freezing. Only the ethanol samples were used for the DNA analyses described in this thesis. The samples preserved for counting have not yet been processed.

We also collected and isolated parasitized copepod individuals with 4 different parasites for sequencing (Figure 13). These parasites were conspicuous enough to be easily noted in live zooplankton samples and were obtained opportunistically during our cruises in 2020.

We measured environmental variables at each station each month using a CTD rosette. A CTD SBE 9 (Seabird electronics, Washington, USA) was used from January through June, and an STD (SAIV, Bergen,

Norway) was used from August through December due to maintenance on the SBE. The STD measured conductivity, temperature and oxygen content, while the CTD also recorded fluorescence and turbidity of the water column. Environmental data from July is missing from our cruises, but was obtained for similar sites for similar dates from Norwegian Institute for Water Research (NIVA): for the stations EP1, BN1 and DK1, data was obtained from a NIVA-cruise at those stations on the same date as our July cruise. For FL1 in July, the aforementioned DK1 data was used, and for IM2, data from NIVA station OF-7 (outer Oslofjord) from July 5 (3 days after our cruise) was used.

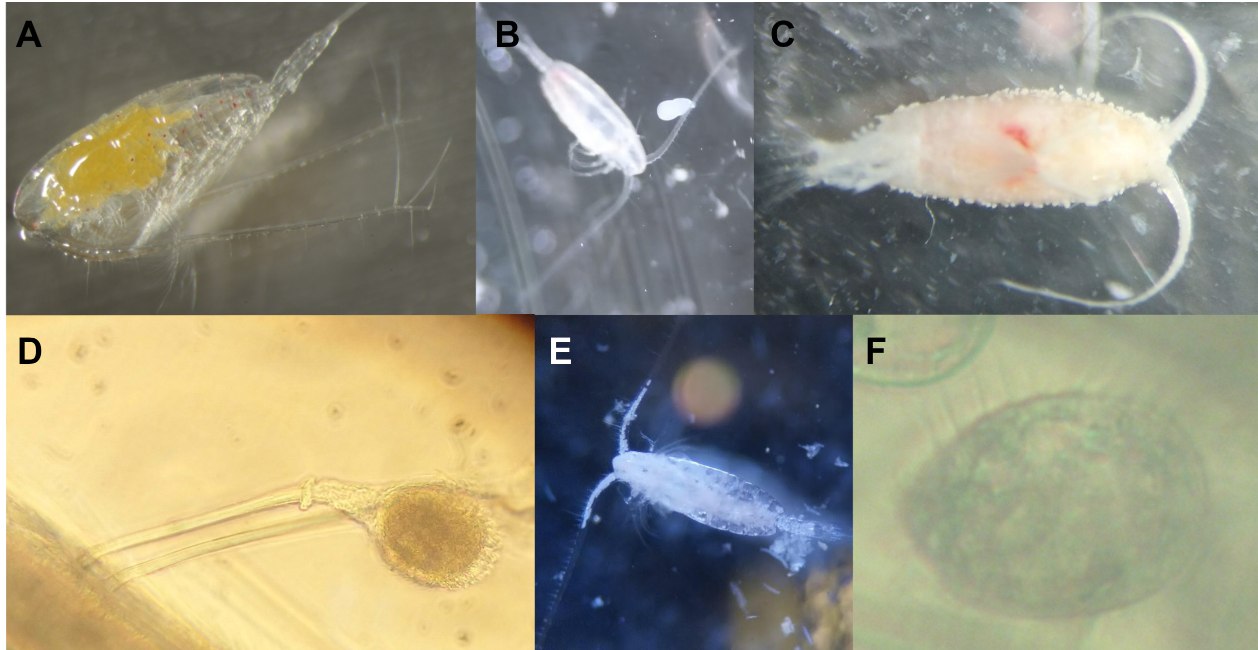


Figure 13: Parasites and hosts that were collected for sequencing. ***A:*** Ichthyophonus sp. *infecting* Calanus sp. *(Sample 56).* ***B:*** Ellobiopsis sp. *infecting* Calanus sp. *(Sample 2).* ***C:*** *Unknown ciliate infecting* Chiridius sp. *(Sample 11).* ***D:*** *Microscope image of the parasite in C.* ***E:*** Calanus sp. *infected with an unknown parasite. At the time of the picture, the copepod was alive. The myriads of parasite cells were moving around within the host (Sample 17).* ***F:*** *Microscope image of the parasite in E.*

Homogenization and DNA extraction

Homogenization of plankton samples

Since the amount of tissue used in DNA extraction is much lower than the total biomass in the samples, we had to extract DNA from a small sub-sample. The subsample should be representative of the full sample, so it is important to thoroughly homogenize before subsampling (van der Loos & Nijland, 2020). In this step we had to process the samples one by one, so to avoid bias we randomized sample order by a pseudo-random number generator.

First, we split each sample with a plankton splitter and filtered away the ethanol so that the whole remaining sample could fit in the 2ml tubes used for homogenization. The fraction of sample used for homogenization varied from the whole sample in the winter samples, to 1/32 in the samples from May and June. Before homogenization we removed all large non-copepod animals, that may have otherwise dominated the biomass in our sub-sample (see Deagle et al., 2018). The removed animals were mostly chaetognaths and the annelid *Tomopteris*, and occasionally a krill or a mysid. We then transferred each sample to one or more 2 ml tubes containing 0.5 g of Zirconium Oxide beads with a diameter of 1.4 mm, which we then topped up with 96% ethanol. The equipment used for sample processing was either washed in chlorine and rinsed with distilled

water or heat-sterilized between each use to avoid cross-contamination. The samples were homogenized using a Precellys 24 bead beater (Bertin instruments, Montigny-le-Bretonneux, France). The bead beating was run for 3 cycles of 10 seconds with a speed of 5000 rpm.

In addition to the 54 samples from the time-series, samples containing parasitized individuals were pooled to 4 samples, so that each sample contained a small number of copepods infected with a single type of parasite (Figure 13). These were put in the same type of tubes as the other samples and processed the same way unless specified otherwise. These parasitized individuals will be referred to as "individual parasite" samples for the rest of this text.

DNA extraction

Before DNA extraction, we randomized sample order again in the same way as described above. For DNA extraction, a sample amount of around 30 mg was required. We centrifuged each tube of homogenized sample and discarded the supernatant. Then we transferred a sub-sample of 25-35 mg from the homogenized sample to a clean 1.5 ml tube with a disposable plastic spatula. For the individual parasite samples, we used all of the material, requiring a slightly different method. Each of these tubes were thoroughly vortexed, and all supernatant was transferred to a new tube, leaving only the beads in the original tube. The new tubes were centrifuged to bring all the biomass to the bottom of the tube, and the supernatant was carefully discarded.

We then extracted DNA from our samples using the DNeasy Blood & tissue kit (QIAGEN, Hilden, Germany) following the manufacturer's instructions. The kit uses a spin-column protocol with the following general steps:

1. Lysing the tissue to make the DNA available and degrade protein
2. Binding the lysate to a silica-based membrane in a spin-column
3. Washing away contaminants
4. Eluting the sample, i.e. releasing it from the membrane and transferring it to a solution.

For lysis, 180 µl of lysis buffer ATL and 20 µl of Proteinase K was added to each sample tube, and each tube was vortexed thoroughly. The samples were incubated at 56 °C for 3 hours, and vortexed every half hour during this step.

To prepare the solution for binding to the membrane, 200 µl of buffer AL and 200 µl of 96% ethanol was added, and the samples were mixed by vortexing. The sample solution was then transferred to the spin column and centrifuged at 8000 rpm for 1 minute. If not all the sample had passed through the membrane, the column was centrifuged again, until the whole sample was through. The DNA was now bound to the membrane, and the flow-through was discarded.

The membranes were washed twice, with the washing buffers AW1 and AW2 respectively. For each step, 500 µl of washing buffer was added to the column, and the samples were centrifuged at 8000 rpm for 1 minute after adding AW1, and 14 000 rpm for 3 minutes after adding AW2 to ensure that the membrane was completely dry before proceeding.

Samples were eluted by adding elution buffer AE to the membrane, incubating for 1 minute at room temperature, and centrifuging at 8000 rpm for 1 minute, collecting the flow-through containing the DNA in a 1.5 ml tube. The elution step was done twice with an elution volume of 100 µl, yielding a total of 200 µl of DNA solution per sample.

DNA quantification with Qubit

The DNA extracts were quantified using a Qubit 3 fluorometer (Thermo Fisher Scientific, Massachusetts, USA). The method quantifies DNA by binding of a fluorescent reagent and measuring the amount of fluorescence from the sample.

First, we made a master mix of Qubit HS reagent diluted 1:200 in Qubit HS buffer. 190-195 µl of the master mix was aliquoted to Qubit assay tubes, and the samples were added to each tube to reach a final volume of 200 µl. In addition, we prepared two tubes containing 190 µl master mix and 10 µl of HS standard #1 and #2 respectively. All samples and standards were vortexed and incubated for 1 minute at room temperature. The standards were read in the Qubit 3 fluorometer to generate a standard curve. Then the samples were measured, returning the concentration in the original sample in units of ng/µl.

Preparing DNA samples for sequencing

One of the limitations of Illumina sequencing is that only relatively short fragments can be sequenced continuously, with a cap of 600 bp on the MiSeq v3 (Illumina, 2021a). This means that to maximize the diversity that is captured with the least amount of sequencing, it is common to only sequence a part of a marker gene marker gene. Marker genes for sequencing typically have both conserved regions for primer binding and variable regions for identifying taxonomy. We chose to use the V4 region of the 18S gene (18SV4), which fills these criteria and has a huge reference library to get taxonomic information from the sequences (Guillou et al., 2013). Additionally, the V4 region has the appropriate length to be sequenced with the Illumina technology.

Box 4: Common issues with PCR-based metabarcoding

- **Amplification bias:** That sequences from some groups of organisms are amplified better in the PCR than sequences from other groups. This occurs when the primers match binding sites in the DNA for some group better than others and yields a biased estimate of relative abundances between groups
- **Chimera:** A new sequence made from the combination of two other sequences. Chimeras are purely artifacts of PCR and do not represent biological reality.
- **Rehybridization:** when abundant template sequences anneal with its complementary sequence instead of the primer in late cycles. This leads to reduced relative abundance of the most abundant sequences, i.e. amplification follows an asymptotic curve.

We isolated the 18SV4 region and prepared it for sequencing using two rounds of polymerase chain reaction (PCR). The first round of PCR isolates the fragment of interest (18SV4) and amplifies it, making millions of copies. The second round adds adapters at the ends of the fragments for binding to the sequencer, as well as unique index sequences used to identify the sample that the sequence originates from (Section “Index-PCR,” p. 25).

Using PCR to isolate the fragments produces some biases and artifacts that need to be adjusted for in the protocol and analyses. Problems include rehybridization, random events in the first few cycles determining the course of the rest of the reaction; and generation of false diversity from *chimeras* (Kanagawa, 2003, see also Box 4). In practice, this means that:

1. The number of reads from sequencing can not be strictly translated to abundance (see Discussion)
2. Parts of the diversity in the sample may originate from chimeras, meaning that these need to be filtered out before analysis (see section on bioinformatics)
3. The first PCR should be performed in multiple replicates for each sample to avoid the impact of random events in early cycles (Kanagawa, 2003, see also section on PCR)

Table 6: Primers used for the first PCR.

Type	Name	Sequence	Median fragment length (bp)	Reference
Anti-metazoan F	574*f	CGGTAAYTCCAGCTCYV	554	Bass & del Campo (2020)
Anti-metazoan R	UNonMet DB	CTTTAARTTTCASYCTTGCG		
18SV4 F	TAReuk454FWD1	CCAGCASCYGCGGTAATTCC	383	Stoeck et al. (2010)
18SV4 R	TAReukREV3	ACTTTCGTTCTTGATYRA		

Primer choice

When sequencing small endobionts together with larger hosts—and the target gene region is present in both host and endobiont—there is a risk that host DNA is so abundant that it overshadows the diversity of the endobionts (Bass & del Campo, 2020). In the samples from this study, copepods will have more genetic material than their relatively smaller parasites and will be overrepresented in the results. To alleviate this, it's possible to design "blocking primers"—primers that bind to a specific sequence and blocks PCR amplification of this segment—but these need to be specifically designed for the host, making them difficult to use in a general zooplankton study (del Campo et al., 2019).

Another option is to use specific primers that amplify the organisms of interest but excludes the groups dominating the biomass. Bass & del Campo (2020) made new 18SV4 primers that are specifically designed to amplify anything but metazoans (anti-metazoan primers, figure 14). We decided to use these primers in an attempt to gain the highest possible resolution of the parasites in our samples. Along with the anti-metazoan primers, we also used a general set of 18SV4 primers (Stoeck et al., 2010), both for being sure that we will capture the full metazoan diversity, and so we could compare the performance of the two primer sets on our samples. The primer sequences, names and approximate fragment lengths are shown in Table 6.

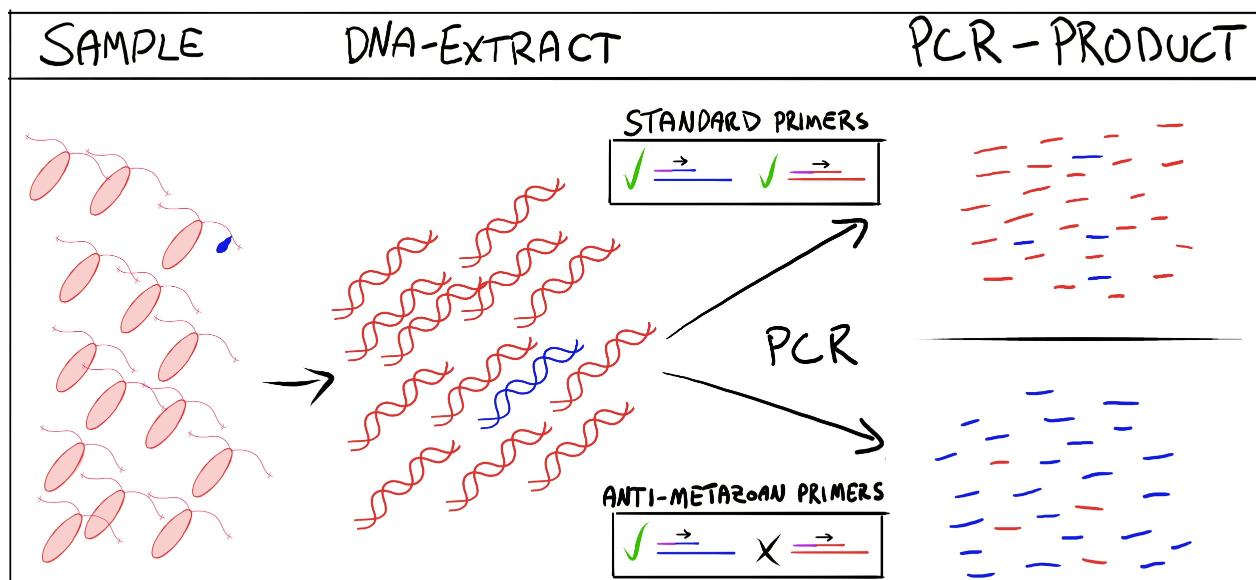


Figure 14: The concept behind the anti-metazoan primers (Bass & del Campo, 2020). The sample has a large biomass of metazoans (red), and a much smaller biomass of target protistian parasites (blue). Using standard primers that amplify both groups equally well, metazoan sequences will dominate the resulting PCR product (top). The anti-metazoan primers does not amplify metazoan sequences, and DNA of the target protists will be abundant in the PCR-product (bottom).

In addition to the primer sequence, each primer also had an adapter sequence preceding it. This adapter is vital in the second PCR where sequencing adapters and unique indexes are added to each sample. This adapter was **GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG** for the forward primers

and **GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG** for reverse. Between the adapter and primer sequences, 4 random nucleotides (N) were added to get better separation of fragments during sequencing. This means that for instance for the anti-metazoan forward primer, the complete sequence was: **TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNN**CGGTAAYTCCAGCTCYV (adapter and Ns emphasized).

Polymerase chain reaction and gel electrophoresis

We performed PCRs separately for the two primer sets in Table 6, using both primer sets for all the samples. For each sample, PCR was performed in triplicate to reduce the impact of random events in the first cycles. After the PCR the triplicates were pooled. The reagents were the same for both primer sets, and are shown in Table 7. For the general 18SV4 primers, we first tested the original protocol of Stoeck et al. (2010). This protocol, however, yielded fragments of different sizes, indicating unspecific binding in the annealing stage. In the end, we used the protocol of Krause et al. (2020) (Table 8). For the anti-metazoan primers, we used the protocol of del Campo et al. (2019) (Table 9).

In the first attempt of PCR around half of the samples showed no band. This may have been because of PCR inhibitors—salts or organic compounds inhibiting the PCR reaction—left in the DNA samples after extraction (Schrader et al., 2012). A simple strategy for removing PCR inhibitors is to dilute the samples, hopefully diluting the inhibitors enough so they do not interfere with the PCR while keeping enough DNA template for the PCR to still be successful (Schrader et al., 2012). We diluted the samples to 1/10th of their original concentration, and all PCRs were successful after this.

Table 7: Reagents for the first PCR. Protocol from https://www.protocols.io/view/illumina-miseq-dual-index-amplicon-sequencing-samp-qytdxwn

Components	Working conc.	Final conc.	volume
5xQ5 reaction buffer	5x	1x	4 µl
F-primer	10 µM	0.25 µm	0.5 µl
R-primer	10 µM	0.25 µm	0.5 µl
dNTP	2 mM	200 µM	2 µl
Q5 HF DNA polymerase	2 U/µl	0.02 U/µl	0.2 µl
Template DNA	-	-	1 µl
Nuclease-free water	-	-	11.8 µl
Sum	-	-	20 µl

Table 8: PCR program for the general 18SV4 primers. Protocol from Krause et al. (2020).

step	temperature	time	cycles
Initial denaturation	98	2min	1
Denaturation	98	20s	30
Annealing	56	30s	
Elongation	72	5min	

Table 9: PCR program for the anti-metazoan 18SV4 primers. Protocol from del Campo et al. (2019).

step	temperature	time	cycles
Initial denaturation	98.0	30s	1
Denaturation	98.0	10s	
Annealing	51.1	30s	35
Elongation	72.0	60s	
Final elongation	72.0	5min	1

After each PCR, we checked the product on a 1% agarose gel. The agarose gel was made by mixing 100 ml 1X TAE buffer with 1 g agarose, and heating the solution in a microwave oven until the reagents were mixed. When the solution had cooled to around 60°C, 4 µl of GelRed Nucleic Acid Gel Stain (Biotium, California, USA) was added for visualizing DNA bands on the gel. The gel mixture was then poured into a tray, and a comb was added to produce wells.

Before loading onto the gel, samples were mixed with 6X loading dye (Abgene, New Hampshire, USA) and distilled H_2O in a variable amount to produce 1X loading dye. The samples with dye were loaded on the gel, with 100bp DNA ladder (New England Biolabs, Massachusetts, USA) added for size reference. The gel was run at 40 V, 120 mA for 35 minutes, and visualized with a GeneGenious Bioimaging System (Syngene, Cambridge, England).

Cleaning of PCR product with magnetic beads

After a PCR, artifacts resulting from dimerization of primers—known as primer dimers— are often present and should be removed before further analysis (Das et al. (1999)). To remove primer dimers, we cleaned the PCR product using AMPure XP beads (Beckman Coulter, California, USA). The beads bind DNA with a preference for long fragments, allowing removal of shorter fragments like primer dimers. They are also paramagnetic—magnetic when in a magnetic field—allowing for easy separation of beads and the liquid they are dissolved in.

First, we mixed the pooled PCR product of each triplicate sample with the beads and incubated it for 5 minutes. For cleaning after the first PCR, a bead:PCR-product rate of 1.8:1 was used. After the second PCR, a ratio of 0.65:1 was used to increase competitive binding to the beads and maximize removal of primer dimers.

After mixing and incubation, the plate with samples was transferred to a DynaMag-96 Side Skirted magnet plate (Invitrogen, California, USA) and incubated for 2 minutes. The magnet separated the beads from the liquid, and the liquid was discarded. Then the beads remaining in the wells of the plate were washed twice with 70% ethanol, by adding 200 µl of ethanol, incubating for at least 30 seconds, and removing the ethanol with a micropipette.

We eluted the purified DNA from the beads by taking the samples away from the magnet plate, thoroughly mixing the beads with elution buffer and incubating for 2 minutes. The samples were placed back on the magnet plate to separate the beads from the solution, and the solution containing purified beads was transferred to a new tube. For the first cleaning, nuclease-free water was used as elution buffer, and for the second 10mM Tris-HCl (pH 8.5) was used.

Index-PCR

All samples needed to be pooled before sequencing. To tell sequences from different samples apart, we added a unique combination of indexes to each fragment by PCR. The indexes were added using a Nextera XT Index Kit (Illumina, California, USA), and are compatible with our adapter sequences (see section on primers).

These kits add different indexes to the forward and reverse part of the DNA fragments, in addition to an adapter for binding in the sequencing. For our 64 samples, all possible combinations of 8 indexes on the forward primer-end and 8 on the reverse end were added, for a total of 64 unique combinations. The same indexes were used for the same sample for the two different primer sets, as these can be separated again in the bioinformatics step. The reagents for the index PCR are listed in Table 11, and the program used is listed in Table 10.

Table 10: PCR program for the indexing PCR.

step	temperature	time	cycles
Initial denaturation	98	2 min	1
Denaturation	98	30s	8
Annealing	55	30s	
Elongation	72	30s	
Final elongation	72	5min	1

Table 11: Reagents for the indexing PCR.

Components	Working conc.	Final conc.	volume
5xQ5 reaction buffer	5x	1x	4 µl
Nextera XT i5	-	-	2 µl
Nextera XT i7	-	-	2 µl
dNTP	2 mM	200 µM	2 µl
Q5 HF DNA polymerase	2 U/µl	0.02 U/µl	0.2 µl
Template DNA	-	-	2 µl
Nuclease-free water	-	-	7.8 µl
Sum	-	-	20 µl

Measuring quantity and sample pooling

After the indexing PCR, we measured the quantity of DNA in each sample with a Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific). The method is fluorometric, similar to Qubit, but is optimized for doing many samples at the same time.

First, we prepared 1X TE buffer by diluting the stock 20X TE with distilled water. Then a 2 ng/µl solution of DNA standard was made by mixing 6 µl of DNA stock with 294 µl of 1X TE. The PicoGreen reagent was prepared by diluting it 1:200 in 1X TE and kept in the dark until use. 98 µl of TE buffer and 2 µl of sample was loaded into each well of a black 96-well-plate with flat, clear bottom. Standards were added to the last wells in a dilution series (Table 12). Finally, 100 µl of the diluted PicoGreen reagent was added to each well and the contents of the well were mixed by pipetting. The plates were incubated in the dark for at least 5 minutes.

We used a Synergy MXII plate reader (BioTek, Vermont, USA) to measure the fluorescence. The samples were excited at 480 nm, and emission was measured at 520 nm. We used the standards to generate a linear regression in Microsoft Excel, and used the equation of the regression to calculate the DNA in our samples.

Table 12: Dilution series of standards in the PicoGreen assay. Each dilution was done in duplicate.

1XTE (µl)	Stock DNA (µl)	PicoGreen dye (µl)	final DNA concentration (ng/mL)	Total DNA (ng)
0	100	100	1000	200
90	10	100	100	20
99	1	100	10	2
100	0	100	0	0

After measuring concentration we cleaned our samples with magnetic beads before pooling our samples in a single tube, in equimolar concentrations. Equimolar pooling meant that the negative controls, with low DNA content, had the largest volumes in the pool. This may have amplified number of contaminant sequences in the data (see Results, p. 32). To make the final pool, we first had to calculate the molar concentration of our sample with the following formula:

$$C_{mol} = \frac{C}{M_N L} * 10^{-6}$$

Where C_{mol} is the molar concentration ($pmol\ \mu l^{-1}$ or $\mu mol\ L^{-1}$), C is the weight per volume concentration ($ng\ \mu l^{-1}$), M_N is the mass of a single nucleotide ($660\ g\ mol^{-1}$) and L is the length of the DNA sequence. 10^{-6} is a conversion factor from ug to pg.

The formula we used for pooling was:

$$V_1 = \frac{C_2 V_2}{C_1 n}$$

Where V_1 is the pooling volume of each sample (μl), V_2 is the desired total volume (μl), C_1 is the sample concentration ($\mu mol\ L^{-1}$), C_2 is the desired final concentration ($\mu mol\ L^{-1}$) and n is the total number of samples. The amount of buffer to add to the pool was calculated with the formula $V_2 - \Sigma_{i=1}^{n} V_{1i}$.

We then added the calculated amounts of buffer (10mM Tris-HCl) and sample to a single 1.5 ml tube. We used the same pipette for all samples to ensure that any inaccuracies in the pipette would affect all samples in the same way. The final pool was cleaned with magnetic beads with a bead:sample ratio of 0.65 and checked on an agarose gel for primer dimers before submitting for sequencing.

Illumina sequencing

The samples were sequenced with an Illumina MiSeq 300bp paired-end run. The sequencing was performed at the Norwegian Sequencing Centre at Oslo University Hospital. Illumina sequencing is based on the principle of sequencing by synthesis, where fluorescent nucleotides emit light as they are added to a DNA chain (Illumina, 2021b). The fluorescence is detected by the sequencer, which translates the light signal into a base (Illumina, 2021b). Illumina sequencing can not detect single DNA fragments, but instead makes multiple copies of all fragments before sequencing (Buermans & den Dunnen, 2014). For paired-end sequencing, the sequencer reads the fragment in both the forward and reverse directions, as well as reading the indexes at the ends of the DNA molecule for later separation of sequences from different samples (Illumina, 2021b). The paired-end sequencing thus theoretically enables DNA fragments of up to 600bp to be sequenced, although sequences tend to have lower quality if they are longer than 500bp (Tan et al., 2019).

Bioinformatics pipeline

The result of sequencing was delivered as two individual fastq-files for each sample, one containing forward reads and the others containing reverse reads. In this section, I describe how these are merged, filtered

based on the two primer sets, cluster to operational taxonomic units (OTUs) and assign the OTUs to entries in a taxonomic database. The computations were performed on resources provided by UNINETT Sigma2 – the National Infrastructure for High-Performance Computing and Data Storage in Norway. A description of the scripts for processing are shown in Appendix B, the full scripts are available at https://github.com/evengar/master-thesis.

The bioinformatics pipeline was adapted from the metabarcoding pipeline of Frédéric Mahé, available at https://github.com/frederic-mahe/swarm/wiki/Fred's-metabarcoding-pipeline. All scripts were modified to work with our data, and the ones that do not have a counterpart in the available pipeline were made from scratch.

Initial processing

The forward and reverse reads were merged using VSEARCH 2.9.1 (Rognes et al., 2016) with the `--fastq_mergepairs` option (script mergepairs.sh). The quality of the merged files was checked with fastqc (Andrews, 2010) and summarized with multiqc (Ewels et al., 2016).

Sequences from both primer sets were still present in each file. These were separated using Cutadapt 2.7 (Martin, 2011). This program cut each sequence based on the primers used in the first PCR, based on whether or not they matched the anti-metazoan reverse primers. The ones matching the general 18SV4 primers were allocated to one file, and those matching the anti-metazoan primers to another (script primer_demultiplex.sh). So that results from both primer sets could be compared, all sequences were cut to the same region, starting after the binding site of the anti-metazoan forward primers, up until the binding site of the general 18SV4 primers.

In a file from Illumina sequencing using general primers, there will typically be a lot of duplicate sequences. To make processing faster, strictly identical sequences can be merged and annotated with the sequence abundance in a process known as dereplication. The sequences were dereplicated (local_derep.sh) in each fastq file separately. In addition, quality information from the fastq files was extracted and fastq-files converted to fasta.

Any sequences present in the negative controls represent contamination and should be filtered out. I tried two different filtering approaches: filtering out the exact sequences that are present in the negative controls and filtering out the OTUs where the negative controls are part of the cluster. The first approach was done at the current processing stage, using awk programming language (Aho et al., 1987) to filter out exact sequence matches from the samples (script filter_negative.sh). This was done for each primer set separately, and both filtered and unfiltered samples were analyzed downstream.

Clustering and chimera detection

Before assigning taxonomy to the sequences, they were clustered into *Operational Taxonomic Units* (OTU). This means that similar sequences are binned together to form groups that act as a proxy for species (Floyd et al., 2002). Clustering to OTUs is done because the same species may have slightly different sequences in the data, either due to intraspecific variance, PCR errors or sequencing errors. When clustering to OTUs, there is a chance of either assigning 2 or more species to 1 OTU (underclustering) or assigning 1 species to 2 or more OTUs (overclustering), and method choices greatly affect how many OTUs you have (Clare et al., 2016). In other words, OTUs can not be directly translated to species but is rather an indication of the diversity of your samples.

The procedures for global dereplication, clustering and chimera checking were run as a single script (derep_cluster.sh). Before clustering, all sequence files were pooled into a single file and dereplicated with VSEARCH.

The clustering was done with Swarm 3.0.0 (Mahé et al., 2014, 2015) with the "fastidious" option enabled and a d-value of 1. Swarm is a clustering algorithm that is not dependent on any global similarity thresholds but on local similarity thresholds between sequences in the cluster. Global similarity thresholds, typically

set (arbitrarily) to 97 % similarity in metabarcoding studies, are poorly suited when investigating a diverse group of organisms with different mutation rates (Brown et al., 2015). Swarm is specifically made to provide accurate clustering even with this kind of data.

Swarm generates OTUs where no sequence has more than *d* differences from another sequence in the OTU and also has an algorithm for breaking up OTUs that are likely to be different species. Swarm v2 introduced the *fastidious* option, which graft low-abundance OTUs onto closely related high-abundance ones via virtual amplicons linking the two, resulting in less overclustering. Each OTU generated by Swarm has a representative sequence, which is the most abundant sequence of the OTU.

After clustering, the OTUs were sorted by decreasing abundance and checked for chimeras with the VSEARCH `--uchime_denovo` method. This algorithm checks if low-abundance sequences in the data (children) can originate from two high-abundance sequences in the data (parents). The chimera search starts at the beginning of the file, adding any sequences that are assigned as chimeras to a "parent" database. The subsequent sequences are checked against this database and possible children of two parents are flagged as chimeras. All OTUs that were assigned as either chimeras or potential chimeras were removed before data analysis.

Taxonomic assignment

The procedure for taxonomic assignment is based on the stampa pipeline by Frédéric Mahé (https://github.com/frederic-mahe/stampa), but modified for our sequences. Before assignment, OTUs containing a single sequence (singletons) were removed (script filter_singletons.sh).

Representative sequences of each OTU were queried against the PR^2 database version 4.13.0 (Guillou et al., 2013). To match the representative sequences, the database sequences were first cut to the relevant V4 region using the primers from the first PCR (script cut_pr2.sh). Then the taxonomic assignment was done using VSEARCH's `--usearch_global` option. After assignment, the queries with multiple best hits were merged to their last common ancestor and sorted by decreasing abundance.

Finally, to connect OTUs and their abundances to the original samples, an OTU table was made (script OTU_contingency.sh). The OTU table also contains information about sequence quality, taxonomic identity with the database and results of chimera checking. OTUs containing a single sequence ("singleton" OTUs), as well as OTUs flagged as chimeras, were removed from the table.

Post-clustering curation with LULU

To avoid overclustering and generate more reliable diversity estimates, I applied the LULU algorithm to our sequence data (Frøslev et al., 2017). The LULU algorithm uses co-occurrence and pairwise similarity data to determine if any low-abundant OTU could be an erroneous variant of a similar, high-abundant OTU. First, I produced a table where all OTUs are compared against each other for pairwise similarities with VSEARCH (script self_match.sh). Then, the LULU algorithm was run on the OTU table using the `lulu` R-package (Frøslev, 2021; script lulu.R). The results were written to a file, and this file was used for data analyses.

Data analysis

All analyses was done in R programming language version R version 4.0.3 (2020-10-10) (R Core Team, 2020) unless specified. General data wrangling was done with the `dplyr` package (Wickham, François, et al., 2020) and other packages in the `tidyverse` set of packages (Wickham, 2019b). Visualizations were done in `ggplot2` (Wickham, Chang, et al., 2020) unless specified, with the colorblind-friendly palette "Safe" from `rcartocolor` (Nowosad, 2019). Rarefaction was done using the `rarecurve()` function in the Vegan package (Oksanen et al., 2020).

Environmental variables

The CTD data was read and parsed using the `oce` R package (Kelley & Richards, 2020) and merged to a single table (script CTD-formatting.R). The STD data was read manually with `readr` (Wickham & Hester, 2020) and merged into a single table (script STD-formatting.R). Finally, the two tables were merged manually with `dplyr` (script ctd-std-join.R). From the merged table, I generated some summary statistics for each station and date (script envvir-summary.R). The depth of the pycnocline was found from the maximum value of the Brunt–Väisälä frequency (i.e. the rate of density change). Then I calculated the mean temperature, salinity and density from above and below the pycnocline. The start time of each measurement was also included as seconds from midnight.

Data for daily sun hours in 2020 was downloaded from the Norwegian Meteorological Institute (https://seklima.met.no/observations/). The observations were from the "Oslo – Blindern" station, which is around 34 km away from our farthest sampling station (IM2 – Elle). The sum of the sun hours of the previous 7 days was calculated for each sampling date.

Primer assessment

The difference between the two primers was investigated by checking the amount of overlapping and unique OTUs. This was visualized with the R package `VennDiagram` (Chen, 2018). Differences in temporal and spatial patterns of taxonomic groups in the primer sets were investigated by comparing numbers of OTUs present, as well as relative numbers of reads for each group in each sample. Unless specified, the summed reads from both primers were used for the rest of the data analysis.

Multivariate analysis

The OTU-location data was analyzed with non-metric multidimensional scaling (NMDS) using the `metaMDS()` function from the `vegan` package in R (Oksanen et al., 2020). First, the OTU table was converted from read count data to presence/absence data, and formatted with each row representing 1 station for a single month, and each column representing an OTU. Then the `metaMDS()` function was run using the Bray-Curtis dissimilarity measure from 100 random starts, with maximum iterations increased to 1000. Otherwise, the default parameters were used. I made generalized additive model (GAM) regressions of NMDS axes 1 and 2 respectively with station ID and sampling month as explanatory variables to find trends in the data. The GAM was done with the `gam()` function from the `mgcv` R-package (Wood, 2021) with cyclic spline, meaning that the regression is constrained to start and end in the same place. The NMDS was visualized with `ggplot2`. In addition, the relationship between the NMDS and different environmental variables from CTD data was investigated by plotting.

Identifying parasites and hosts

For each of the individual parasite samples, I identified potential hosts and parasite OTUs from read counts of metazoans and non-metazoans respectively in the samples. The 4 OTUs with the most reads from each sample were chosen as targets. If the database search from PR^2 gave ambiguous results, the closest match in GenBank was found with BLAST (NCBI Resource Coordinators, 2018). For potential parasites, each hit that matched a parasitic group, as well as any other OTUs that matched the PR^2 accession numbers of those hits, was tagged as parasites in our data. For the hosts, the OTUs and all matching PR^2 accessions were tagged with the last common ancestor of the GenBank hits.

I did a general search for parasite genera in the database hits, using the genus names in Table 1 in the `str_detect()` function from the `stringr` R-package (Wickham, 2019a). The results of the search were combined with the tagged parasites from the individual parasite samples to produce the final parasite table.

The PR^2 database does not have a "Copepoda" taxonomic level, so I first subset the data to class Maxillipoda. Then I used the PR^2 database R-package (Vaulot, 2021) to find the corresponding taxonomic annotation in

the Silva database (Quast et al., 2013). If the OTU had a “Copepoda” hit in the Silva database, the OTU was tagged as a copepod in our data. The Silva database was also used to find out if a copepod was calanoid or cyclopoid if there was no species match in either database.

Co-occurence analysis

The parasite-host co-occurrence was investigated with χ^2 contingency analyses. Subsets of hosts with a read abundance of more than 100 and the top 8 parasites were used. For each potential host-parasite relationship, a presence-absence contingency table was made. Each contingency table was analyzed with the `chisq.test()` function in R. The resulting p-values were adjusted for multiple testing with Holm-Bonferroni correction using the `p.adjust()` function.

Analyzing existing data sets

The data from the BioMarKs project was received directly from researchers involved in the project (Dominik Forster, Bente Edvardsen), but can also be accessed from the European Nucleotide Archive (ENA, https://www.ebi.ac.uk/ena/). The sequences were processed in the same way as our samples with two exceptions:

1. For taxonomic assignment, PR^2 version 4.12.0 was used instead of 4.13.0, since the analyses were performed before PR^2 4.13.0 was released.
2. Since the BioMarKs project used the older 454 sequencing rather than Illumina, the sequencing was naturally much shallower than ours. To avoid risk of removing real diversity, singletons were kept.

To find parasites in the BioMarKs data, the search string and accession numbers described above for our data were used.

A processed version of the data from the Tara Oceans project (de Vargas et al., 2017) was downloaded from PANGAEA (https://www.pangaea.de), along with sample information and environmental data. This data was used for creating the map in figure 9 with the R-packages sf (Pebesma, 2020) rnaturalearth (South, 2017) and ggplot2.

Results

The sequencing yielded 8,069,477 sequences to be processed and assigned taxonomy (Figure S1). In this section, I begin by assessing the data's quality and descriptive value and comparing the two primer sets we have used. Then, I identify the parasites in the individual parasite samples and use that information together with database assignments to find patterns of parasites in our data and the BioMarKs data.

All figures with time-of-year on the x-axis contain a greyed-out area in April 2020, representing the missed field sampling that month due to the COVID-19 pandemic. In all figures that include the Oslofjord stations, the legend is ordered from the innermost (EP1) to the outermost (IM2) station. Some of the visualizations of quality statistics are in the supplementary material (see Appendix A.)

Overview of the metabarcoding data

The sequences had good overall quality after paired-end merging, with only a single sample (Sample 29, anti-metazoan primers) flagged with a warning by FastQC (Figure 15). A warning in FastQC occurs if any base has either a median Phred-score below 25 or first quartile below 10. The longer sequences generated by the anti-metazoan primers had a considerable quality dip around the middle compared to the regular 18SV4 primers due to shorter overlaps between forward and reverse reads.

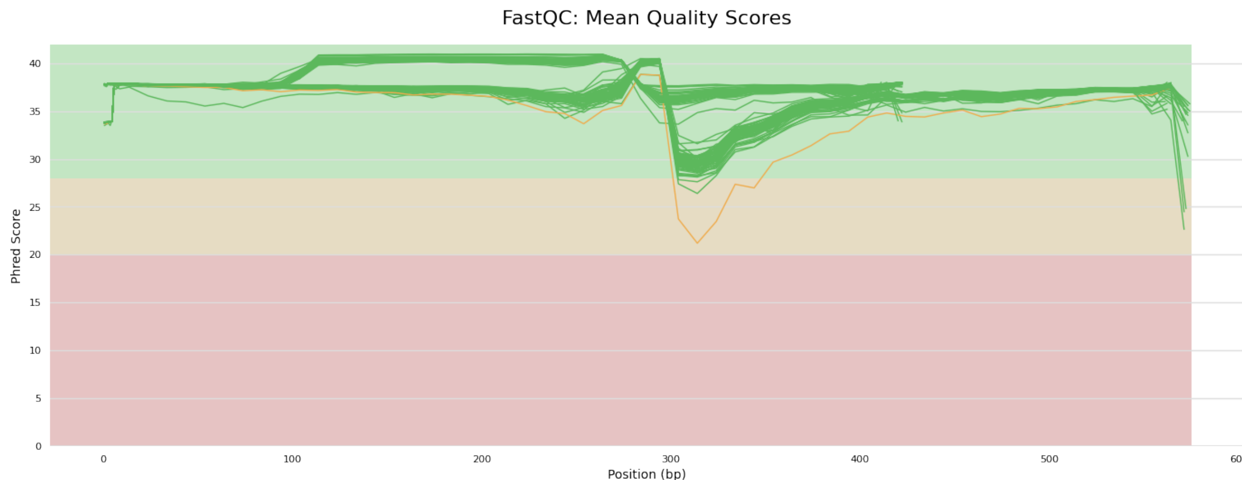


Figure 15: Per base quality scores from FastQC after merging of paired ends. The lines represent the mean score of the bases in a given position along the sequence, with each line representing 1 sample. The fields in the graph represent base qualities that are "very good" (green), "reasonable" (yellow) and "poor" (red). The yellow line is sample 29 with anti-metazoan primers. FastQC results were combined using multiQC to produce this graph.

There was a general problem with contamination in the data. The first approach to removing sequences present in the negative controls—filtering on the OTU level after clustering—removed 94% of reads and 11% of OTUs in the data. The second approach—filtering on exact sequence identity before clustering—removed 6% of unique sequences and 73% of reads from the 18SV4 sequences, and 4% of unique sequences and 23% of reads from the anti-metazoan sequences. To avoid substantial data loss, I used the results filtered using the second approach for further analyses.

The total number of OTUs after filtering, clustering, singleton removal and chimera removal was 24,794. Lulu-curation discarded 16,349 of these, leaving 8,445 OTUs in the final data. After curation, sequences from the 18SV4 primers had a read total of 482,618 and a mean of 8,321 reads per sample, and those from the anti-metazoan primers had a total of 589,061 reads, averaging 10,156 per sample (Figure 16). The taxonomy was dominated by metazoans, with many of the time-series samples also having a high relative read abundance (proportion of total reads in the sample) of the algal groups Dinoflagellata and Ochrophyta, among others (Figures 17 and 18).

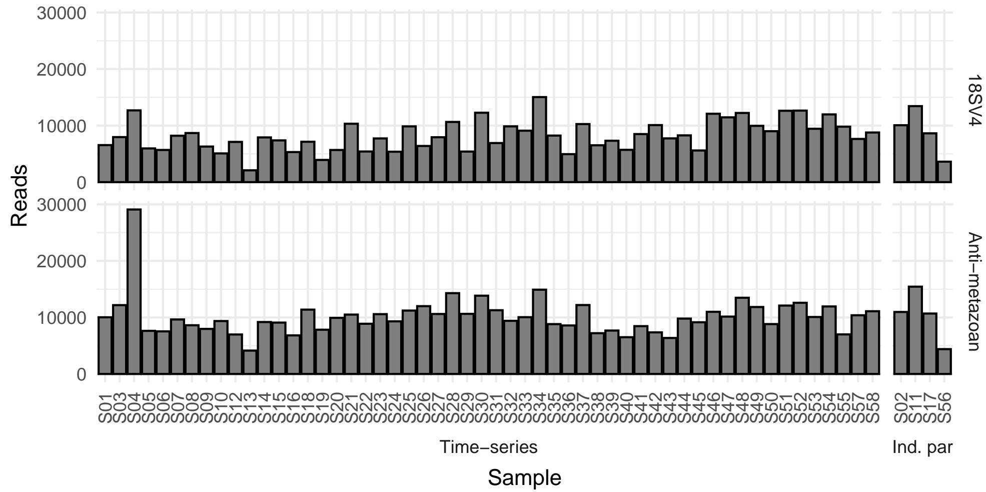


Figure 16: Total reads per sample per primer set. Read counts are after removing sequences that are strictly identical with those in the negative controls.

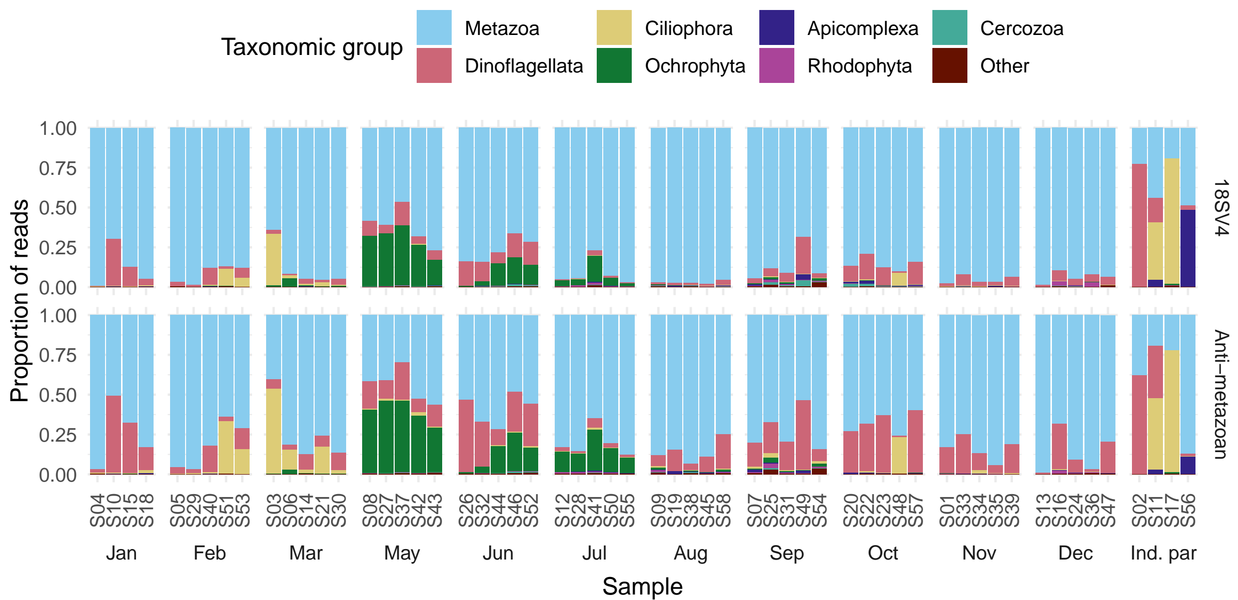


Figure 17: Overview of the largest taxonomic groups present in the data. Samples 2, 11, 17 and 56 are individual parasite samples, and the rest are time-series samples.

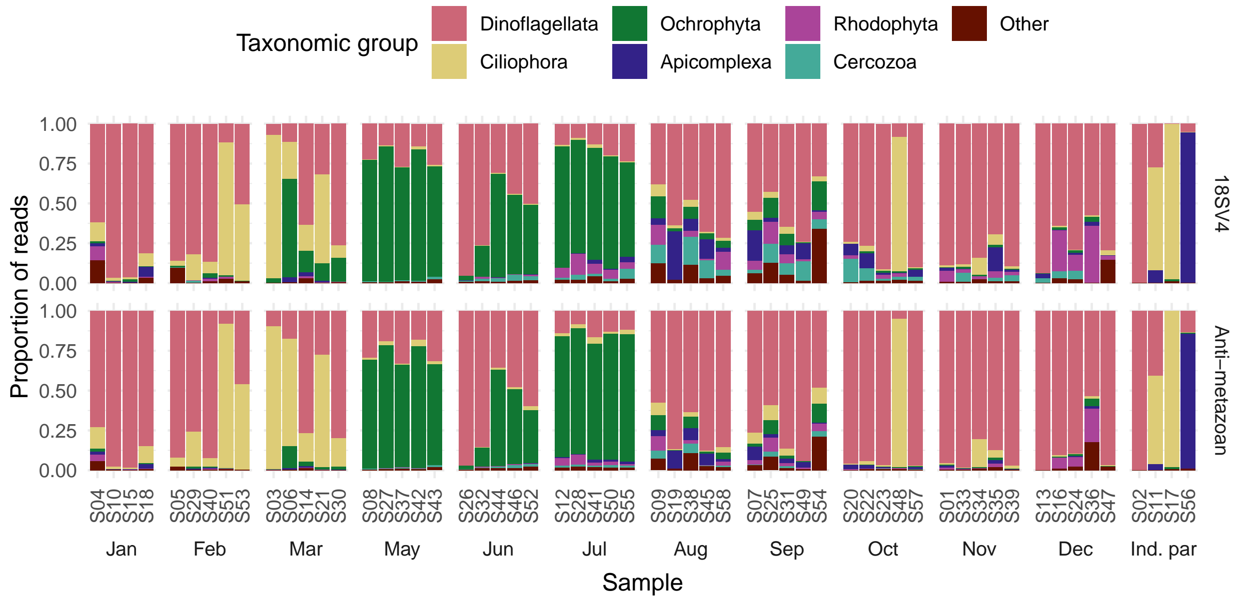


Figure 18: Proportion of reads for the non-metazoan groups in the samples.

The majority of the retained sequences were of similar length, had an identity of more than 90% with the reference database, and overall good quality (Figure S2). I further filtered the data by removing sequences with a quality score of more than 0.002 and taxonomic identity of less than 80%. The length of the 18SV4 region can vary (Brown et al., 2015; Nickrent & Sargent, 1991), so I did not do any further filtering on sequence length. This final filtering removed 1,048 OTUs, and the final table contained 7,397 OTUs.

Rarefaction curves are traditionally used in ecology for determining if the sample size is sufficient (Heck et al., 1975). Rarefaction curves show expected species number as a function of sample size. If the curve reaches an asymptote, you do not expect a significant increase in the number of species if you increase the sample size (Heck et al., 1975). Similarly, rarefaction can be used to assess if the sequencing depth is good enough in metabarcoding studies (Grey et al., 2018). Our samples' curves level off but do not quite reach an asymptote (Figure 19 A). At the genus level, or when grouping by month, the slopes at the end of the curve are shallower (Figure 19 B-C). If we had increased sequencing depth, we would expect to find between 2 and 8 more OTUs per 100 new reads (de Vargas et al., 2015), indicating that our study could benefit from a bit deeper sequencing.

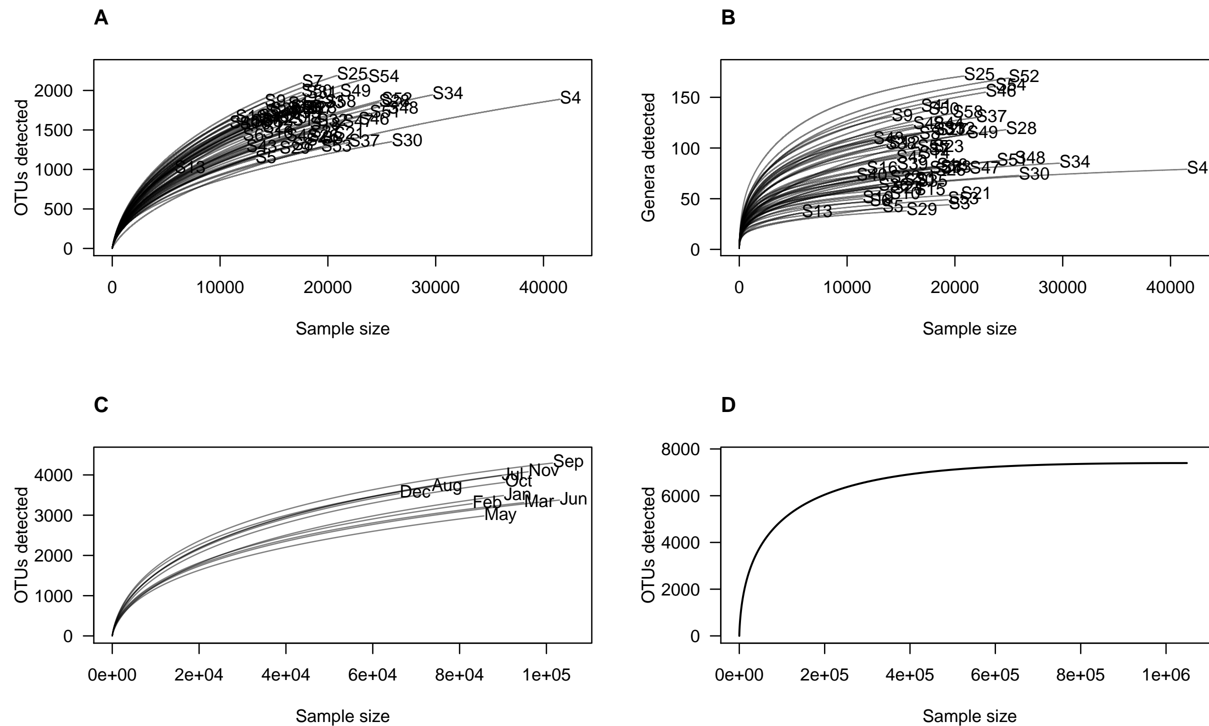


Figure 19: Rarefaction curves of the metabarcoding samples at OTU level (A) and genus level (B), samples grouped by month at OTU level (C) and curve of all samples summed (D). Range of slopes at the end of each line: ***A:*** *0.021-0.081 (S4 and S13, respectively).* ***B:*** *0.00036-0.0027 (S4 and S43, respectively).* ***C:*** *0.012-0.021 (June and December, respectively).* ***D:*** *4.4e-06. Curves can be extrapolated (with caution) to indicate how many new groups will be discovered when sample size increases, i.e. a slope of 0.02 at the OTU level means that we expect to find 2 new OTUs per 100 new reads (de Vargas et al., 2015).*

Primer assessment

The purpose of the anti-metazoan primers (Bass & del Campo, 2020) was to get better resolution of non-metazoan taxa by inhibiting amplification of metazoan sequences. However, the samples were still dominated by metazoans regardless of primer, and the differences between primer sets were smaller than expected. The majority of the OTUs from the sequencing were amplified by both primer sets (72%, Figure 20 A). While the anti-metazoan primers captured a slightly higher diversity of non-metazoans, the community composition between the two was largely the same (Figure 20 B). The most common OTUs were captured using both primers, and only relatively rare sequences differed between primer sets. The anti-metazoan primers captured more rare non-metazoan OTUs, most with 100 reads or less (Figure S3). The number of OTUs from each group in each sample was linearly correlated between primers, meaning that they uncover more or less the same diversity (Figure 21). Based on this assessment, I used the combined data from anti-metazoan and regular 18SV4 primers (made by summing the reads for each sample and OTU) for the rest of the data analyses.

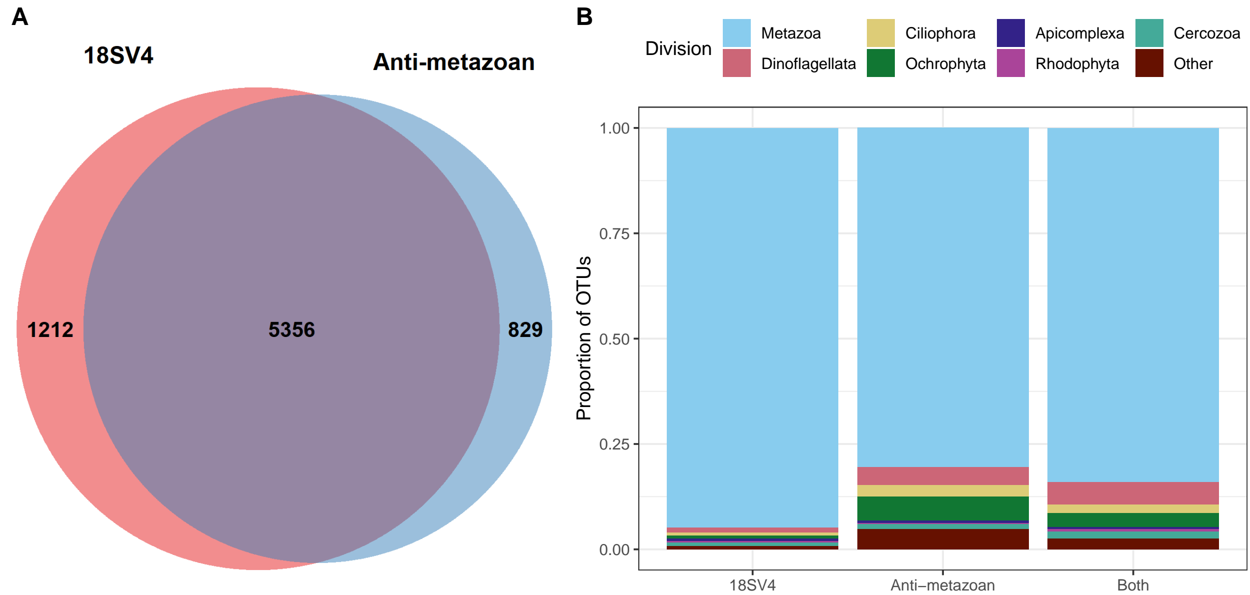


*Figure 20: **A:** Venn diagram of overlapping OTUs between the regular and anti-metazoan 18SV4 primers. **B** Proportion of OTUs by taxonomic group that was exclusive to either primer, or found in both. The "Other" category is all other groups than the top 7 shown.*

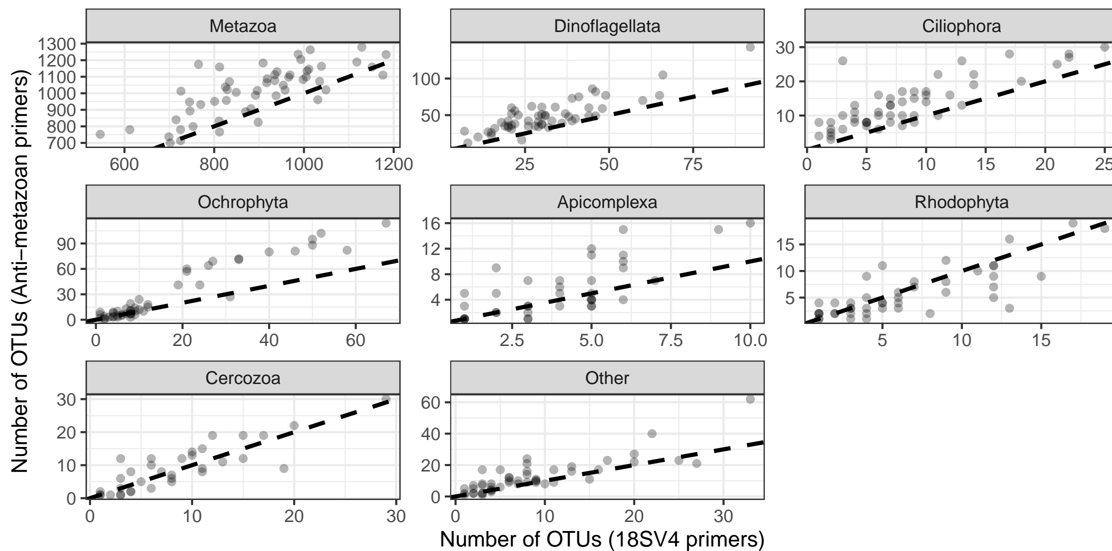


Figure 21: Relationship of OTUs from the same samples in the two different primer sets. The stippled line represents a 1:1 relationship.

Multivariate analysis

The NMDS captured the seasonal cycle in the oslofjord. The two first dimensions show that variation in samples is cyclic through the year (Figure 22 A). Samples from the same month are consistently close to each other, and January 2020 was very similar to December 2020. The samples from May through July and the rest of the samples, respectively, formed two distinct groups in the NMDS plot. The NMDS also

corresponded strongly with the patterns in critical environmental variables, further supporting that it has captured the seasonal variation in the samples (Figure 22 B-D).

Additionally, the NMDS captured the differences between stations, where the innermost and outermost stations (EP1 and IM2 respectively) are positioned furthest apart (Figure 23). The other stations position in-between, following a rough pattern of inner to outer placement in the fjord. It is also worth noting that the outermost station IM2 has less variation in NMDS3 than the others, which may indicate that it has less variety in species composition throughout the year.

The NMDS analysis fit the data well overall. It converged with 100 random starts (the minimum amount passed to `metaMDS()`) and has a stress score of 0.07. While there are no hard rules for acceptable stress scores, a score of less than 0.1 usually indicates a good ordination (see Dexter et al., 2018 for a discussion on this). The distances between points in the NMDS are highly correlated with the calculated distance in the distance matrix, indicating that the ordination is a good fit for the data (Figure S4).

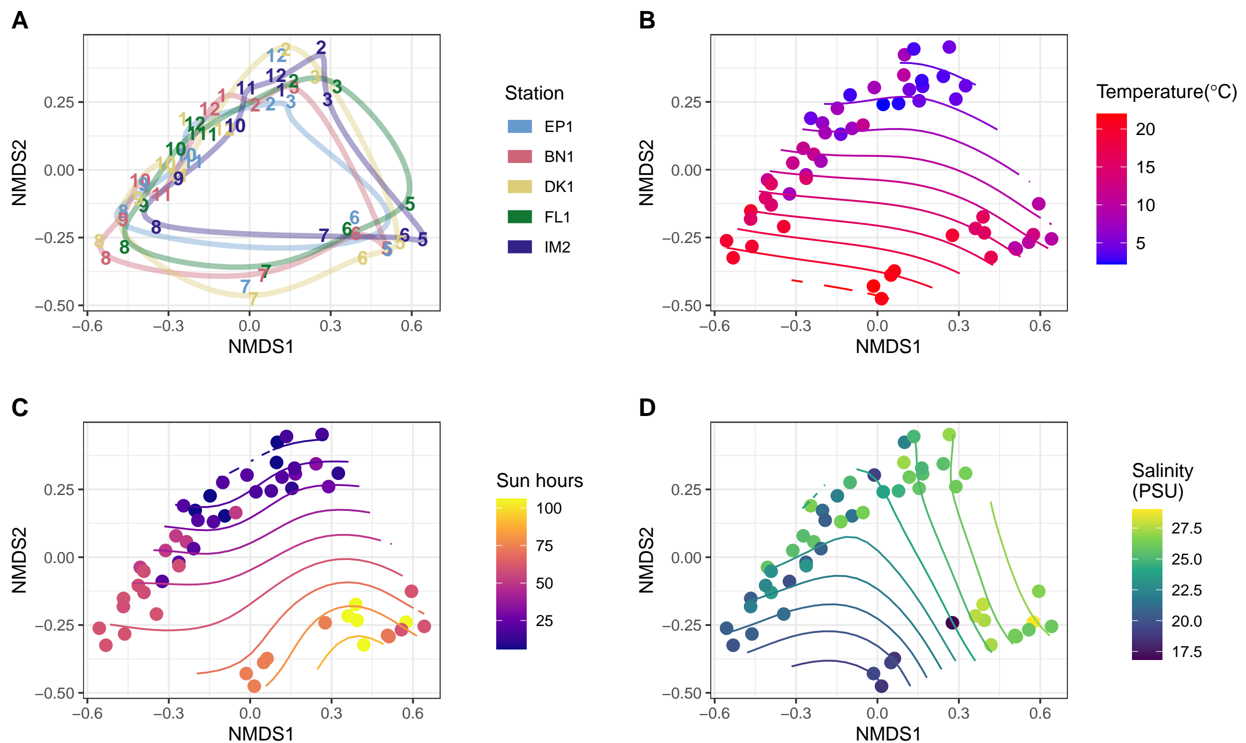


Figure 22: The first 2 dimensions in the NMDS of the time-series samples based on the metabarcoding data. A: The numbers are the sampling months, and the color is the station. The lines are generalized additive models (GAMs) with cyclic cubic splines of the two NMDS axes by station and month, plotted against each other. B: Points colored by temperature in the upper 10 meters. C: Points colored by the sum of sun hours 7 days prior to sampling. D: Points colored by salinity in the upper 10 meters. In plots B-D, the lines represent the gradient in environmental variables and can be interpreted the same way as contour lines on a map.

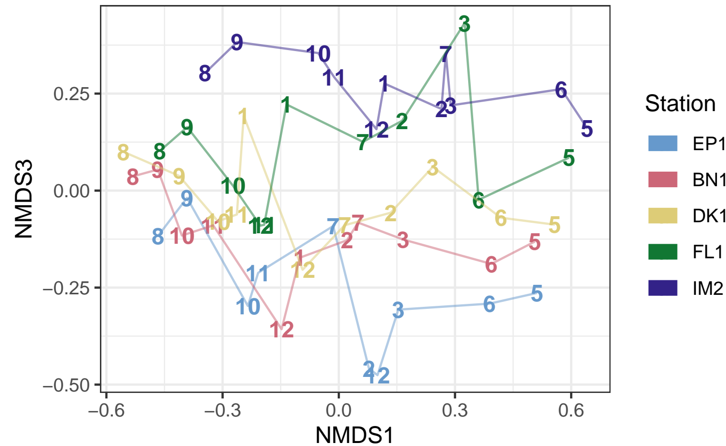


Figure 23: The first and third dimension of the NMDS. The lines are connecting the samples from the same station, and bear no further significance. Note that in the legend, station IDs are listed from the innermost (EP1) to the outermost (IM2) station.

Patterns of abundant taxonomic groups

The largest taxonomic groups in the samples had a marked seasonality in the number of OTUs present. The patterns were also largely consistent between stations in the same month (Figure 24). The two purely parasitic taxa Gregarinomorphea (Apicomplexa) and Syndiniales (Dinoflagellata) had peak OTU abundances in late fall, with Syndiniales having an additional peak in June (Figure 25).

Although the mesh size of our net hauls was 200 µm, a sizable proportion of our data was from phytoplankton groups (Figure 24). Some of these may come from the gut content of copepods. In addition, both large dinoflagellates and chain-forming diatoms are large enough to be included in our samples. Mainly free-living groups represent a large part of the Dinoflagellata and Ocrophyta reads in our data (75% and 96%, respectively, Figure S7). The patterns for these two groups are consistent with known phytoplankton succession patterns, where diatoms are most abundant at the beginning of the spring bloom (around March-April in the Oslofjord) and dinoflagellates are most abundant in the fall (Gran-Stadniczeñko et al., 2019; Kiørboe & Nielsen, 1994).

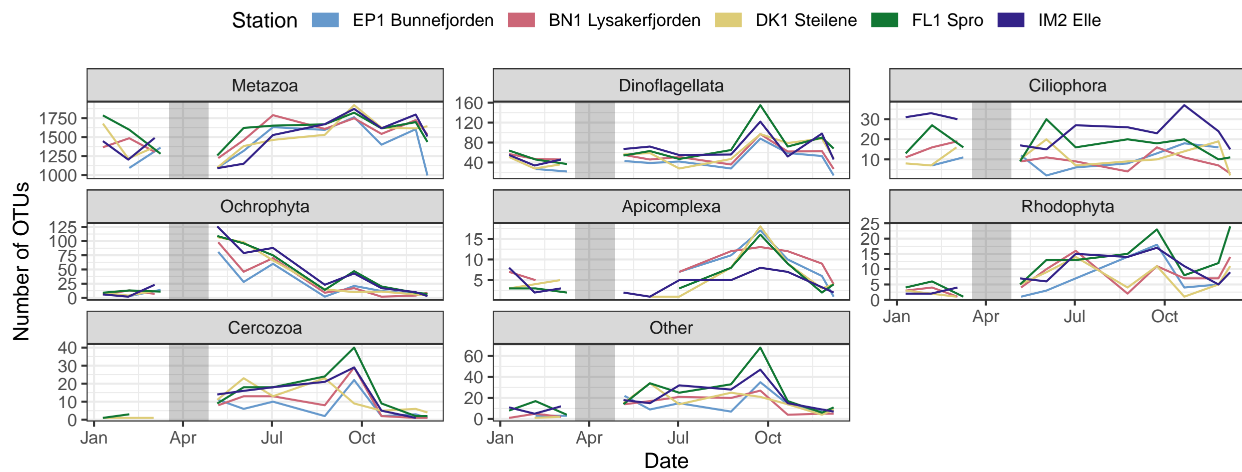


Figure 24: Seasonal patterns in the OTU counts of the largest taxonomic groups. Note: different y-axis for each group.

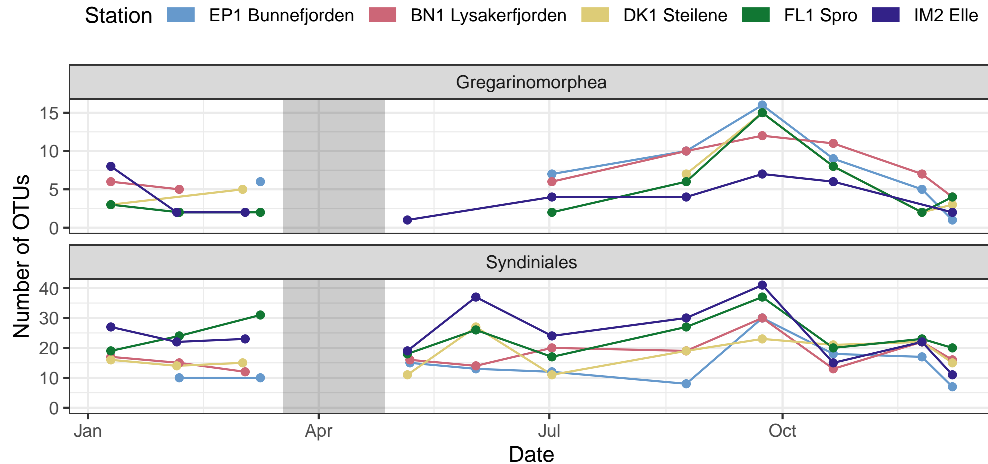


Figure 25: Seasonal patterns of the two exclusively parasitic groups Gregarinomorphea (Apicomplexa) and Syndiniales (Dinoflagellata)

Many of the copepod genera in the data showed distinct seasonal patterns in relative read abundance (percent of total reads in the sample). The genera *Temora* and *Paracalanus* had the most distinct peaks in early spring and late summer, respectively. *Metridia*, *Pseudocalanus*, *Acartia* and *Centropages* also had an apparent seasonality across all stations (Figure 26). These seasonal patterns correspond well to earlier observations from the Skagerrak area (Kiørboe & Nielsen, 1994; Zervoudaki et al., 2009). In addition, the large copepods in genus *Euchaeta* had noticeably higher read abundance at the station IM2 throughout the year. The *Euchaeta* sequences in our data was closely related to the genera *Paraeuchaeta* and *Chiridius* in BLAST (not shown), and the *Euchaeta* hits may encompass both of these. IM2 is the deepest station and the only one located outside of the Drøbak sill (see Methods), and is a known habitat for these large copepods (Schøyen & Kaartvedt, 2004; e.g., Skarra & Kaartvedt, 2003).

The negative controls (blanks) in our data had many contaminants representing actual groups in the plankton—like copepods and dinoflagellates—indicating some cross-contamination between samples (data not shown). The blanks had the highest read depth of our samples, possibly inflating this problem (see Figure S1. This problem was mitigated by removing all sequences in the blanks from the data, though not eliminated entirely. Our results may have some false positives due to this, and some read and OTU counts may be overestimated. However, the data has a clear seasonal structure that is similar to past studies, and taxon proportions are consistent between primer sets (Figure 17), indicating that our data still has some descriptive value.

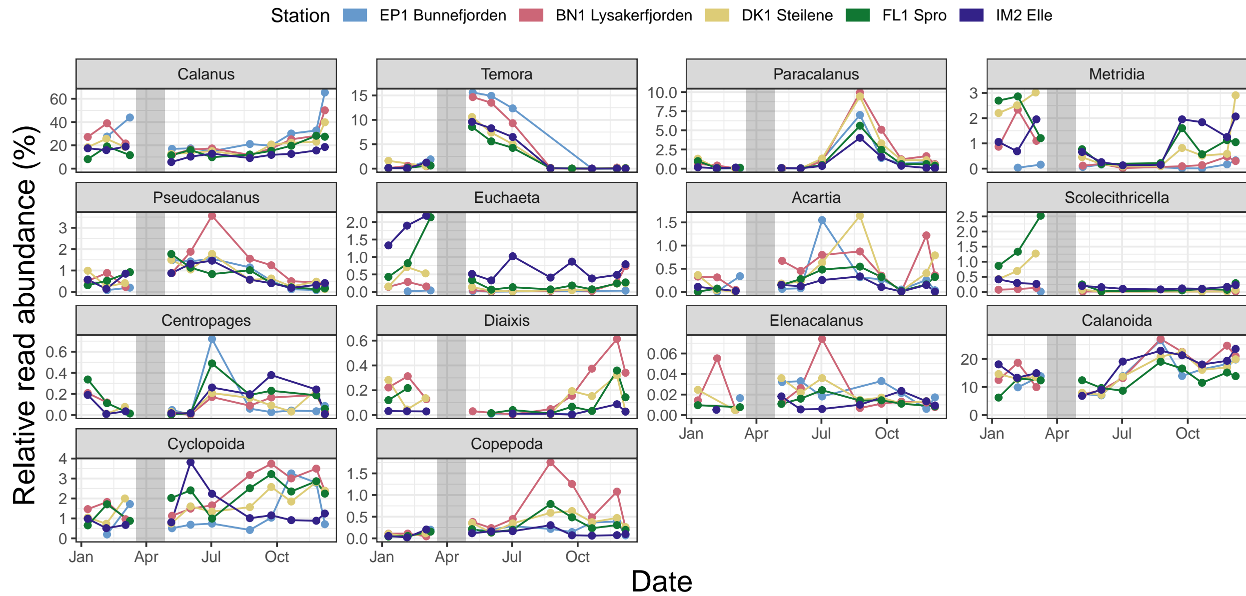


Figure 26: Patterns of the top 14 copepod genera and groups in the data based on relative read abundance (percent of total reads in the sample). All groups occur in more than 30 of the 54 samples. The groups "Calanoida", "Cyclopoida" and "Copepoda", contain sequences where taxonomy could not be determined down to genus. Reads were normalized by dividing with the total of each sample, and summed by genus and sample. For the genera Centropages and Diaxis, a single outlier was removed to make patterns clearer. Note: Different y-axes for each group

Identification of individual parasites

From the 4 individual parasite samples, I identified 5 potential parasites (Table 13). Sample 11 contained 2 parasitic lineages which made up a large proportion of reads, so both of these were included. None of the candidate parasites had an exact species match in either database, but most had an unspecific match of 100% or close to 100%, meaning these are not entirely novel sequences. These tentative parasites corresponded to 52 accession numbers in PR^2, and any hit in our data matching at least one of these accession numbers was flagged as potential parasites.

Table 13: Top hits from PR^2 and GenBank for the most abundant non-metazoans in the individual parasite samples. The genbank hit excludes all non-species level hits, e.g. "Uncultured Eukaryote Clone".

Sample	PR^2	Genbank hit	Genbank identity (%)
S2	Dinophyceae	*Ellobiopsis sp*	94.38
S11	Suctorian ciliate	*Acineta flava*	85.07
S11	*Hematodinium sp.*	*Hematodinium sp.*	93.37
S17	*Chromidina sp.*	*Chromidina sp.*	96.95
S56	Gregarine (Apicomplexa)	*Haliclona oculata*	81.40

Patterns of parasite occurrence

In addition to the 5 tentative parasites identified from the individual parasite samples, 3 genera of known copepod parasites were found in the data: *Blastodinium*, *Syndinium* and *Vampyrophrya*. The total number of unique parasite OTUs was highest in March, and lowest from May to July (Figure 27 A). Most of the parasite OTUs were relatively rare, with all except 1 *Ellobiopsis* OTU accounting for less than 1% of reads in average per sample, and the majority accounting for less than 0.01% (Figure 27 B).

OTU occurrence followed a seasonal pattern for some parasite groups, with several having the highest OTU count in early fall to winter (Figure 28). *Syndinium sp.*, the suctorian ciliate and *Vampyrophrya pelagica* had mostly only 1 OTU present each month, but that OTU was present for most of the year (Figure S5). The parasites *Hematodinium sp.* and the suctorian ciliate were primarily found in the station IM2, with only sporadic occurrences at the other stations (Figures 29 and S5).

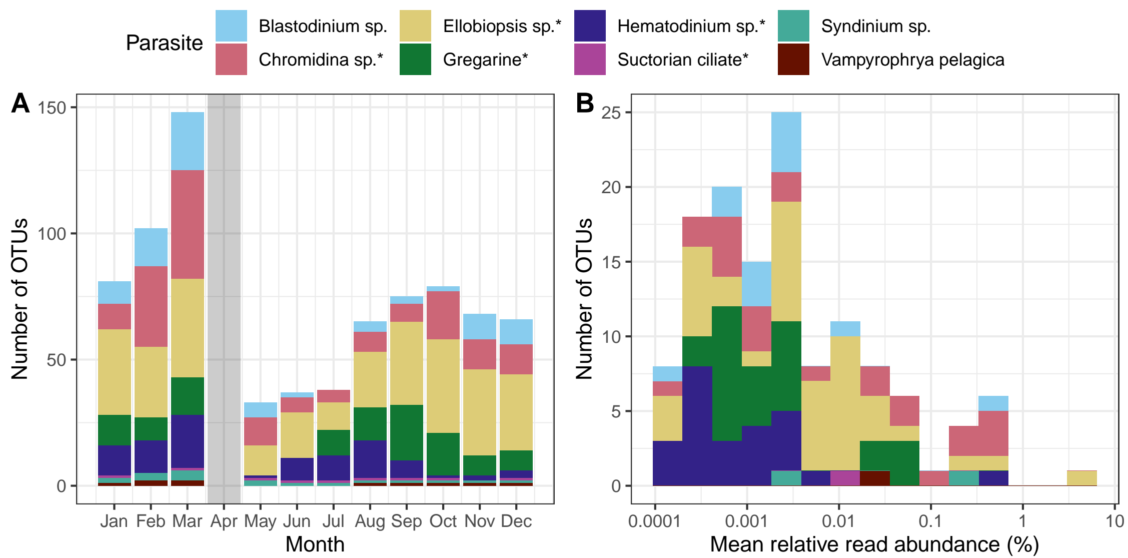


Figure 27: ***A:*** *Total unique parasite OTUs per month.* ***B:*** *Histogram of the rarity of the detected parasites in the samples, the x-axis is the mean percentage of reads per OTU across all samples. The parasites that were identified from the individual parasite samples are marked with an asterisk.*

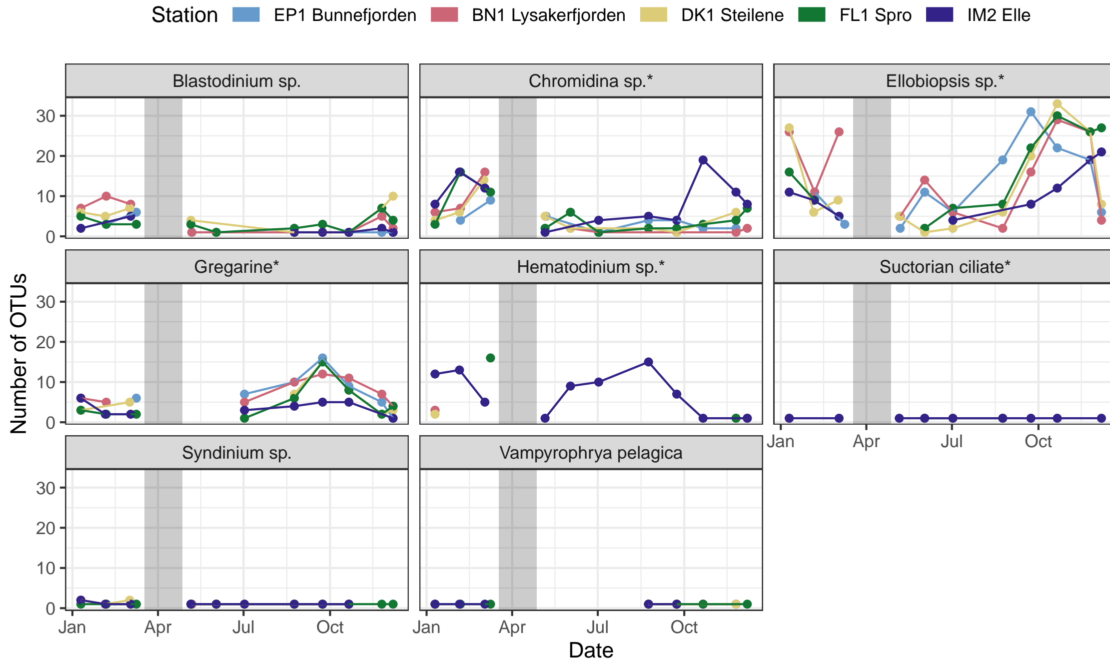


Figure 28: Number of OTUs per station per month for the copepod parasites. The parasites that were identified from the individual parasite samples are marked with an asterisk. Note: patterns are not visible in the three last groups due to overplotting; see Figures 29 and S5.

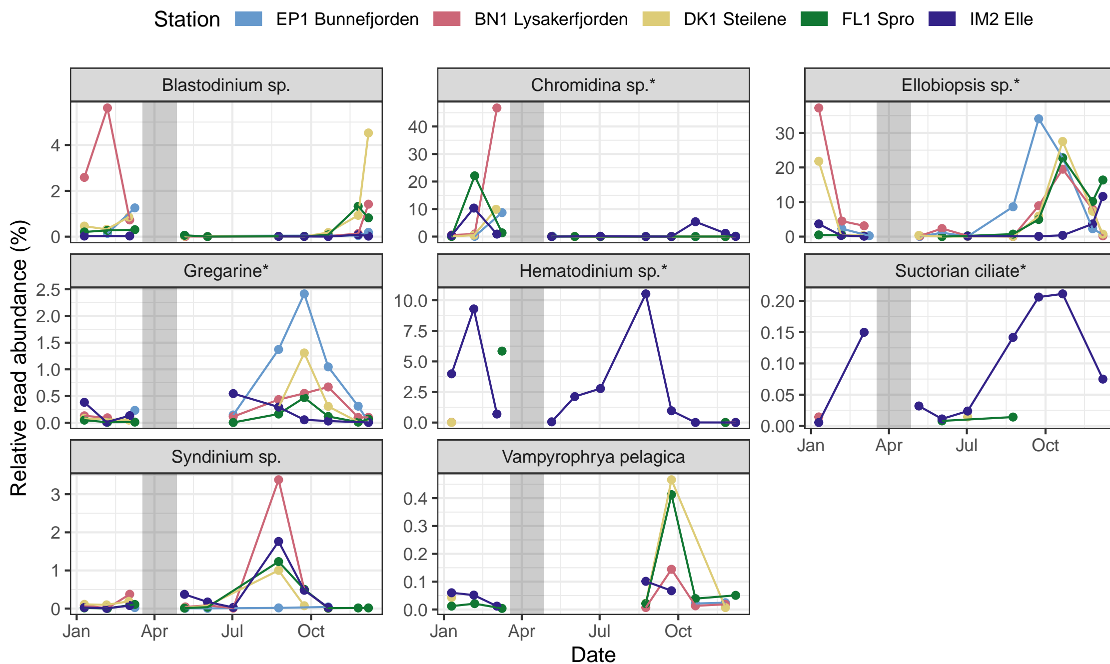


Figure 29: Relative read abundances (percent of total reads in the sample) of the copepod parasites per station per month.

Co-occurrence analysis

The χ^2 contingency test showed a single significant co-occurrence pattern between parasite and host OTUs (Figures 30 and S6) ($p < 0.05$). The parasite OTU 111 was assigned to the Gregarine parasite. The copepod OTU 215 could unfortunately not be identified beyond being a calanoid copepod. The parasite was only present when the host was also present, and the two OTUs appeared together in more than half of the samples where the host was present (Table 14).

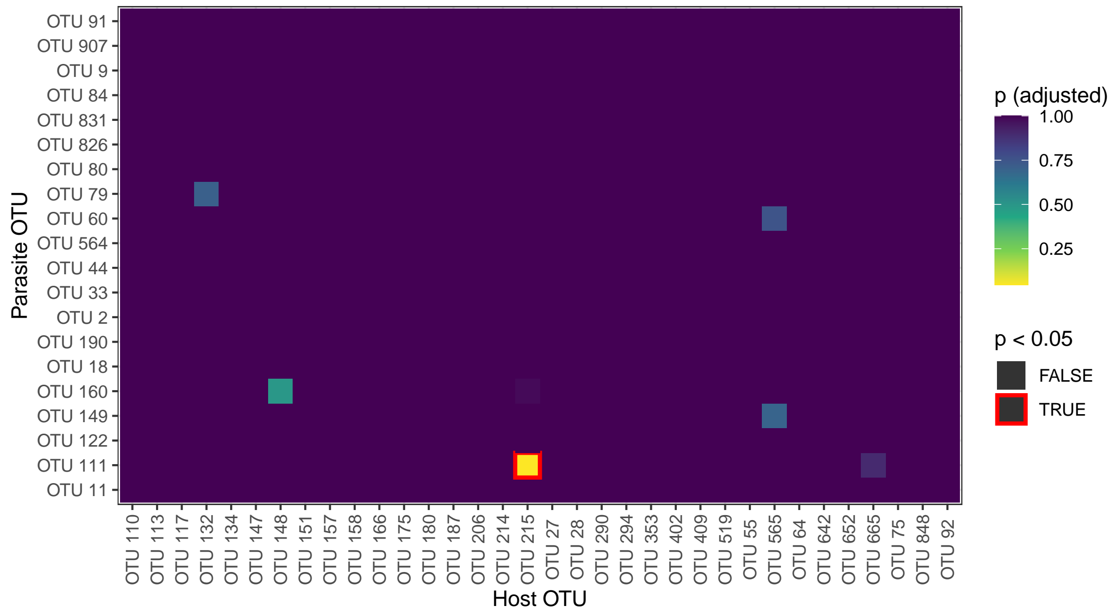


Figure 30: P-values from χ^2 contingency test of presence-absence data of the top parasite and copepod OTUs. The test shows a single significant relationship ($p < 0.05$), between OTU 111 which is assigned to the gregarines, and OTU 215, which is an unidentified calanoid copepod. P-values are adjusted for multiple testing with Holm-Bonferroni correction.

Table 14: Contingency table of OTUs 111 (Gregarine) and 215 (Calanoida)

	Parasite	
	Abs.	Pres.
Host absent	13	0
Host present	13	28

Searching in BioMarKs data

Around half of the samples from the BioMarKs project contained parasite reads, with distinct differences between locations (Table 15). Notably, no parasites were found at the Varna station (Black Sea), and Naples (Mediterranean Sea) had the highest number of OTUs. In addition to a different number of samples, the sampling dates and environmental conditions were quite different between locations (Table 16), which may contribute to the variation in number and composition of parasites.

Table 15: Total samples and samples with parasites by location from the BioMarKs project.

Location	Total samples	Samples w/parasites (count)	Samples w/parasites (%)
Blanes	11	7	63.64
Gijon	4	4	100.00
Naples	46	42	91.30
Oslo	44	16	36.36
Roscoff	9	3	33.33
Varna	25	0	0.00
Total	139	72	51.80

In the 8 parasite groups found, the search identified 5 from their taxonomic assignment in the PR^2 database and 3 from the accessions in our samples (Figure 31). The Oslo and Naples sampling sites had a similar number of samples but differed considerably in both the number of parasite OTUs and parasite composition. It is important to note that the *Ichthyophonus* hits may not correspond to a copepod parasite at all, given the results from the individual parasite samples.

The sequences from regular DNA and cDNA (DNA made from RNA by reverse transcription) had largely similar results. While DNA in the samples may originate from dead organisms, cDNA is more likely to represent organisms that were alive at the time of sampling due to the short degradation time of RNA molecules.

Table 16: Environmental data of BioMarKs samples. Adapted from Logares et al. (2012) table S1. DCM = deep chlorophyll maximum.

Location	Sampling date	Station depth	DCM depth	Surface temperature (°C)	Surface salinity (PSU)	Bottom layer salinity (PSU)	ChlA (µg/l)
Blanes	Feb 2010	20	-	12.5	37.5	38.2	1.0
Gijon	Sep 2010	110	40	20.2	35.7	36.6	7.0
Naples	Oct 2009	75	23	22.8	37.7	37.9	1.4
Naples	May 2010	75	35	19.2	37.2	37.9	1.2
Oslo	Sep 2009	100	8	15.0	25.0	35.0	3.2
Oslo	Jun 2010	100	9	15.0	22.0	35.0	1.9
Roscoff	Apr 2010	60	-	9.9	34.9	34.9	0.5
Varna	May 2010	400	40	21.5	16.0	22.0	8.0

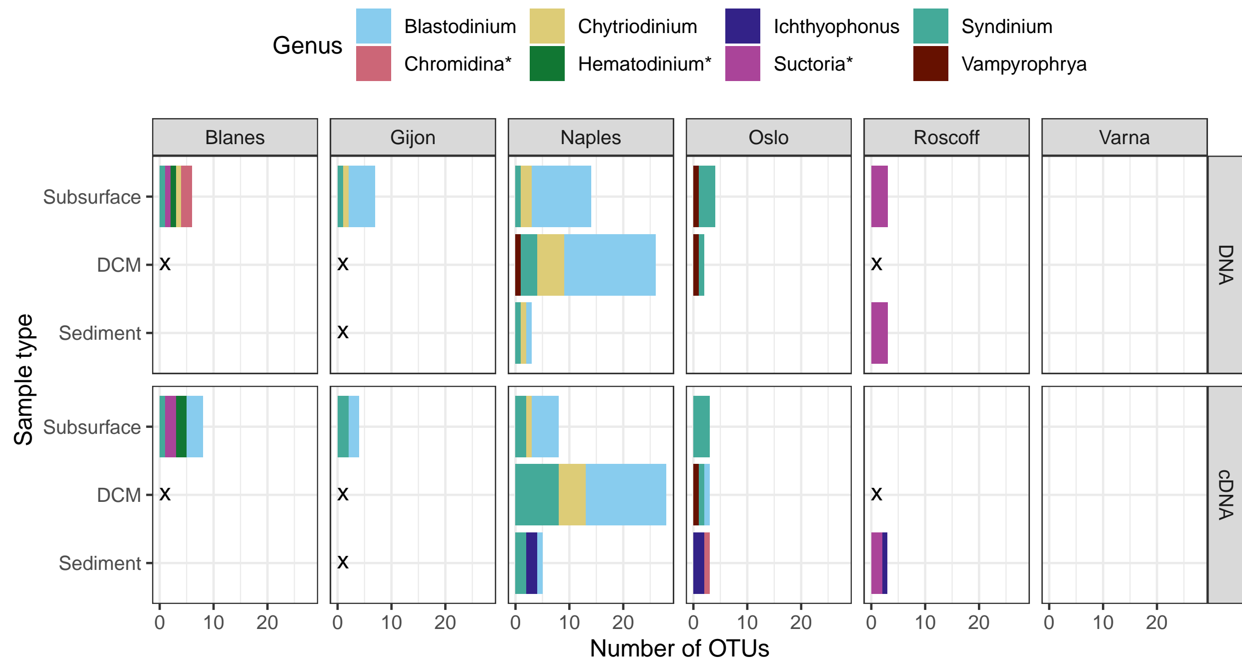


Figure 31: Parasite genera in the BioMarKs data set. Genera marked with asterisks were identified by the accessions from our samples. An "x" in the plot indicates that the combination of sample type and location does not exist. Subsurface is the upper 5 meters of the water column, deep chlorophyll maximum (DCM) is a different depth at each station and year, see Table 16

Discussion

I set out to investigate *where* and *when* parasites were present in their copepod hosts. In doing this, finding out *who* the parasites were became instrumental, both in reviewing existing literature and obtaining sequences of previously unknown parasites. I found out that the parasites are present anywhere you look and seem to have both seasonal and spatial patterns in occurrence. Our data unfortunately had issues with cross-contamination, which means that samples may contain false positives. The data nevertheless has valid information content (see later discussion), but the contamination issues should be kept in mind when investigating the data.

First, I discuss the newfound sequences from the individual parasite samples before discussing the seasonal and spatial patterns in the data. Then I assess metabarcoding as a tool for studying parasitism in copepods in general and the quality of our study in particular. Finally, I discuss the future of studying parasitism in copepods and what kind of studies will be important to advance the field.

Who are the parasites? — Revisited

This study shows how limited knowledge we have about which organisms parasitize copepods. By metabarcoding 4 samples of copepods with parasites, we have detected 5 potential parasites that have no exact sequence match in any database. One of these, *Chromidina sp.*, has to my knowledge never been recorded as a copepod parasite before. The parasite described as *Ichthyophonus sp.* in earlier literature (e.g., Torgersen et al., 2002) seems to have a different taxonomic identity than previously assigned. The other parasites are known from copepods or at least other crustaceans, but their exact identity is not clear. Below I briefly discuss each of them.

Chromidina sp. infecting *Calanus sp.*

The parasite in sample 17 was identified as an apostome ciliate in the genus *Chromidina*. When discovered, a myriad of cells of this parasite was swimming around inside the body cavity of live *Calanus* copepods.[1] Parasites in the genus *Chromidina* have not been found in copepods in any published study but are known parasites of cephalopods (Souidenne et al., 2016). *Chromidina spp.* is suggested to have an arthropod intermediate host, but this connection has never been confirmed (Hochberg, 1982; cited in Souidenne et al., 2016)[2]. Interestingly, sequences of *Cromidina* were among the most abundant sequences in *Calanus pacificus*, *Metridia pacifica* and *Eucalanus bungii* in a recent master's thesis investigating copepod eukaryote microbiome by metabarcoding (Savage, 2020). Our study complements previous studies and provides the first visual observation of *Chromidina sp.* in a copepod, assuming our taxonomic assignment is correct.

The evidence together suggests that calanoid copepods act as intermediate hosts for *Chromidina spp.*. Future studies should isolate the parasite we have observed and do morphological and genetic investigations to confirm its identity. Since parasites of intermediate hosts often manipulate host behavior (Poulin, 1994), behavioral changes should be investigated in infected copepods. Such studies, coupled with transmission experiments and existing knowledge of cephalopod feeding, can shed some light on the life cycle of *Chromidina*.

Gregarine parasite in *Calanus sp.*

I am confident that the parasite in sample 56 is the parasite described as *Ichthyophonus sp.* by Torgersen et al. (2002), which in turn is consistent with the early descriptions of *Ichthysporidium* (later *Ichthyophonus*)

[1] A video of this is available at https://youtu.be/o37XVWudGuw

[2] I was unable to locate the original text, but Hochberg (1982) is cited on this by Landers (2010), Souidenne et al. (2016) and Savage (2020).

in the literature (Chatton, 1920; Jepps, 1937). However, our molecular data suggest that this parasite is an Apicomplexan in the subclass Gregarinasina (gregarines). Gregarines are mentioned as parasites on copepods multiple times throughout the literature but are described as gut-inhabiting parasites with a distinct body plan (Jepps, 1937; Sano et al., 2016), thus not resembling the hyphae-like organism we found filling the body cavity of *Calanus*. The parasite sequences had no close matches in GenBank, so it is unlikely that the parasite is related to the fish parasite *Ichthyophonus hoferi*. To remain somewhat consistent with the literature—and for lack of a more precise taxonomic assignment—the parasite will be referred to as "*Ichthyophonus*-like gregarine" for the remainder of this text.

Torgersen et al. (2002) observed that the parasite altered copepod vertical migration patterns so that large amounts of parasitized individuals were present in the surface at daytime. This corresponds with our observations, as we could reliably sample the parasite by towing plankton nets in the uppermost surface in Drøbak in October 2020. This behavioral modification—coupled with the conspicuous coloration the parasite gives—suggests that *Calanus* may be an intermediate host for the *Ichthyophonus*-like gregarine (see Introduction).

To increase our understanding of the *Ichthyophonus*-like gregarine, finding the potential final host is essential. One place to start could be to search for similar sequences in publicly available fish gut (or better, tissue) data sets to find out if any particular fish eat the parasitized copepods. Experiments investigating transmission between copepods, quantifying host manipulation, or investigating how infection affects host vital traits, will aid in understanding the parasite-host relationship. Ongoing studies indicate differences in respiration rate and pigmentation between infected and uninfected individuals (Eliassen et al., in prep).

Suctorian ciliates on *Chiridius sp.*

The epibionts on *Chiridius sp.* in sample 11 were identified as a ciliate of the subclass Suctoria. A wide range of suctorian ciliates are known epibionts on copepods (Fernandez-Leborans & Tato-Porto, 2000b), and *Chiridius spp.* are known hosts of these (Fernandez-Leborans & Tato-Porto, 2000a). These are generally considered commensals, where the ciliates use the copepods as a substrate for attachment and feed using specialized tentacles (Fernandez-Leborans & Tato-Porto, 2000a). However, growth of suctorian ciliates may affect the fitness of the copepod by increasing its surface area and thus lowering its sinking speed[3] (Weissman et al., 1993). The sequences from our samples did not have a good database match at the species level, indicating that this is a suctorian that has yet to be sequenced.

A similar epibiont has been found on *Chiridius armatus* earlier near Drøbak, i.e., the same host and location as our finding (Olsen et al., 2000). Olsen et al. (2000) found indications that infected *C. armatus* had an increased feeding rate, implying some cost for the host. Although neither Olsen et al. (2000) nor this thesis investigated morphology thoroughly, the two findings look similar in pictures (Figure 13 C-D, Olsen et al., 2000 Fig. 1). They could very well be the same epibiont, especially considering that both host and location are the same.

In the same sample, a large number of reads were assigned to *Hematodinium sp.*. This genus is known to parasitize crustaceans but is not reported from copepods (Shields, 1994). It is curious that these sequences showed up in our *Chiridius* samples, considering that the target parasite was definitely a ciliate judging only by rough morphology (compare our Figure 13 with, e.g., Figure 3 from Fernandez-Leborans & Tato-Porto, 2000b). A likely explanation is that one or more of the *Chiridius* individuals by chance were infected with a second parasite that we did not detect visually, which was an unknown *Hematodinium* parasite. Another possible explanation is that one of the *Chiridius* recently had ingested a parasitized prey animal and that the *Hematodinium* signal comes from gut content.

Ellobiopsis sp. on *Calanus sp.*

Ellobiopsis is a genus containing several copepod parasites (see Introduction). The parasite sequences in our sample only had around 94% similarity with any known *Ellobiopsis* species in Genbank. In the PR^2 database,

[3]Reminding us of the blurred lines between different modes of symbiosis (see Box 1).

they were not assigned to Ellobiopsis at all, even though several representatives exist in the database. This may indicate that the *Ellobiopsis* in the Oslofjord is a different species or variant of the *Ellobiopsis* found in, e.g., the Mediterranean Sea. For future research, at least the complete 18S gene of the Oslofjord *Ellobiopsis* should be sequenced, in conjunction with morphological investigations, to determine if it is different from *Ellobiopsis* described in the literature.

When are the parasites there?

A general answer to the question of "when?" could be "always." Of the 8 copepod parasites identified in this study, half were present in all the sampling months, and the rest were present in at least two-thirds of the months (Figure 27 A). I will discuss some specific patterns in occurrence in this section.

In general, there are fewer parasite OTUs present in May through July, and more throughout the winter, peaking in March (Figure 27 A). This is largely consistent with the observed parasite patterns from the literature (see Introduction). The low occurrence in the summer months, after the spring bloom, could be biased due to the high algal biomass in these samples (see later discussion). The number of OTUs at least shows that more diversity is detected in late summer through winter. However, it is uncertain whether this diversity stems from different species or intraspecific genetic variation (see Brown et al., 2015).

For the *Ichthyophonus*-like gregarine, there is a slight discrepancy between our results and the observations of Torgersen et al. (2002). Torgersen et al. (2002) reported the parasite to peak in late summer and be absent from October to May; our data has detected the parasite in all months but May and June. That we found the parasite in winter, while Torgersen et al. (2002) did not, may be attributed to differences in sampling methods. Torgersen et al. (2002) looked for parasites visually and primarily by surface tows, possibly only detecting heavily infected specimens. On the other hand, our method may have detected early stages of infection that are not conspicuously colored, as well as parasitized individuals not located in the surface. That our study did not detect the parasite at all in June is harder to explain. It may be because they are relatively rare with a patchy distribution (Torgersen et al., 2002) and because the biomass was very high in our summer samples (see later discussion).

Parasites often need a host threshold density to establish in a population, making peak parasite prevalence lag slightly behind peak host density (May, 1983). In our data, we can see that many of the parasite groups have peak occurrence in early autumn—around September to October (Figures 28 and 29). This pattern also goes for the larger parasitic groups Syndiniales and Gregarines (Figure 25). Many of the copepod groups have peaks around July to August both in our data (Figure 26) and earlier studies (Kiørboe & Nielsen, 1994; Zervoudaki et al., 2009). Many parasites in our data seem to respond to host density, with a time-lag similar to that described by May (1983). This is consistent with the results of Skovgaard & Saiz (2006), where *Syndinium sp.* and *Blastodinium sp.* had the same response to host abundance.

Another effect that could explain some of the seasonal variations is the effect proposed by Marshall et al. (1934): that parasites accumulate during the host's lifetime so that older individuals have a higher chance of being parasitized. This could happen if host interaction with the free-living stage of the parasite and subsequent infection is rare, and the longer an organism lives, the more opportunities for infection. The parasite could also use a long time to develop after infection so that older hosts have more developed parasites. Some copepods, e.g., *Calanus finmarchicus*, which have overwintering stages and can be relatively long-lived (Hirche, 1983), may be subject to this effect if it exists. One could speculate that the parasites with increased OTU and relative read abundances in winter—*Blastodinium sp.*, *Chromidina sp.* and *Ellobiopsis sp.* (Figures 28 and 29)[4]—parasitize overwintering *Calanus*. The *Chromidina* and *Ellobiopsis* found in our study were identified from parasitized *Calanus* individuals, which means that *Calanus* is a potential host, although not necessarily the only one.

To find out more about either of the patterns above, we need to investigate host range and specificity for the parasites and get reliable estimates of parasite prevalence. Sequencing several samples containing smaller numbers of copepods of the same species will give insight into both of these metrics, as outlined

[4]Some patterns are easier to see on a logarithmic scale; see Figure S8 in Appendix A.

earlier (Section on advantages and limitation of methods). From morphology-based studies, it seems like the parasites all have several hosts (Table 2), but molecular investigation may shed some light on this. Similarly, earlier prevalence estimates may be inaccurate due to difficulty in detection (Skovgaard & Saiz, 2006), which molecular methods can mitigate.

Where are the parasites?

Parasites were present everywhere, being found at all stations but one in our geographically narrow study in the Oslofjord and from the much broader BioMarKs project combined. There are some distinct differences between locations in both studies that I will discuss here.

The Oslofjord

From Figures 28 and 29, it is apparent that there is more seasonal variation than variation between stations in the Oslofjord. There are, however, some differences between the station IM2 (Elle) and the others. IM2 is the deepest of our stations and the only station sampled that is located outside of the Drøbak sill (Methods, p. 17). This means that it has more deep-water exchange with the ocean and typically hosts a more oceanic fauna (Baalsrud & Magnusson, 2002). It also stands out in the NMDS (Figure 23), in that it seems to have slightly less variation in species composition throughout the year. Except for occasional deviations in relative read abundances, no other station stands out regarding parasite occurrence.

The two potential parasites (or commensals, see earlier discussion) *Hematodinium sp.* and the suctorian ciliate are almost exclusively present at IM2. The sequences of these parasites were found in *Chiridius sp.*, a large, omnivorous copepod often residing in deep waters (Schøyen & Kaartvedt, 2004). Large copepods are in general most abundant in deeper waters, where they presumably can escape visual predation at daytime (Robertis, 2002), and both *Chiridius* and the even larger *Paraeuchaeta norvegica* are common at IM2 (Schøyen & Kaartvedt, 2004; Skarra & Kaartvedt, 2003). Sequences of the *Chiridius* in our individual parasite sample were assigned to the genus *Euchaeta*, which consistently had the highest relative read abundances at IM2 (Figure 26).

The observed patterns may indicate that both the tentative *Hematodinium* parasite and the suctorian ciliate are host-specific to *Chiridius*. The distribution of parasites is inevitably linked to host distribution (e.g., Hance et al., 2007). Thus, the availability of *Chiridius* may limit the distribution of the parasites. The ciliate is large and conspicuous and should be possible to detect visually even in fixed samples. We may discover more about this host-symbiont relationship when we count the conserved copepods from our fixed samples.

BioMarKs data

The sampling sites of the BioMarKs project have a much larger geographic distribution than our study (Figure 8), with a considerable variation in environmental factors. It is not unexpected, then, that the occurrence of parasites differs more between stations in BioMarKs than in our data. It is also important to consider that the BioMarKs sampling methods were different from ours, combining filtered water samples and sediment samples. Thus, they probably represent more of the free-living diversity (e.g., spores, resting stages) of copepod parasites. Due to the significant differences in both sampling strategy and number of samples from each station (Table 15), it is not easy to make direct comparisons between sites. There are, however, still some patterns in the results worth discussing.

The Varna location stands out from the others, as not a single parasite OTU was found across 25 samples. The station is in the Black Sea, which has some unique environmental conditions. It is divided into an oxygenated upper layer (around 0-50m) and an anoxic deep layer, and only connects to the ocean through the narrow Bosporos strait (Stewart et al., 2007). The water in the Black Sea is brackish and has a much lower salinity in both upper and lower layers than the other stations in the BioMarKs project (see Table

16). The Black Sea represents an environment where few, if any, studies on copepod parasites have been conducted (the closest being the adjacent Sea of Marmara, see Figure 7). This means that even if parasites are present, they are too different to match any known parasites in the PR^2 database. It is also possible—if perhaps unlikely—that there are no copepod parasites in the Black Sea. Studies need to be conducted in the Black Sea specifically, as it is so different from other systems that general parasite patterns may not apply there.

The Blanes and Naples locations had a completely different parasite composition, despite both being in the Mediterranean Sea. Interestingly, Blanes had a higher number of unique genera identified (although certainly less unique OTUs) than Naples despite the difference in sampling effort. Some of the differences, especially the higher OTU count in Naples, could be due to different sample sizes. Differences in depth, temperature and sampling month may also account for the variation in parasite presence between these two locations. Our study indicates that season can have a significant effect on parasite occurrence. Since the Blanes sampling was conducted at a different time of year than the two Naples samplings, this may be the source of some variation. Furthermore, the Blanes station is very shallow and probably has a different copepod community than the Naples station, possibly affecting which parasites are present.

The data from Oslo and Naples have around the same amount of samples, both from two sampling sessions in two different years, making them more easily comparable than the rest of the locations. The two locations have a very different composition and abundance of parasite OTUs. Most studies—and consequently most DNA sequences available—of copepod parasites are from the Mediterranean, which may account for some of these differences. Nonetheless, the Naples samples have 28 unique *Blastodinium* OTUs compared to Oslo's 1 OTU. This fits the known patterns in *Blastodinium* distribution, where several (photosynthetic) species are found at lower latitudes, while only the non-photosynthetic *Blastodinium hyalinum* can be found in northern waters (Skovgaard et al., 2012). It is also interesting to note that mostly the same genera are found in both locations, except the egg parasite *Chythriodinium* in Naples, and a single *Chromidina* OTU in Oslo. It is difficult to say whether the differences seen here are specific to the two locations, or if they represent a general pattern in occurrence. The latitudinal variation in copepod parasites needs investigation, especially by studying systems at higher latitudes, where few studies have been done so far.

Interestingly, there are some parasite OTUs detected in the sediment samples from BioMarKs. Disregarding *Ichthyophonus*—which, as discussed, is probably not the copepod parasite of Torgersen et al. (2002) —the suctorian ciliate and the genera *Chromidina* and *Syndinium* had hits in the sediment. This is not unheard of, as Cleary & Durbin (2016) similarly found that parasite sequences were abundant in the sediments in Antarctic waters, especially the groups Apicomplexa and Syndiniales. According to Anderson & May (1981), having long-lived free-living infective stages is one way for a parasite to remain in a population, even though the host density is low for large parts of the year. From our results from the Oslofjord, it appears that all three parasites in the BioMarKs sediment samples have a marked seasonality in occurrence. It could be possible that the parasites maintain their population by having resting stages in the sediment when host density is low. Another possible explanation is that the sequences from the sediment are dead parasites (or parasites of dead hosts) that have sunk to the bottom. However, many parasite sequences from the sediment were from RNA, which indicates that the organisms were alive at the time of sampling. As more sediment samples are sequenced or processed in more quantitative ways, the occurrence and function of parasites in the sediment will hopefully be better understood.

Metabarcoding as a tool

Our data shows that metabarcoding is a promising tool for studying parasites of copepods, especially regarding where and when they occur. We discovered 8 potential parasites on copepods in our study across all months and sampling stations, which might not have been found through traditional methods. In this section, I discuss our study design and how well the method is suited to this field of study, with its advantages and limitation.

Our study had cross-contamination issues, and despite efforts to remove contamination, there may still be an unknown number of false positives. At worst, false positives may lead to drawing wrong conclusions (Ficetola

et al., 2016). Contamination is often homogenous across samples, meaning that heavy contamination should make all samples more similar (McKnight et al., 2019). Our samples had clear seasonal structure (e.g. Figure 22), and similar patterns were seen across stations throughout the year (see Results, pages 38 and 41). In addition, the patterns in copepod and phytoplankton groups correspond to those from earlier literature. In conclusion, although they should not be trusted blindly, the data are still useful for investigating parasite patterns. Since sufficient material still exists from the original samples, the contamination issues can hopefully be remedied by re-sequencing.

Although the data capture the patterns in abundant groups, there is no guarantee that patterns in rare genera and species are accurate. Detection of rare species is heavily dependent on sampling effort, and when increasing sample sizes or the number of replicates, there is generally an increase in the number of rare species detected (Magurran, 2004). Detection of rare species could be an issue, especially in our samples from late spring to summer. These samples were full of chain-forming diatoms and had larger total biomass than samples from the rest of the year (not quantified). The DNA extracts are a sub-sample of a sub-sample (Methods, p. 20), and we used a smaller fraction of the total in the spring and summer sub-samples for DNA extraction. Therefore, high algal biomass may have made rare species more difficult to detect in these samples (Deagle et al., 2018). In practice, this means that for rare species, an absence in the data does not necessarily mean that the species was not present (Ficetola et al., 2015).

Since we performed the PCRs using several technical replicates, and the sequencing depth, at least at the genus level, was good (Figure 19), I conclude that detection probabilities are likely most affected by the steps prior to DNA extraction. More replicates—technical or true—should be extracted for the samples with the largest total biomass for future studies. This can aid in the detection of rare species, which most of the copepod parasites in our samples seem to be (Figure 27).

Another difficulty regarding detection in metabarcoding data is how to quantify abundance. In this study, I primarily present presence/absence data to investigate parasite occurrence patterns, with some exceptions. DNA from different groups is amplified differently in the PCR (amplification bias), and without having empiric investigations of these biases, converting read counts to biomass estimates is not straightforward (reviewed in Deagle et al., 2019). Conversely, converting read counts to presence-absence data may overestimate the importance of rare groups (Deagle et al., 2019). However, since this study specifically investigates those rare groups, it is reasonable to analyze most of our data this way.

At some points in the text, I have used relative read abundance (RRA)—i.e., reads expressed as a percentage of the total reads in the sample—to discuss some patterns in the data. When doing this, it is vital to keep in mind what this statistic tells us—and what it does not. Even if there had been no biases (in amplification or otherwise), a higher RRA does not necessarily mean more individuals or biomass. RRA is, as the name implies, *relative* compared to the other reads in the same sample. That means that if the RRA of a species is higher in one sample than another, it is presumably only more abundant compared to the other species in the sample. For a parasite-specific example of how this can lead to wrong interpretations, consider two samples where a parasite has the same prevalence in its host. If a lot of non-host organisms are present in one of the samples, the RRA of the parasite will be lower even though the true prevalence is the same. While RRA can still be somewhat helpful to investigate patterns, one should be careful with drawing too strong conclusions from it.

With the current study design, it is not possible to estimate the prevalence of parasites and host specificity, regardless of which metric is used to quantify abundance. For a more quantitive application of metabarcoding, individual species can be isolated from the samples and sequenced separately with several replicates. This design has been employed in copepod microbiome and diet studies, and provides high resolution of copepod-symbiont or copepod-prey interactions (Moisander et al., 2015; see Ray et al., 2016; Savage, 2020; Zamora-Terol et al., 2020). Another alternative is to size-fraction the sample before sequencing (see Djurhuus et al., 2018), which may give some crude information about host specificity.

We used the anti-metazoan 18SV4 primers of Bass & del Campo (2020) in this study to get the best possible taxonomic resolution of non-metazoan taxa. The universal 18SV4 primers were used for comparison reasons and to ensure that we also captured the metazoan diversity. However, most sequences were metazoan in all samples regardless of primer, probably due to the way we designed our study. Taxonomic preference of the

anti-metazoan primers could not mitigate the significant biomass difference between metazoans and other organisms in our samples (Bass & del Campo, 2020; see also Ray et al., 2016).

Neither of the primers gave sufficient taxonomic resolution for metazoans. For example, the 18SV4 marker could not separate the different species of the ecologically important genus *Calanus*. This lack of taxonomic resolution is perhaps most evident in the co-occurrence analysis, where the host in the single significant result could only be confirmed to be a calanoid copepod. Thus, something that could have given information about host specificity was obscured. Other genes, like the ribosomal 28S or the mitochondrial Cytochrome Oxidase I (COI) and Cytochrome b (Cyt b), have more interspecific variation in copepods and may provide better resolution (Blanco-Bercial et al., 2011; Djurhuus et al., 2018; Hirai et al., 2020).

In conclusion, the use of two different primer sets had no apparent benefit in this study. This is partly because the anti-metazoan primers failed to block metazoans in our samples as effectively as in the *in silico* tests of Bass & del Campo (2020). In addition, the two primer sets used amplified sequences from the same genomic region, so for the taxa that were found, the primers did not complement each other on the taxonomic assignments. A better approach for future studies would be to use an evolutionary independent gene, like COI or 28S, in conjunction with the anti-metazoan primers (Djurhuus et al., 2018; e.g., Hirai et al., 2020). Using two independent primer sets can reduce amplification biases, and supplementing 18S with a more variable gene will provide more information on the copepod communities (Blanco-Bercial et al., 2011; Drummond et al., 2015; Tang et al., 2012).

To summarize, metabarcoding is a method with many possibilities for researching parasites on copepods. It finds organisms that can be difficult to detect visually (see Skovgaard & Saiz, 2006), and processing many samples is relatively easy and quick compared to visual detection methods. Specifically, our method of bulk-processing plankton samples was fairly quick—around 15-30 minutes of pre-processing per sample—while still consistently detecting parasites in our samples. With additional considerations, like a more appropriate combination of primers and adjustments in sub-sampling for biomass differences, this method can give a good insight into long time series data with a reasonable amount of work.

Going further

There is so much uncharted terrain in the field of copepod parasites that one can do just about any study imaginable and discover something new. However, just as important as what to study is *how* to study it, i.e., what method and study design to use. I have shown here that metabarcoding is promising but that we still a few challenges to overcome and design choices to make before it can become a great way to study these systems. Here, I will outline some of the things I think will be most important going forward.

A significant hurdle for using metabarcoding at the moment is the lack of reference sequences in the databases. In our Oslofjord study, we found more parasites from the sequences we identified than those already available. Reliably sequencing as many parasites as possible will be tremendously helpful for both future metabarcoding studies and for possible metastudies searching for parasites in available data sets. This should be among the top priorities to advance our understanding of copepod-parasite systems.

When more reference sequences become available, accessing the information already available in existing metabarcoding data sets may initially be as fruitful as generating new data. I have only scratched the surface of available data by analyzing the BioMarKs data set, but it still shows the potential in re-analyzing existing data. Projects like metaPR2[5] and Ocean Barcode Atlas (Vernette et al., 2021) aims to make these kinds of studies easier and should be a natural starting point for investigating global patterns of copepod parasites.

Our study design did not provide any prevalence estimates or host specificity for the copepod parasites. More quantitative methods or study designs need to be used to obtain such estimates. One possibility is to sequence smaller numbers of isolated copepod species (see earlier discussion). Another is to use fluorescence-based methods, like those employed by Alves-de-Souza et al. (2011). They used a combination of a general DNA-binding fluorophore to identify parasitized individuals and the more targeted fluorescence in-situ hybridization

[5]Daniel Vaulot, https://daniel-vaulot.fr/project/metapr2/

(FISH) to quantify spores from the surrounding water. FISH requires parasite sequences to design probes that specifically bind to parasites in the sample, further highlighting the need to obtain sequences from as many parasites as possible.

An important question remaining, in addition to those investigated in this thesis, is: How do the parasites affect copepod individuals? This question is best investigated by experiments with live animals and parasite spores, investigating fitness costs of parasitism, modes of infection and spore fitness traits. One challenge in doing this is that one would need reliable access to live infected individuals, which requires knowledge of where and when to find them. Both Torgersen et al. (2002) and ourselves were able to collect the *Ichthyophonus*-like gregarine reliably in the surface waters of the Oslofjord, making this parasite a natural starting point for experiments.

All these proposed studies together can aid in answering the overarching question of how parasites affect zooplankton communities. Results of basic research on occurrence, prevalence, infection and spore characteristics can be used in mathematical models, attempting to give general insight into the system. Employing molecular methods like metabarcoding and FISH, supplemented by traditional visual methods, will be crucial to gain this insight.

Conclusion

The fact that parasites are everywhere all the time has implications for the copepod communities. The effects are perhaps easiest to imagine for *Syndinium* infection, which is lethal to the copepod, and *Vampyrophrya pelagica*, which also directly affects mortality. However, the presumably sterilizing *Ellobiopsis* and *Blastodinium* may also affect host densities (Anderson & May, 1981; see Lafferty & Kuris, 2009), and the same goes for the other parasites detected in this study, although their effects on their hosts are not fully (if at all) known. Kiørboe & Nielsen (1994) noted that there was an increase in copepod mortality in early fall, which coincides with the increased diversity and relative read abundance of copepod parasites around that time. One can speculate that parasitism in some way contributes to this mortality. Parasites seem to be ubiquitous in the zooplankton, and the question for the future is *how much*—not *if*—these parasites affect some of the most abundant animals on Earth.

References

Aho, A. V., Kernighan, B. W., & Weinberger, P. J. (1987). *The AWK programming language*. Addison-Wesley Longman Publishing Co., Inc.

Albaina, A., & Irigoien, X. (2006). Fecundity limitation of Calanus helgolandicus, by the parasite Ellobiopsis sp. *Journal of Plankton Research*, *28*(4), 413–418. https://doi.org/10.1093/plankt/fbi129

Alves-de-Souza, C., Cornet, C., Nowaczyk, A., Gasparini, S., Skovgaard, A., & Guillou, L. (2011). Blastodinium spp. Infect copepods in the ultra-oligotrophic marine waters of the Mediterranean Sea. *Biogeosciences*, *8*(8), 2125–2136. https://doi.org/10.5194/bg-8-2125-2011

Andersen, S. B., Gerritsma, S., Yusah, K. M., Mayntz, D., Hywel-Jones, N. L., Billen, J., Boomsma, J. J., & Hughes, D. P. (2009). The Life of a Dead Ant: The Expression of an Adaptive Extended Phenotype. *The American Naturalist*, *174*(3), 424–433. https://doi.org/10.1086/603640

Anderson, R. M., & May, R. M. (1981). The population dynamics of microparasites and their invertebrate hosts. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *291*(1054), 451–524. https://doi.org/10.1098/rstb.1981.0005

Andrews, S. (2010). *FastQC: A Quality Control Tool for High Throughput Sequence Data.*

Artüz, M. L. (2016). Parasites (Ellobiopsis chattoni Caullery, 1910) on Copepoda with two new host records, from Sea of Marmara, Turkey. *Marine Biodiversity Records*, *9*(1), 11. https://doi.org/10.1186/s41200-016-0006-9

Auld, S. K., & Tinsley, M. C. (2015). The evolutionary ecology of complex lifecycle parasites: Linking phenomena with mechanisms. *Heredity*, *114*(2), 125–132. https://doi.org/10.1038/hdy.2014.84

Baalsrud, K., & Magnusson, J. (2002). Indre Oslofjord: Natur og miljø. *Fagrådet for Vann Og Avløpsteknisk Samarbeid i Indre Oslofjord, Oslo.*

Bagge, L. E. (2019). Not As Clear As It May Appear: Challenges Associated with Transparent Camouflage in the Ocean. *Integrative and Comparative Biology*, *59*(6), 1653–1663. https://doi.org/10.1093/icb/icz066

Baker, D. M., Freeman, C. J., Wong, J. C. Y., Fogel, M. L., & Knowlton, N. (2018). Climate change promotes parasitism in a coral symbiosis. *Isme Journal*, *12*(3), 921–930. https://doi.org/10.1038/s41396-018-0046-8

Bass, D., & del Campo, J. (2020). Microeukaryotes in animal and plant microbiomes: Ecologies of disease? *European Journal of Protistology*, *76*, 125719. https://doi.org/10.1016/j.ejop.2020.125719

Blanco-Bercial, L., Bradford-Grieve, J., & Bucklin, A. (2011). Molecular phylogeny of the Calanoida (Crustacea: Copepoda). *Molecular Phylogenetics and Evolution*, *59*(1), 103–113. https://doi.org/10.1016/j.ympev.2011.01.008

Briseño-Avena, C. (2015). *Fine-scale spatial and temporal plankton distributions in the Southern California Bight : Lessons from in situ microscopes and broadband echosounders.* [PhD thesis]. UC San Diego.

Brown, E. A., Chain, F. J. J., Crease, T. J., MacIsaac, H. J., & Cristescu, M. E. (2015). Divergence thresholds and divergent biodiversity estimates: Can metabarcoding reliably describe zooplankton communities? *Ecology and Evolution*, *5*(11), 2234–2251. https://doi.org/10.1002/ece3.1485

Buermans, H. P. J., & den Dunnen, J. T. (2014). Next generation sequencing technology: Advances and applications. *Biochimica Et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1842*(10), 1932–1941. https://doi.org/10.1016/j.bbadis.2014.06.015

Cattley, J. G. (1948). Sex reversal in copepods. *Nature*, *161*(4102), 937–937. https://doi.org/10.1038/161937a0

Chatton, É. (1920). Les Peridiniens parasites : Morphologie, reproduction, ethologie. *Arch. Zool. Exp. Gen.*, *59*, 1–475.

Chen, H. (2018). *VennDiagram: Generate high-resolution venn and euler plots*. https://CRAN.R-project.org/package=VennDiagram

Clare, E. L., Chain, F. J. J., Littlefair, J. E., & Cristescu, M. E. (2016). The effects of parameter choice on defining molecular operational taxonomic units and resulting ecological analyses of metabarcoding Data1. *Genome*. https://doi.org/10.1139/gen-2015-0184

Cleary, A. C., & Durbin, E. G. (2016). Unexpected prevalence of parasite 18S rDNA sequences in winter among Antarctic marine protists. *Journal of Plankton Research*, *38*(3), 401–417. https://doi.org/10.1093/plankt/fbw005

Coats, D., Bachvaroff, T., Handy, S., Kim, S., Gárate-Lizárraga, I., & Delwiche, C. (2008). Prevalence and phylogeny of parasitic dinoflagellates (genus Blastodinium) infecting copepods in the Gulf of California. *CICIMAR Oceánides*, *23*(1-2), 67–77.

Colebrook, J. M. (1979). Continuous Plankton Records: Seasonal cycles of phytoplankton and copepods in the North Atlantic ocean and the North Sea. *Marine Biology*, *51*(1), 23–32. https://doi.org/10.1007/BF00389027

Combes, C. (1996). Parasites, biodiversity and ecosystem stability. *Biodiversity & Conservation*, *5*(8), 953–962. https://doi.org/10.1007/bf00054413

Das, S., Mohapatra, S. C., & Hsu, J. T. (1999). Studies on primer-dimer formation in polymerase chain reaction (PCR). *Biotechnology Techniques*, *13*(10), 643–646. https://doi.org/10.1023/A:1008924615839

de Bary, H. A. (1879). Die Erscheinung der Symbiose. *Strasbourg, France*.

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

de Vargas, C., Audic, S., Tara Oceans Consortium, C., & Tara Oceans Expedition, P. (2017). *Total V9 rDNA information organized at the OTU level for the Tara Oceans Expedition (2009-2012)*. PANGAEA. https://doi.org/10.1594/PANGAEA.873275

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

Deagle, B. E., Thomas, A. C., McInnes, J. C., Clarke, L. J., Vesterinen, E. J., Clare, E. L., Kartzinel, T. R., & Eveson, J. P. (2019). Counting with DNA in metabarcoding studies: How should we convert sequence reads to dietary data? *Molecular Ecology*, *28*(2), 391–406. https://doi.org/10.1111/mec.14734

del Campo, J., Pons, M. J., Herranz, M., Wakeman, K. C., Valle, J. del, Vermeij, M. J. A., Leander, B. S., & Keeling, P. J. (2019). Validation of a universal set of primers to study animal-associated microeukaryotic communities. *Environmental Microbiology*, *21*(10), 3855–3861. https://doi.org/10.1111/1462-2920.14733

Dexter, E., Rollwagen-Bollens, G., & Bollens, S. M. (2018). The trouble with stress: A flexible method for the evaluation of nonmetric multidimensional scaling. *Limnology and Oceanography: Methods*, *16*(7), 434–443. https://doi.org/10.1002/lom3.10257

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

Drits, A., & Semenova, T. (1985). Some aspects of interrelations between marine copepods with algae of the genus Blastodinium (Peridinales). In *Zoologichesky Zhurnal* (Vol. 64, pp. 453–455). Mezhdunarodnaya Kniga 39 Dimitrova ul., 113095 Moscow, Russia.

Drummond, A. J., Newcomb, R. D., Buckley, T. R., Xie, D., Dopheide, A., Potter, B. C., Heled, J., Ross, H. A., Tooman, L., Grosser, S., Park, D., Demetras, N. J., Stevens, M. I., Russell, J. C., Anderson, S. H., Carter, A., & Nelson, N. (2015). Evaluating a multigene environmental DNA approach for biodiversity assessment. *GigaScience*, *4*(s13742-015-0086-1). https://doi.org/10.1186/s13742-015-0086-1

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Fahmi, A. M., & Hussain, M. (2003). Two groups of parasites on the body of copepods. *Journal of Arid Environments*, *54*(1), 149–153. https://doi.org/10.1006/jare.2001.0883

Fernandez-Leborans, G., & Tato-Porto, M. L. (2000a). A Review of the Species of Protozoan Epibionts on Crustaceans. Ii. Suctorian Ciliates. *Crustaceana*, *73*(10), 1205–1237.

Fernandez-Leborans, G., & Tato-Porto, M. L. (2000b). Presence of two suctorian ciliate species of the genera Acineta and Pelagacineta, epibionts on marine copepod crustaceans. *Cytobios*, *103*(404), 139–148.

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet-Covex, C., Barba, M. D., Gielly, L., Lopes, C. M., Boyer, F., Pompanon, F., Rayé, G., & Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*, *15*(3), 543–556. https://doi.org/10.1111/1755-0998.12338

Ficetola, G. F., Taberlet, P., & Coissac, E. (2016). How to limit false positives in environmental DNA and metabarcoding? *Molecular Ecology Resources*, *16*(3), 604–607. https://doi.org/10.1111/1755-0998.12508

Fields, D. M., Runge, J. A., Thompson, C., Shema, S. D., Bjelland, R. M., Durif, C. M. F., Skiftesvik, A. B., & Browman, H. I. (2015). Infection of the planktonic copepod Calanus finmarchicus by the parasitic dinoflagellate, Blastodinium spp: Effects on grazing, respiration, fecundity and fecal pellet production. *Journal of Plankton Research*, *37*(1), 211–220. https://doi.org/10.1093/plankt/fbu084

Floyd, R., Abebe, E., Papert, A., & Blaxter, M. (2002). Molecular barcodes for soil nematode identification. *Molecular Ecology*, *11*(4), 839–850. https://doi.org/10.1046/j.1365-294X.2002.01485.x

Frøslev, T. G. (2021). *Lulu: Post clustering curation of amplicon data for reliable biodiversity metrics*. https://github.com/tobiasgf/lulu

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

Fumagalli, M., Sironi, M., Pozzoli, U., Ferrer-Admettla, A., Pattini, L., & Nielsen, R. (2011). Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution. *PLOS Genetics*, *7*(11), e1002355. https://doi.org/10.1371/journal.pgen.1002355

Gallienne, C. P., & Robins, D. B. (2001). Is Oithona the most important copepod in the world's oceans? *Journal of Plankton Research*, *23*(12), 1421–1432. https://doi.org/10.1093/plankt/23.12.1421

Gifford, D. J. (1991). The Protozoan-Metazoan Trophic Link In Pelagic Ecosystems. *The Journal of Protozoology*, *38*(1), 81–86. https://doi.org/10.1111/j.1550-7408.1991.tb04806.x

Gleichsner, A. M., Cleveland, J. A., & Minchella, D. J. (2016). One stimulus two responses: Host and parasite life-history variation in response to environmental stress. *Evolution*, *70*(11), 2640–2646. https://doi.org/10.1111/evo.13061

Gómez, F. (2012). A quantitative review of the lifestyle, habitat and trophic diversity of dinoflagellates (Dinoflagellata, Alveolata). *Systematics and Biodiversity*, *10*(3), 267–275. https://doi.org/10.1080/14772000.2012.721021

Gómez, F., López-García, P., Nowaczyk, A., & Moreira, D. (2009). The crustacean parasites Ellobiopsis Caullery, 1910 and Thalassomyces Niezabitowski, 1913 form a monophyletic divergent clade within the Alveolata. *Systematic Parasitology*, *74*(1), 65–74. https://doi.org/10.1007/s11230-009-9199-1

Gómez, F., Moreira, D., & López-García, P. (2009). Life cycle and molecular phylogeny of the dinoflagellates Chytriodinium and Dissodinium, ectoparasites of copepod eggs. *European Journal of Protistology*, *45*(4), 260–270. https://doi.org/10.1016/j.ejop.2009.05.004

Gran-Stadniczeñko, S., Egge, E., Hostyeva, V., Logares, R., Eikrem, W., & Edvardsen, B. (2019). Protist Diversity and Seasonal Dynamics in Skagerrak Plankton Communities as Revealed by Metabarcoding and Microscopy. *Journal of Eukaryotic Microbiology*, *66*(3), 494–513. https://doi.org/10.1111/jeu.12700

Grey, E. K., Bernatchez, L., Cassey, P., Deiner, K., Deveney, M., Howland, K. L., Lacoursière-Roussel, A., Leong, S. C. Y., Li, Y., Olds, B., Pfrender, M. E., Prowse, T. A. A., Renshaw, M. A., & Lodge, D. M. (2018). Effects of sampling effort on biodiversity patterns estimated from environmental DNA metabarcoding surveys. *Scientific Reports*, *8*(1), 8843. https://doi.org/10.1038/s41598-018-27048-2

Grimes, B., & Bradbury, P. (1992). The Biology of Vampyrophrya-Pelagica (chatton and Lwoff, 1930), a Histophagous Apostome Ciliate Associated with Marine Calanoid Copepods. *Journal of Protozoology*, *39*(1), 65–79. https://doi.org/10.1111/j.1550-7408.1992.tb01285.x

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., del Campo, J., Dolan, J. R., Dunthorn, M., Edvardsen, B., Holzmann, M., Kooistra, W. H. C. F., Lara, E., Le Bescot, N., Logares, R., … Christen, R. (2013). The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, *41*(Database issue), D597–D604. https://doi.org/10.1093/nar/gks1160

Guiry, M. D., & Guiry, G. M. (2020). *AlgaeBase. World-wide electronic publication, National University of Ireland, Galway (taxonomic information republished from AlgaeBase with permission of M.D. Guiry). Blastodinium Chatton, 1906.* http://www.marinespecies.org/aphia.php?p=taxdetails&id=562641.

Hance, T., van Baaren, J., Vernon, P., & Boivin, G. (2007). Impact of extreme temperatures on parasitoids in a climate change perspective. In *Annual Review of Entomology* (Vol. 52, pp. 107–126). Annual Reviews. https://doi.org/10.1146/annurev.ento.52.110405.091333

Heck, K. L., Belle, G. van, & Simberloff, D. (1975). Explicit Calculation of the Rarefaction Diversity Measurement and the Determination of Sufficient Sample Size. *Ecology*, *56*(6), 1459–1461. https://doi.org/10.2307/1934716

Hirai, J., Tachibana, A., & Tsuda, A. (2020). Large-scale metabarcoding analysis of epipelagic and mesopelagic copepods in the Pacific. *PLOS ONE*, *15*(5), e0233189. https://doi.org/10.1371/journal.pone.0233189

Hirche, H.-J. (1983). Overwintering of Calanus finmarchicus and Calanus helgolandicus. *Marine Ecology Progress Series*, *11*, 281–290. https://doi.org/10.3354/meps011281

Hochberg, F. G. (1982). The" kidneys" of cephalopods: A unique habitat for parasites. *Malacol*, *23*, 121–134.

Hoffman, E. G., & Yancey, R. M. (1966). *Ellobiopsidae of Alaskan coastal waters.*

Horiguchi, T., Harada, A., & Ohtsuka, S. (2006). Taxonomic studies on parasitic dinoflagellates in Japan. *Bulletin of the Plankton Society of Japan*, *53*(1), 21–29.

Ianora, A., Dicarlo, B. S., Mazzocchi, M. G., & Mascellaro, P. (1990). Histomorphological changes in the reproductive condition of parasitized marine planktonic copepods. *Journal of Plankton Research*, *12*(2), 249–258. https://doi.org/10.1093/plankt/12.2.249

Ianora, A., Mazzocchi, M. G., & Scotto di Carlo, B. (1987). Impact of parasitism and intersexuality on Mediterranean populations of Paracalanus parvus (Copepoda: Calanoida). *Diseases of Aquatic Organisms*, *3*(1), 29–36.

Illumina. (2021a). *MiSeq Specifications | Key performance parameters.* https://www.illumina.com/systems/sequencing-platforms/miseq/specifications.html.

Illumina. (2021b). *An Introduction to Next-Generation Sequencing Technology.*

Jepps, M. W. (1937). Memoirs: On the Protozoan Parasites of Calanus Finmarchicus in the Clyde Sea Area. *Journal of Cell Science*, *s2-79*(316), 589–658.

Kanagawa, T. (2003). Bias and artifacts in multitemplate polymerase chain reactions (PCR). *Journal of Bioscience and Bioengineering*, *96*(4), 317–323.

Karsenti, E., Acinas, S. G., Bork, P., Bowler, C., Vargas, C. D., Raes, J., Sullivan, M., Arendt, D., Benzoni, F., Claverie, J.-M., Follows, M., Gorsky, G., Hingamp, P., Iudicone, D., Jaillon, O., Kandels-Lewis, S., Krzic, U., Not, F., Ogata, H., … Consortium, the T. O. (2011). A Holistic Approach to Marine Eco-Systems Biology. *PLOS Biology*, *9*(10), e1001177. https://doi.org/10.1371/journal.pbio.1001177

Kelley, D., & Richards, C. (2020). *Oce: Analysis of oceanographic data.* https://dankelley.github.io/oce

Kimmerer, W. J., & McKinnon, A. D. (1990). High mortality in a copepod population caused by a parasitic dinoflagellate. *Marine Biology*, *107*(3), 449–452. https://doi.org/10.1007/bf01313428

Kiørboe, T. (1997). Population regulation and role of mesozooplankton in shaping marine pelagic food webs. *Hydrobiologia*, *363*(1), 13–27. https://doi.org/10.1023/A:1003173721751

Kiørboe, T. (2011). What makes pelagic copepods so successful? *Journal of Plankton Research*, *33*(5), 677–685. https://doi.org/10.1093/plankt/fbq159

Kiørboe, T., & Nielsen, T. G. (1994). Regulation of zooplankton biomass and production in a temperate, coastal ecosystem. 1. Copepods. *Limnology and Oceanography*, *39*(3), 493–507. https://doi.org/10.4319/lo.1994.39.3.0493

Kofoid, C. A. (1931). Report of the biological survey of Mutsu Bay. 18. Protozoan fauna of Mutsu Bay. Subclass Dinoflagellata ; tribe Gymnodinioidae. *Tohoku Imp. Univ. Sci. Rep. Ser. 4*, *6*, 1–43.

Krause, J. W., Brzezinski, M. A., Largier, J. L., McNair, H. M., Maniscalco, M., Bidle, K. D., Allen, A. E., & Thamatrakoln, K. (2020). The interaction of physical and biological factors drives phytoplankton spatial distribution in the northern California Current. *Limnology and Oceanography*, *65*(9), 1974–1989. https://doi.org/10.1002/lno.11431

Kuris, A. M. (1974). Trophic Interactions: Similarity of Parasitic Castrators to Parasitoids. *The Quarterly Review of Biology*, *49*(2), 129–148. https://doi.org/10.1086/408018

Lafferty, K. D., Allesina, S., Arim, M., Briggs, C. J., De Leo, G., Dobson, A. P., Dunne, J. A., Johnson, P. T. J., Kuris, A. M., Marcogliese, D. J., Martinez, N. D., Memmott, J., Marquet, P. A., McLaughlin, J. P., Mordecai, E. A., Pascual, M., Poulin, R., & Thieltges, D. W. (2008). Parasites in food webs: The ultimate missing links. *Ecology Letters*, *11*(6), 533–546. https://doi.org/10.1111/j.1461-0248.2008.01174.x

Lafferty, K. D., Dobson, A. P., & Kuris, A. M. (2006). Parasites dominate food web links. *Proceedings of the National Academy of Sciences*, *103*(30), 11211–11216. https://doi.org/10.1073/pnas.0604755103

Lafferty, K. D., & Kuris, A. M. (2009). Parasitic castration: The evolution and ecology of body snatchers. *Trends in Parasitology*, *25*(12), 564–572. https://doi.org/10.1016/j.pt.2009.09.003

Lagadeuc, Y., Bouté, M., & Dodson, J. J. (1997). Effect of vertical mixing on the vertical distribution of copepods in coastal waters. *Journal of Plankton Research*, *19*(9), 1183–1204. https://doi.org/10.1093/plankt/19.9.1183

Landers, S. C. (2010). The fine structure of the tropho-tomont of the parasitic apostome Chromidina (Ciliophora, Apostomatida). *Protistology*, *6*(4), 271–279.

Leinaas, H. P., Jalal, M., Gabrielsen, T. M., & Hessen, D. O. (2016). Inter- and intraspecific variation in body- and genome size in calanoid copepods from temperate and arctic waters. *Ecology and Evolution*, *6*(16), 5585–5595. https://doi.org/10.1002/ece3.2302

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron, S., Ignacio-Espinosa, J. C., Roux, S., Vincent, F., Bittner, L., Darzi, Y., Wang, J., Audic, S., Berline, L., Bontempi, G., Cabello,

A. M., Coppola, L., Cornejo-Castillo, F. M., … Raes, J. (2015). Determinants of community structure in the global plankton interactome. *Science*, *348*(6237), 1262073. https://doi.org/10.1126/science.1262073

Loeng, H., & Drinkwater, K. (2007). An overview of the ecosystems of the Barents and Norwegian Seas and their response to climate variability. *Deep Sea Research Part II: Topical Studies in Oceanography*, *54*(23), 2478–2500. https://doi.org/10.1016/j.dsr2.2007.08.013

Logares, R., Audic, S., Santini, S., Pernice, M. C., de Vargas, C., & Massana, R. (2012). Diversity patterns and activity of uncultured marine heterotrophic flagellates unveiled with pyrosequencing. *The ISME Journal*, *6*(10), 1823–1833. https://doi.org/10.1038/ismej.2012.36

López-Escardó, D., Paps, J., de Vargas, C., Massana, R., Ruiz-Trillo, I., & del Campo, J. (2018). Metabarcoding analysis on European coastal samples reveals new molecular metazoan diversity. *Scientific Reports*, *8*(1), 9106. https://doi.org/10.1038/s41598-018-27509-8

Lundsør, E., Stige, L. C., Sørensen, K., & Edvardsen, B. (2020). Long-term coastal monitoring data show nutrient-driven reduction in chlorophyll. *Journal of Sea Research*, *164*, 101925. https://doi.org/10.1016/j.seares.2020.101925

Magurran, A. E. (2004). *Measuring biological diversity*. Blackwell Publishing.

Mahé, F., Rognes, T., Quince, C., Vargas, C. de, & Dunthorn, M. (2014). Swarm: Robust and fast clustering method for amplicon-based studies. *PeerJ*, *2*, e593. https://doi.org/10.7717/peerj.593

Mahé, F., Rognes, T., Quince, C., Vargas, C. de, & Dunthorn, M. (2015). Swarm v2: Highly-scalable and high-resolution amplicon clustering. *PeerJ*, *3*, e1420. https://doi.org/10.7717/peerj.1420

Marshall, S. M., Nicholls, A. G., & Orr, A. P. (1934). On the biology of Calanus finmarchicus. V. Seasonal distribution, size, weight and chemical composition in Loch Striven in 1933, and their relation to the phytoplankton. *Journal of the Marine Biological Association of the United Kingdom*, *19*(2), 793–828.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Mauchline, J. (1998). *The biology of calanoid copepods* (J. H. Blaxter, B. Douglas, & P. A. Tyler, Eds.; Vol. 33). Academic Press.

May, R. M. (1983). Parasitic Infections as Regulators of Animal Populations: The dynamic relationship between parasites and their host populations offers clues to the etiology and control of infectious disease. *American Scientist*, *71*(1), 36–45.

McKnight, D. T., Huerlimann, R., Bower, D. S., Schwarzkopf, L., Alford, R. A., & Zenger, K. R. (2019). microDecon: A highly accurate read-subtraction tool for the post-sequencing removal of contamination in metabarcoding studies. *Environmental DNA*, *1*(1), 14–25. https://doi.org/10.1002/edn3.11

Minchella, D. J. (1985). Host life-history variation in response to parasitism. *Parasitology*, *90*(1), 205–216. https://doi.org/10.1017/S0031182000049143

Moisander, P. H., Sexton, A. D., & Daley, M. C. (2015). Stable Associations Masked by Temporal Variability in the Marine Copepod Microbiome. *PLOS ONE*, *10*(9), e0138967. https://doi.org/10.1371/journal.pone.0138967

NCBI Resource Coordinators. (2018). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, *46*(D1), D8–D13. https://doi.org/10.1093/nar/gkx1095

Nickrent, D. L., & Sargent, M. L. (1991). An overview of the secondary structure of the V4 region of eukaryotic small-subunit ribosomal RNA. *Nucleic Acids Research*, *19*(2), 227–235. https://doi.org/10.1093/nar/19.2.227

Nowosad, J. (2019). *Rcartocolor: CARTOColors palettes*. https://github.com/Nowosad/rcartocolor

Ohtsuka, S., Hora, M., Suzaki, T., Arikawa, M., Omura, G., & Yamada, K. (2004). Morphology and host-specificity of the apostome ciliate Vampyrophrya pelagica infecting pelagic copepods in the Seto Inland Sea, Japan. *Marine Ecology Progress Series*, *282*, 129–142.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2020). *Vegan: Community ecology package*. https://CRAN.R-project.org/package=vegan

Olsen, E. M., Jørstad, T., & Kaartvedt, S. (2000). The feeding strategies of two large marine copepods. *Journal of Plankton Research*, *22*(8), 1513–1528. https://doi.org/10.1093/plankt/22.8.1513

Orenstein, E. C. (2018). *Automated analysis of oceanographic image data* [PhD thesis]. UC San Diego.

Oulhen, N., Schulz, B. J., & Carrier, T. J. (2016). English translation of Heinrich Anton de Bary's 1878 speech, ' Die Erscheinung der Symbiose ' (' De la symbiose '). *Symbiosis*, *69*(3), 131–139. https://doi.org/10.1007/s13199-016-0409-8

Pasternak, A. F., Arashkevich, Y. G., & Sorokin, Y. S. (1984). The role of the parasitic algal genus Blastodinium in the ecology of planktic copepods. *Oceanology*, *24*, 748–751.

Pebesma, E. (2020). *Sf: Simple features for r*. https://CRAN.R-project.org/package=sf

Poulin, R. (1994). The evolution of parasite manipulation of host behaviour: A theoretical analysis. *Parasitology*, *109*(S1), S109–S118. https://doi.org/10.1017/S0031182000085127

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1), D590–D596. https://doi.org/10.1093/nar/gks1219

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ray, J. L., Althammer, J., Skaar, K. S., Simonelli, P., Larsen, A., Stoecker, D., Sazhin, A., Ijaz, U. Z., Quince, C., & Nejstgaard, J. C. (2016). Metabarcoding and metabolome analyses of copepod grazing reveal feeding preference and linkage to metabolite classes in dynamic microbial plankton communities. *Molecular Ecology*, *25*(21), 5585–5602.

Robertis, A. D. (2002). Size-dependent visual predation risk and the timing of vertical migration: An optimization model. *Limnology and Oceanography*, *47*(4), 925–933. https://doi.org/10.4319/lo.2002.47.4.0925

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, *4*, e2584. https://doi.org/10.7717/peerj.2584

RStudio Team. (2021). *RStudio: Integrated development environment for r*. RStudio, PBC. http://www.rstudio.com/

Sanders, R., Henson, S. A., Koski, M., De La Rocha, C. L., Painter, S. C., Poulton, A. J., Riley, J., Salihoglu, B., Visser, A., Yool, A., Bellerby, R., & Martin, A. P. (2014). The Biological Carbon Pump in the North Atlantic. *Progress in Oceanography*, *129*, 200–218. https://doi.org/10.1016/j.pocean.2014.05.005

Sano, M., Miyamoto, H., & Nishida, S. (2016). Thiriotia euchirellae n. Sp., A new gregarine species (Apicomplexa: Eugregarinorida) from the mesopelagic copepod Euchirella rostrata in Sagami Bay, Japan. *Marine Biodiversity*, *46*(4), 753–760. https://doi.org/10.1007/s12526-015-0423-8

Santhakumari, V., & Saraswathy, M. (1979). On the ellobiopsidae, parasitic protozoa from zooplankton. *Mahasagar*, *12*(2), 83–92.

Savage, R.-L. (2020). Characterizing the diversity of the eukaryotic microbiome in marine crustacean zooplankton. (Master's thesis). *Electronic Thesis and Dissertation Repository*.

Schrader, C., Schielke, A., Ellerbroek, L., & Johne, R. (2012). PCR inhibitors occurrence, properties and removal. *Journal of Applied Microbiology*, *113*(5), 1014–1026. https://doi.org/10.1111/j.1365-2672.2012.05384.x

Schøyen, M., & Kaartvedt, S. (2004). Vertical distribution and feeding of the copepod Chiridius armatus. *Marine Biology*, *145*(1), 159–165. https://doi.org/10.1007/s00227-004-1294-x

Sewell, R. B. S. (1951). The epibionts and parasites of the planktonic Copepoda of the Arabian Sea. *Scientific Reports of the Murray Expedition.*, *9*(4), 255–394.

Shields, J. D. (1994). The parasitic dinoflagellates of marine crustaceans. *Annual Review of Fish Diseases*, *4*(0), 241–271. https://doi.org/10.1016/0959-8030(94)90031-0

Skarra, H., & Kaartvedt, S. (2003). Vertical distribution and feeding of the carnivorous copepod Paraeuchaeta norvegica. *Marine Ecology Progress Series*, *249*, 215–222. https://doi.org/10.3354/meps249215

Skovgaard, A. (2005). Infection with the dinoflagellate parasite Blastodinium spp. In two Mediterranean copepods. *Aquatic Microbial Ecology*, *38*(1), 93–101. https://doi.org/10.3354/ame038093

Skovgaard, A. (2014). Dirty tricks in the plankton: Diversity and role of marine parasitic protists. *Acta Protozool*, *53*(1), 51–62.

Skovgaard, A., & Daugbjerg, N. (2008). Identity and Systematic Position of Paradinium poucheti and Other Paradinium-Like Parasites of Marine Copepods Based on Morphology and Nuclear-Encoded SSU rDNA. *Protist*, *159*(3), 401–413. https://doi.org/10.1016/j.protis.2008.02.003

Skovgaard, A., Karpov, S. A., & Guillou, L. (2012). The parasitic dinoflagellates Blastodinium spp. Inhabiting the gut of marine, planktonic copepods: Morphology, ecology, and unrecognized species diversity. *Frontiers in Microbiology*, *3*. https://doi.org/10.3389/fmicb.2012.00305

Skovgaard, A., Massana, R., Balagué, V., & Saiz, E. (2005). Phylogenetic Position of the Copepod-Infesting Parasite Syndinium turbo (Dinoflagellata, Syndinea). *Protist*, *156*(4), 413–423. https://doi.org/10.1016/j.protis.2005.08.002

Skovgaard, A., Massana, R., & Saiz, E. (2007). Parasitic species of the genus Blastodinium (Blastodiniphyceae) are peridinioid dinoflagellates. *Journal of Phycology*, *43*(3), 553–560. https://doi.org/10.1111/j.1529-8817.2007.00338.x

Skovgaard, A., & Saiz, E. (2006). Seasonal occurrence and role of protistan parasites in coastal marine zooplankton. *Marine Ecology Progress Series*, *327*, 37–49. https://doi.org/10.3354/meps327037

Skovgaard, A., & Salomonsen, X. M. (2009). Blastodinium galatheanum sp nov (Dinophyceae) a parasite of the planktonic copepod Acartia negligens (Crustacea, Calanoida) in the central Atlantic Ocean. *European Journal of Phycology*, *44*(3), 425–438. https://doi.org/10.1080/09670260902878743

Solberg, I., Røstad, A., & Kaartvedt, S. (2015). Ecology of overwintering sprat (Sprattus sprattus). *Progress in Oceanography*, *138*, 116–135. https://doi.org/10.1016/j.pocean.2015.08.003

Sorensen, R. E., & Minchella, D. J. (2001). Snailtrematode life history interactions: Past trends and future directions. *Parasitology*, *123*(7), S3–S18. https://doi.org/10.1017/S0031182001007843

Souidenne, D., Florent, I., Dellinger, M., Justine, J. L., Romdhane, M. S., Furuya, H., & Grellier, P. (2016). Diversity of apostome ciliates, Chromidina spp. (Oligohymenophorea, Opalinopsidae), parasites of cephalopods of the Mediterranean Sea. *Parasite*, *23*. https://doi.org/10.1051/parasite/2016033

South, A. (2017). *Rnaturalearth: World map data from natural earth*. https://github.com/ropenscilabs/rnaturalearth

Soyer, M.-O. (1970). Etude ultrastructurale de l'endoplasme et des vacuoles chez deux types de Dinoflagellés appartenant aux genres Noctiluca (Suriray) et Blastodinium (Chatton). *Zeitschrift für Zellforschung Und Mikroskopische Anatomie*, *105*(3), 350–388. https://doi.org/10.1007/BF00335462

Stewart, K., Kassakian, S., Krynytzky, M., DiJulio, D., & Murray, J. W. (2007). Oxic, suboxic, and anoxic conditions in the Black Sea. In *The Black Sea flood question: Changes in coastline, climate, and human settlement* (pp. 1–21). Springer.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H.-W., & Richards, T. A. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology*, *19*(s1), 21–31. https://doi.org/10.1111/j.1365-294X.2009.04480.x

Sunagawa, S., Acinas, S. G., Bork, P., Bowler, C., Eveillard, D., Gorsky, G., Guidi, L., Iudicone, D., Karsenti, E., Lombard, F., Ogata, H., Pesant, S., Sullivan, M. B., Wincker, P., & Vargas, C. de. (2020). Tara Oceans: Towards global ocean ecosystems biology. *Nature Reviews Microbiology*, *18*(8), 428–445. https://doi.org/10.1038/s41579-020-0364-5

Tan, G., Opitz, L., Schlapbach, R., & Rehrauer, H. (2019). Long fragments achieve lower base quality in Illumina paired-end sequencing. *Scientific Reports*, *9*(1), 2856. https://doi.org/10.1038/s41598-019-39076-7

Tang, C. Q., Leasi, F., Obertegger, U., Kieneke, A., Barraclough, T. G., & Fontaneto, D. (2012). The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna. *Proceedings of the National Academy of Sciences*, *109*(40), 16208–16212. https://doi.org/10.1073/pnas.1209160109

Thomas, F., & Poulin, R. (1998). Manipulation of a mollusc by a trophically transmitted parasite: Convergent evolution or phylogenetic inheritance? *Parasitology*, *116*(5), 431–436. https://doi.org/10.1017/S003118209800239X

Timofeev, S. (2002). Effect of the parasitic dinoflagellate Ellobiopsis chattoni (Protozoa: Mastigophora) on the winter mortality of the calanoid copepod Calanus finmarchicus (Crustacea: Copepoda) in the Norwegian Sea. *Parazitologiia*, *36*(2), 158–162.

Tompkins, D. M., Dobson, A. P., Arneberg, P., Begon, M. E., Cattadori, I. M., Greenman, J. V., Heesterbeek, J. A. P., Hudson, P. J., Newborn, D., & Pugliese, A. (2002). Parasites and host population dynamics. *The Ecology of Wildlife Diseases*, 45–62.

Torgersen, T., Karlsbakk, E., & Kaartvedt, S. (2002). Deviating vertical distribution and increased conspicuousness of parasitized Calanus. *Limnology and Oceanography*, *47*(4), 1187–1191. https://doi.org/10.4319/lo.2002.47.4.1187

van der Loos, L. M., & Nijland, R. (2020). Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. *Molecular Ecology*.

Vane, F. R. (1952). The distribution of Blastodinium hyalinum in the North Sea. *Challenger Soc*, *3*, 23–24.

Vaulot, D. (2021). *pr2database: PR2 database*.

Verity, P., & Smetacek, V. (1996). Organism life cycles, predation, and the structure of marine pelagic ecosystems. *Marine Ecology Progress Series*, *130*, 277–293. https://doi.org/10.3354/meps130277

Vernette, C., Henry, N., Lecubin, J., Vargas, C. de, Hingamp, P., & Lescot, M. (2021). The Ocean barcode atlas: A web service to explore the biodiversity and biogeography of marine organisms. *Molecular Ecology Resources*, *21*(4), 1347–1358. https://doi.org/10.1111/1755-0998.13322

Walkusz, W., & Rolbiecki, L. (2007). Epibionts (Paracineta) and parasites (Ellobiopsis) on copepods from Spitsbergen (Kongsfjorden area). *Oceanologia*, *49*(3), 369–380.

Weissman, P., Lonsdale, D. J., & Yen, J. (1993). The effect of peritrich ciliates on the production of Acartia hudsonica in Long Island Sound. *Limnology and Oceanography*, *38*(3), 613–622. https://doi.org/10.4319/lo.1993.38.3.0613

Wickham, H. (2019a). *Stringr: Simple, consistent wrappers for common string operations*. https://CRAN.R-project.org/package=stringr

Wickham, H. (2019b). *Tidyverse: Easily install and load the tidyverse*. https://CRAN.R-project.org/package=tidyverse

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., & Dunnington, D. (2020). *ggplot2: Create elegant data visualisations using the grammar of graphics*. https://CRAN.R-project.org/package=ggplot2

Wickham, H., François, R., Henry, L., & Müller, K. (2020). *Dplyr: A grammar of data manipulation*. https://CRAN.R-project.org/package=dplyr

Wickham, H., & Hester, J. (2020). *Readr: Read rectangular text data*. https://CRAN.R-project.org/package=readr

Wickstead, J. H. (1963). A new record of Ellobiopsis chattoni (Flagellata incertae sedis) and its incidence in a population of Undinula vulgaris var. Major (Crustacea Copepoda). *Parasitology*, *53*(1-2), 293–296. https://doi.org/10.1017/S0031182000072760

Windsor, D. A. (1998). Controversies in parasitology, most of the species on earth are parasites. *International Journal for Parasitology*, *28*(12), 1939–1941.

Wood, S. (2021). *Mgcv: Mixed GAM computation vehicle with automatic smoothness estimation*. https://CRAN.R-project.org/package=mgcv

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Chapman; Hall/CRC. http://yihui.org/knitr/

Xie, Y. (2020). *Bookdown: Authoring books and technical documents with r markdown*. https://github.com/rstudio/bookdown

Zamora-Terol, S., Novotny, A., & Winder, M. (2020). Reconstructing marine plankton food web interactions using DNA metabarcoding. *Molecular Ecology*, *29*(17), 3380–3395. https://doi.org/10.1111/mec.15555

Zervoudaki, S., Nielsen, T. G., & Carstensen, J. (2009). Seasonal succession and composition of the zooplankton community along an eutrophication and salinity gradient exemplified by Danish waters. *Journal of Plankton Research*, *31*(12), 1475–1492. https://doi.org/10.1093/plankt/fbp084

Øresland, V. (1991). Feeding of the carnivorous copepod Euchaeta antarctica in Antarctic waters. *Marine Ecology Progress Series*, *78*(1), 41–47.

Appendix A: Supplementary figures

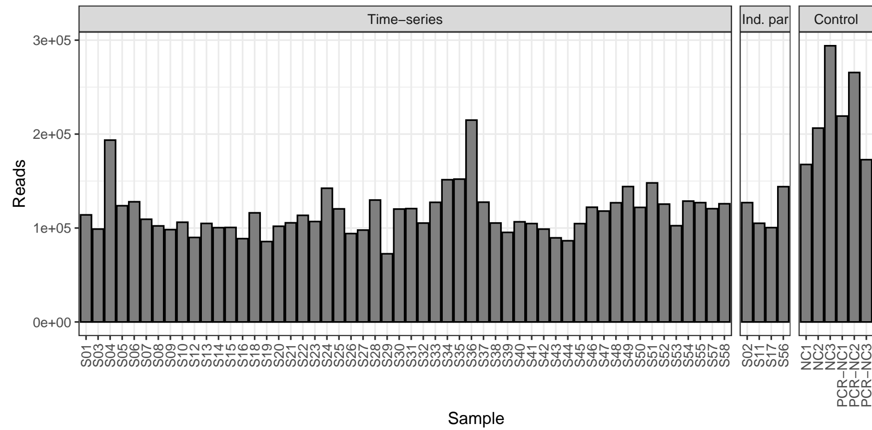


Figure S1: Raw read counts from illumina sequencing from time-series, individual parasite and control samples.

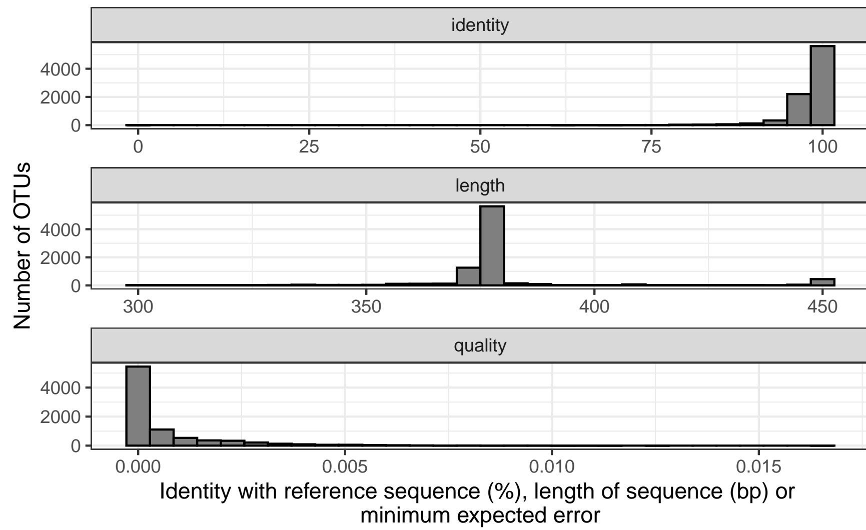


Figure S2: Quality histograms of the curated OTU table.

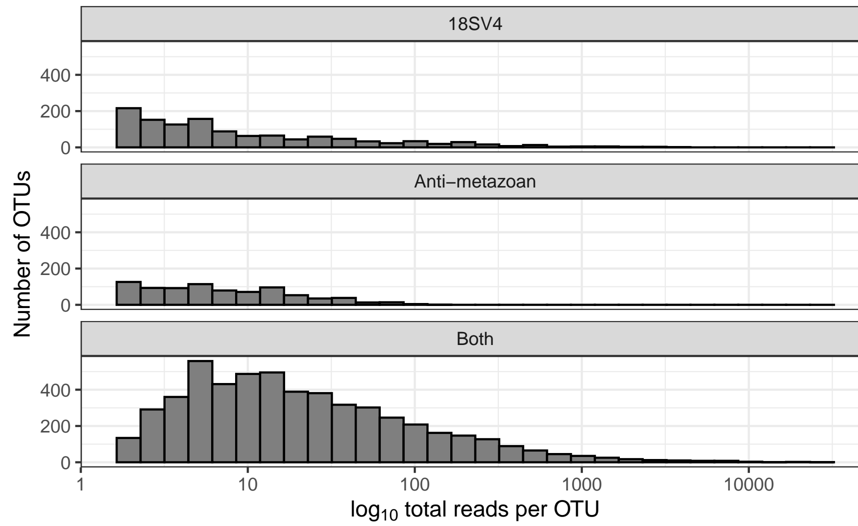


Figure S3: Distribution of total reads per OTU, grouped by whether they are exclusive to either primer set or common to both.

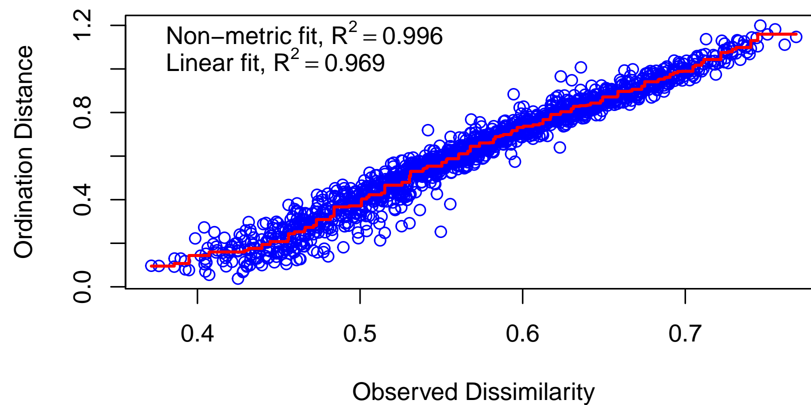


Figure S4: Shepard plot of the NMDS analysis, comparing the distances in the distance matrix to the distance in the ordination.

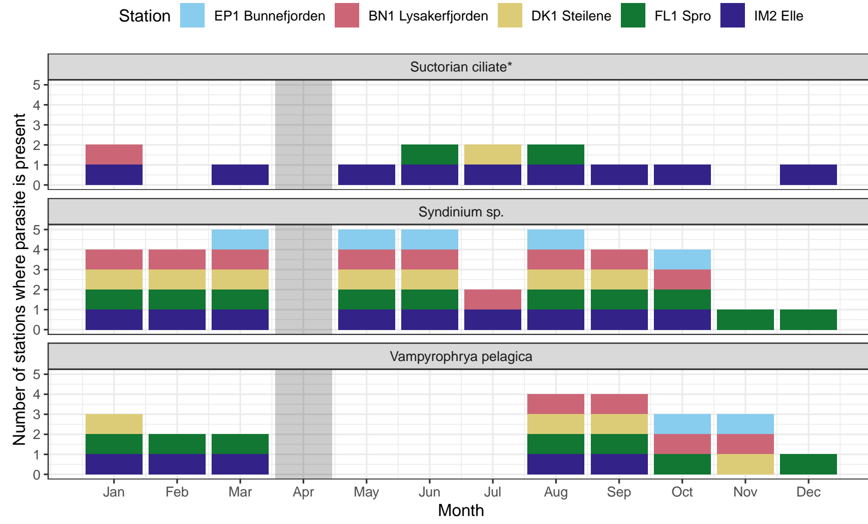


Figure S5: Occurrences per month of the taxa with overplotting issues in figure 28

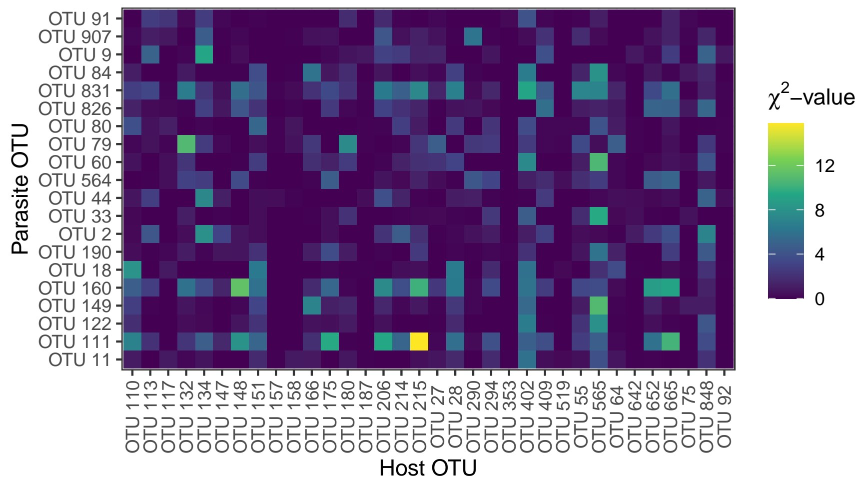


Figure S6: Chi-squared values from contingency tests of presence/absence of parasites vs. hosts.

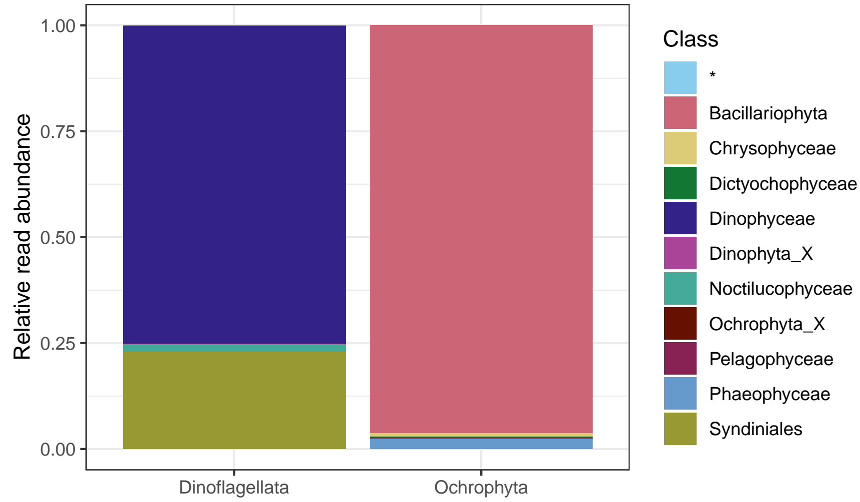


Figure S7: Relative abundances of classes in divisions Dinoflagellata and Ocrophyta. 75% of dinoflagellates were assigned to Dinophyceae, and 96% of Ochrophyta was assigned to Bacillariophyta (diatoms). The '' category are OTUs with unresolved taxonomy at class level.*

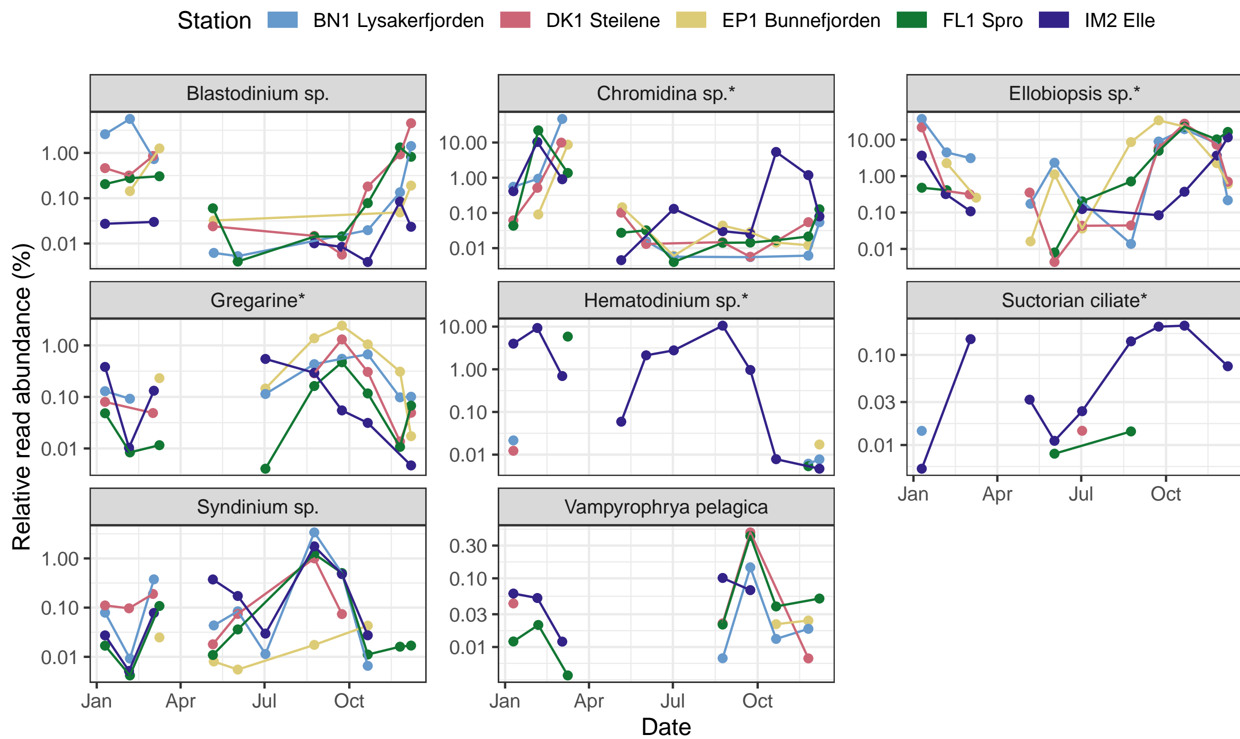


Figure S8: Relative read abundance of potential parasites with logarithmic y-axes.

Appendix B: Scripts

The full scripts along with an explanation of what they do can be found at https://github.com/evengar/master-thesis. Each script is also briefly described here.

File extension legend:

- R script: `.R`
- Python script: `.py`
- Shell script: `.sh`

Bioinformatic processing

mergepairs.sh

Reads all the forward and reverse fastq-files in a directory and merges each paired-end sequence.

primer_demultiplex.sh

Splits all files into two: one containing only sequences gained from the regular 18SV4 primers, and one containing those from the anti-metazoan primers.

local_derep.sh

Dereplicates each sample file, merging strictly identical sequences and annotating abundance.

filter_negative.sh

Reads the files of the negative controls, and removes any exact matches from the sample files.

derep_cluster.sh

Dereplicates all sequences globally, clusters with Swarm and checks for chimeras with VSEARCH's `--uchime_denovo` option.

cut_pr2.sh

Cuts the PR^2 database to match our fragments with Cutadapt. This is an example with the general 18SV4 primers, a similar script was made with the anti-metazoan primers.

filter_singletons.sh

Filters out OTUs with only a single representative sequence.

stampa.sh

Queries the sequences against the cut PR^2 database. Then it merges multiple best hits with the script stampa_merge.py and sorts by decreasing abundance. The script stampa_merge.py is written by Frédéric Mahé, available for download from: https://github.com/frederic-mahe/stampa/raw/master/stampa_merge.py

OTU_contingency.sh

Creates a contingency table of OTUs and samples. It calls the script OTU_contingency_table.py, which is written by Frédéric Mahé, available at https://github.com/frederic-mahe/swarm/wiki/Fred's-metabarcoding-pipeline#build-the-otu-table.

self_match.sh

Checks the pairwise similarities between all OTUs in the data.

lulu.R

Runs the LULU algorithm on an OTU table. Sums the two primers before curation, splits them afterwards.

Data analysis

CTD-formatting.R

Reads all CTD data from a folder and combines to one table. Requires that the file name structure is "YYYY-MM-DD.ID.cnv."

STD-formatting.R

Manually reads .csv files and combines to a single table. Example shown for September data.

ctd-std-join.R

Joins the formatted CTD and STD data.

envvir-summary.R

Summarizes environmental variables from the joined CTD and STD data.

OTUtab-functions.R

Contains some utility functions for analyzing metabarcoding data.