



UiT The Arctic University of Norway

Department of Arctic and Marine Biology

eDNA metabarcoding of amoebic gill disease (AGD) pathogens exposes potential vectors and reservoirs

Nathan Mertz

BIO-3950 Master's Thesis in Biology.....Marine Ecology and Resource Biology.....May 2020



Faculty of Biosciences, Fisheries and Economics,
Department of Arctic and Marine Biology

eDNA metabarcoding of amoebic gill disease (AGD) pathogens exposes potential vectors and reservoirs

Nathan Edward Mertz, UiT The Arctic University of Norway, Tromsø, Norway

BIO-3950 Master Thesis in Biology, Marine Ecology and Resource Biology, May 2020

Supervisors:

Kim Præbel, UiT The Arctic University of Norway, Tromsø, Norway

Owen Simon Wangenstein Fuentes, UiT The Arctic University of Norway, Tromsø, Norway

**HAVBRUKS-
STASJONEN**
I T R O M S Ø



(This page was intentionally left blank)

Abstract

Aquaculture worldwide is under constant scrutiny and financial pressure to maintain fish health, while providing an essential food source to a growing populace. To assist in monitoring mariculture sites and preventing harmful diseases, I tested a recently proposed method of trace pathogen detection using eDNA metabarcoding. The target pathogen, *Paramoeba perurans*, is the etiological agent of AGD, a respiratory infection which causes significant mortalities and treatment costs, affecting a growing percentage of the salmon aquaculture industry each year. 14 reference sequences of the COI Leray gene fragment from the family, genus, and species level were curated in order to identify DNA extracted and amplified from environmental samples collected at an active commercial scale aquaculture facility in Arctic Norway. The 12 *Paramoeba* sp. assigned MOTUs that were detected did not directly match any references, but showed considerable genetic relatedness to *P. pemaquidensis*, a known co-infecter, and displayed significant spatiotemporal trends within the sampled area and time series. Over a grid of 14 sampling points, distance from the farm, transect direction, and depth all effected the relative abundance of *Paramoeba* pathogens detected, and combined with known physical factors, provided evidence that incubation of these pathogens may have been occurring within the fish population despite no clinical signs being observed. Potential reservoirs for the pathogens in sediments and biofouling were evidenced by temporal changes in relative abundance coinciding with commercial activities at the start of production, and the existence of reads in the sediment prior to salmon placement. Continuous detection of *Paramoeba* throughout winter months indicate a lower temperature tolerance than was previously recorded for this genus, as low as 2.5°C. Phylogenetic assessment of detected MOTUs revealed 12 divergent haplotypes with varying degrees of relatedness to reference *Paramoeba* spp. and each other, exposing a novel and diverse assemblage of *Paramoeba* in this region. These discoveries highlight the capability of COI Leray metabarcoding to identify trace pathogens and assess spatiotemporal trends in their relative abundance and diversity, encouraging continued monitoring of Skogshamn and other aquaculture facilities going forward.

Table of Contents

Abstract	i
List of Figures and Tables.....	iv
Foreword.....	v
INTRODUCTION	1
<i>Origins of aquaculture</i>	1
<i>Modern advances in salmonid aquaculture</i>	1
<i>Major setbacks</i>	3
<i>Amoebic gill disease</i>	5
<i>Phylogenetics</i>	11
<i>eDNA metabarcoding</i>	14
<i>Objectives</i>	16
<i>Hypotheses</i>	16
METHODS	17
<i>Development of COI/18S reference sequences</i>	17
<i>Description of aquaculture site</i>	17
<i>Sampling design</i>	20
<i>Sample collection</i>	22
Seawater collection & filtration.....	22
Sediment collections	22
Other samples and data	23
<i>Laboratory practices</i>	23
Sediment and feed sample extractions.....	23
Water sample extractions.....	24
PCR amplification, library preparation, sequencing, and bioinformatics.....	24
<i>Statistical analysis</i>	25
<i>Phylogenetic analysis</i>	26

RESULTS	26
<i>Culture mitogenome assembly</i>	26
<i>COI barcode assignments</i>	27
<i>Distribution of reads between sediments and water</i>	28
<i>Distribution of sediment COI reads over space and time</i>	29
<i>Distribution of Paramoeba reads in water samples</i>	32
<i>Phylogenetic analysis</i>	35
DISCUSSION.....	37
<i>MOTU prevalence between sediment and water</i>	38
<i>Effects dictating spatial distribution of Paramoeba</i>	39
<i>Temporal clues to potential vectors and environmental tolerances</i>	41
<i>Diversity and genetic origins of Arctic Paramoebae</i>	43
CONCLUSION.....	44
PERSPECTIVES	45
Acknowledgements.....	47
References.....	48
Appendix A – Reference Sequences.....	57
Appendix B – Sampling Protocols.....	59
Appendix C – Sediment Extraction Protocol.....	62
Appendix D – Water Extraction Protocol.....	65
Appendix E – COI Leray-XT metabarcoding Protocol	68
Appendix F – Programming script for RStudio.....	71
Appendix G – Phylogenetic sequence alignment	79

List of Figures and Tables

Figure 1 Farmed Atlantic salmon production weight by country(1990-2018e)	2
Figure 2 Scanning electron micrographs of gill filaments from Atlantic salmon.....	7
Figure 3 Location of the Skogshamn aquaculture research station and sampling points	18
Figure 4 Solbergfjord bathymetry and commercial scale salmon aquaculture farm map	19
Figure 5 Circular barplot of Skogshamn current velocities and directions by depth.....	20
Figure 6 Relative read abundance of <i>Paramoeba</i> sp. by sample type	29
Figure 7 Relative read abundance of <i>Paramoeba</i> sp. in sediments by distance	30
Figure 8 Relative read abundance of <i>Paramoeba</i> sp. in sediments by directional transect.....	31
Figure 9 Relative read abundance of <i>Paramoeba</i> sp. in sediments by depth	31
Figure 10 Relative read abundance of <i>Paramoeba</i> sp. in sediments by time	31
Figure 11 Relative read abundance of <i>Paramoeba</i> sp. in surface water by distance from aquaculture pens over all sampling dates.....	32
Figure 12 Relative read abundance of <i>Paramoeba</i> sp. in surface water on Sept 12th by distance from aquaculture pens.....	33
Figure 13 Relative read abundance of <i>Paramoeba</i> sp. and sea surface temperature by date ..	34
Figure 14 Relative read abundance of <i>Paramoeba</i> sp. in surface water by directional.....	35
Figure 15 Dendrogram of detected <i>Paramoeba</i> COI Leray-XT MOTUs and reference sequences of related Paramoebidae species	36
Table 1 Skogshamn sampling schedule.	21
Table 2. Skogshamn COI Leray-XT MOTUs.....	28
Table 3 <i>Paramoeba</i> reference sequences.....	57

Foreword

This thesis is written and submitted as the completion of the second year of the Master of Biology program from the department of Arctic and Marine Biology at UiT The Arctic University of Norway. Topics discussed herein are the result of research conducted on behalf of the Research Group for Genetics (RGG) at the Norwegian College of Fishery Science from May 2019 to May 2020 with funding from Havbrukstasjon i Tromsø. My involvement included: contributing to the project's experimental design, acquiring and assembling sampling equipment, conducting the majority of fieldwork and eDNA extractions from sediments, feed, and water samples, accumulating and curating reference sequences, running statistical and phylogenetic analyses, producing tables and figures, and writing of the manuscript. Other significant contributions to the project were made by many members of RGG in planning, fieldwork, and laboratory aspects, most notably:

Kim Præbel, supervisor, conceived the idea, procured funding, assisted in experimental design of the thesis, conducted fieldwork, designed eDNA laboratory extraction methods, obtained cultures and their sequences, and oversaw all other aspects of the project.

Owen Wangenstein, co-supervisor, assisted in experimental design, fieldwork, library prep, PCR, sequencing, and ran metabarcoding pipelines when UiT was forced to close to students during the COVID-19 pandemic.

Gledis Guri, a fellow master student, helped in the design of overall sampling strategy and contributed equally to field and lab work.

Marta Turon, post-doc, assisted significantly with field and lab work, managed the sample datasets, corresponded with farm technicians, and gave guidance in data analysis methods.

Julie Bitz-Thorsen, lab manager, assisted with fieldwork, performed DNA extractions on remaining sediments and water filter samples, conducted PCRs, library preps, and sequencing when UiT was forced to close to students during the COVID-19 pandemic.

Shripathi Baht, bioinformatics specialist, performed shotgun sequence genome assemblies

(This page was intentionally left blank)

INTRODUCTION

Origins of aquaculture

Aquaculture - the act of growing fish in captivity, be it for personal consumption, sale, decoration, or otherwise - has existed for thousands of years. In Norway, a country with extensive marine and freshwater fish resources and a low population density, aquaculture remained a relative unknown until around 1850. At this time, the first hatchings of cultured brown trout (*Salmo trutta* L. 1758) were taking place, followed a few decades later by the import of rainbow trout (*Onchorhynchus mykiss* Walbaum 1792) from Denmark, and the growth of freshwater culture in dams and lakes¹. Experiments feeding and growing wild Atlantic cod (*Gadus morhua* L. 1758) larvae in captivity took place at the Institute of Marine Research early in the 1930s, but the big boom came at the end of WWII when entrepreneurial interest in aquaculture grew significantly. Then, the first successful translocations of trout species and Atlantic salmon (*Salmo salar* L. 1758) from freshwater to saltwater were carried out, enabling faster growth². By 1969, the first sea cages were constructed and successfully tested, birthing an industry that would change the world of seafood forever^{1,3}.

Modern advances in salmonid aquaculture

From its early origins in Norway, mariculture of salmonids developed from a cooperative accumulation of experienced based knowledge from within the small fishing communities where it was taking place⁴. The important lessons learned about the biology of the species being grown and the ways to decrease mortalities during each life stage were assembled slowly through the 1960s and 70s, but successes were shared, and progress made⁵. Certain problems eventually required scientific study, with survivorship of early life stage hatchery larvae and the timing and chemical basis for smoltification being major priorities. Breakthroughs in the understanding of osmoregulatory changes and the use of blood chemistry to monitor smoltification progress in parr were vital to the start of a more industrialized hatchery process⁶. A shift was also made around this time in breeding priorities. While previous breeding was done locally based on wild characteristics, Harald Skjervold, a professor from the Norwegian School of Agriculture and expert in breeding and genetics, started a salmon breeding program with the principles he had established based on terrestrial livestock. Using wild fish from 41 rivers throughout Norway, the station he founded went on to selectively breed salmon for specific traits, including improved larval survival and disease resistance^{7,8}. These advances brought about a large boom in hatchery production capabilities in the early 1980s⁶. As a result

of the rapidly increasing fish volume in sea pens across Norway, the demand for fishmeal spiked. The rising costs of feeding thousands of salmon and the level of waste observed from wet feed use spurred research on dry feed formulation, including characteristics such as fat vs protein content, density, and the effects of using cheaper plant based ingredients⁹. The combined effort in each of these fields, as well as the cooperation between fish farmers, breeders, and researchers, drove Norway to become a global leader in all areas of anadromous fish culture early on.

This leadership has meant that Norwegian interests have played a large part in the expansion of Atlantic salmon aquaculture as it has spread on a global scale. The species, native to the higher latitudes of the North Atlantic, quickly became farmed within its range along the eastern Canadian seaboard, and eventually Iceland, Scotland, Ireland, and the Faroe Islands became significant producers of the native species as well. On the other side of the world where Atlantic salmon are non-native, operations in Chile, Australia, British Columbia, and Washington State also began farming Norwegian-bred salmon on an industrial scale and looked to their Norwegian suppliers for guidelines and best practices. Production in these countries followed a similar surge in volume after initial establishment (Figure 1). This rapid growth in the mariculture industry has had its benefits and drawbacks for Norwegian interests as well as international, from increased success with production regimes, sex manipulation, and sterilization, to further difficulties associated with feed costs, market flooding, and disease outbreak^{5,10-12}.

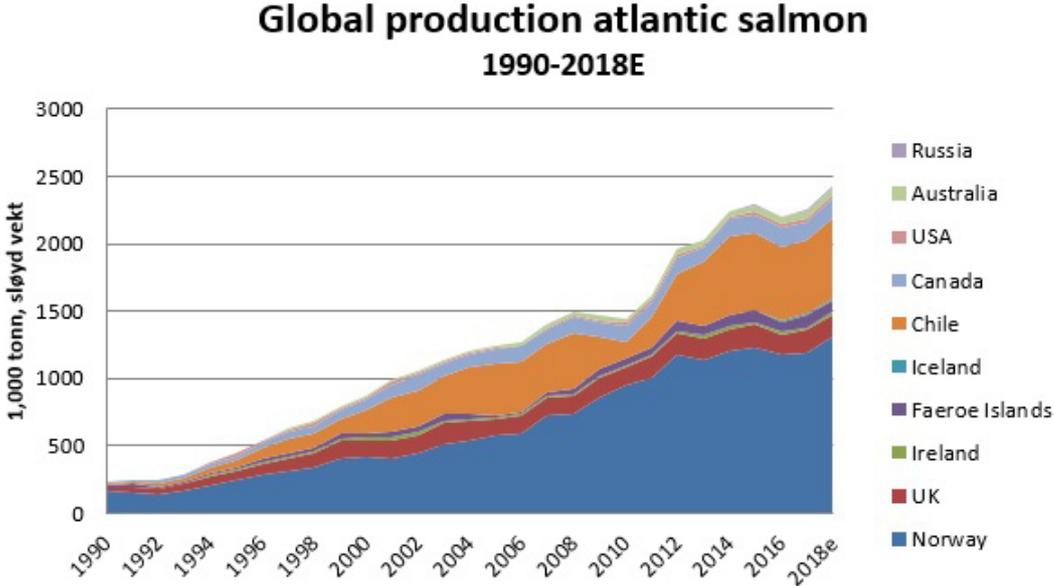


Figure 1 Farmed Atlantic salmon production weight by country(1990-2018e)¹³

Major setbacks

Over the last half century, several major setbacks to salmonid aquaculture have occurred, primarily due to rapid disease and parasite spread, but also concerning social issues such as environmental impact and fish welfare. Disease outbreaks specifically have caused serious problems for the industry from the start of intensive hatchery practices back in the 1950s through to today¹⁴. The first bacterial disease encountered by a budding aquaculture practice in Norway was furunculosis, a disease caused by bacteria commonly found in freshwater systems worldwide, especially those with low water quality or high densities of fish, such as rivers used in the early years of hatchery production¹⁵. Vibriosis diseases have also caused serious distress within the aquaculture industry over many years due to their rampant and deadly nature. Typical vibriosis, caused by *Vibrio anguillarum* Bergeman 1909, is the most common and was the first to be recognized as pathogenic in fish. The disease causes rapid mortality in young juvenile salmon during warmer temperatures due to internal hemorrhaging, and can be very resistant to antibiotics¹⁵. The latter trait of drug resistance was resolved by the creation of a vaccine, but other closely related pathogens *Aliivibrio salmonicida* and *Moritella viscosa* have plagued aquaculture in Norway and around the world in the decades since.

Beyond the damage to aquaculture industry's profitability caused by expensive disease treatment and mortality of fish, the sector's public image took a significant blow as a result of the many bacterial outbreaks and resulting treatment measures that took place in the late 20th century. In order to decrease the frequency and size of infectious disease related die offs and save fish who showed signs of illness, broad spectrum antibiotics became widely used. Concerns about the environmental impacts of antibiotic use in open sea pens, as well as the risk of disease spread to declining wild salmonid populations, have negatively affected the marketability of farmed salmon in Europe and North America for the last 3 decades¹⁶. Despite major shifts in the veterinary approach to aquaculture fish health towards preventative care through vaccines and reduced handling stress, plus a huge reduction in the overall mortality rate of fish in all stages of culture, the public image of the salmonid aquaculture industry is still struggling to recover^{17,18}.

Salmon louse has been the most harmful pathogen to salmon aquaculture in recent years¹⁹. *Lepeophtheirus salmonis* (Krøyer 1837), the species of parasitic copepod native to Norwegian waters, has been rampant in its seasonal spread through farms throughout Norway since the initiation of extensive aquaculture systems. In addition to being harmful to the farmed fish by

direct parasitism of feeding on the mucous membrane and dermal tissues, sea louse and other parasitic copepods are known to act as transmission vectors for bacterial and viral diseases^{20,21}. Therefore, the treatment of *L. salmonis* outbreaks and development of methods to prevent the parasite's spread have been a priority in Norway for decades. Simple, but costly methods such as freshwater and chemical bathing treatments, dominated for many years, and are still common-use, but implementation of cleanerfish cohabitation and advanced laser technology methods to remove the parasites as they grow on the fish are also now widely accepted preventative measures²². Despite the massive effort by industry and research institutions worldwide, the rapid life cycle of louse, with to six generations per year, and close proximity of farms within salmon producing regions ensures that these parasites will continue to prey on cultured fish²³.

Complex gill diseases (CGDs), caused by a broad group of harmful organisms including amoebas, bacteria, and algae, are another serious health problem for aquaculture worldwide because no prophylactic treatment methods are currently available²⁴. The diseases, which can often cause high mortalities, can only be mitigated by intensive bathing methods to rinse pathogens away, like those used for salmon lice. Most often, these diseases are the result of an infection of the gill structures, which causes an immune response that limits the capability of the gills to uptake oxygen. The various gill diseases usually cause lethargy and reduced appetite in infected fish, eventually leading to susceptibility to other diseases and mortal conditions. Infections frequently involve multiple species, such as the bacterium *Tenacibaculum maritimum* (Wakabayashi *et al.* 1986), *Branchiomonas cisticola* Toenshoff *et al.* 2012, and *Piscichlamydia salmonis* Draghi *et al.* 2004, which may coinfect immediately, or in succession, making study of the specific causal agents extremely difficult²⁵. In some cases, the disease can be initially caused by non-parasitic organisms, such as algae or even just large volumes of suspended particulates in the water column that can become embedded in the gills. Microalgae which have caused significant CGDs recently, such as *Pseudochattonella verruculosa* Hara & Chihara 1994 and *Chrysochromulina leadbeateri* Estep *et al.* 1984 will often do the most harm to the fish during spring blooms, when large swaths of phytoplankton can drift through farms, releasing harmful toxins that damage the fishes' gill structures as they photosynthesize and reproduce^{26,27}. Amoebic gill disease (AGD), on the other hand, is caused by aggressive unicellular eukaryotic parasites, which target the gills as an easily accessible location to embed themselves, feed, and reproduce when optimal conditions occur. AGD, the most significant gill

disease in regard to both fish health and industry economics, is the primary focus of this study, and is described in more detail below.

Amoebic gill disease

The onset of AGD in Atlantic salmon was first observed by mariculturists in Tasmania, Australia in the 1980s, shortly after extensive aquaculture practices began in the region in 1984. The gill disease was diagnosed in young salmon in sea cages during warmer water temperatures in the fall months of 1985, resulting in mortalities of 2% per day and 50% overall during outbreaks in subsequent years²⁸. These events spurred research efforts to quickly identify the pathogen, which was declared *Paramoeba* sp., a genus of amoeba containing an endosymbiont, after histological examination of infected gill tissues. Further study resulted in the culprit being labeled *Paramoeba pemaquidensis* Page 1970, based on the size and structure of the organism and its pseudopodia, or arm-like projections of the cell membrane characteristic to amoeba²⁹. Simultaneous to these events in Australia, outbreaks of gill disease were ravaging coho salmon, *Oncorhynchus kisutch* (Walbaum 1792), aquaculture operations in British Columbia and Washington State along the west coast of North America. Identification based on histological mounts of infected gill structures, as well as cultured amoeba came to the same result³⁰. Measured specimens from both regions were noted for changes in appearance between liquid vs agar cultures and their actively infectious counterparts and were labeled locomotive and transitional forms. Despite establishing many cultures of these morphologically different forms, both in vitro and in vivo infection experiments were never successful, meaning that Koch's postulates were never confirmed for the potential pathogen^{28,30}. These difficulties reestablishing infection severely limited the ability of researchers to study the mechanism of transmission for AGD and prevent its spread both locally and globally. Approximately 20 years after the first outbreaks, due to modern advances in molecular tools for species identification, the true etiological agent of AGD, *Paramoeba perurans* (Young *et al.* 2007) was discovered. By isolating amoeba samples directly from infected gill tissues and using novel next generation sequencing methods, researchers at the University of Tasmania extracted, amplified, and sequenced rDNA from the 18s and 28s genes from all the species present during an active infection. After assembling sequencing reads from the two genes, it was discovered that a yet unnamed species of the genus existed and possibly caused the infections, but had not responded to all previous methods of isolation and culturing³¹. Upon this discovery, specific culture methods were developed for *P. perurans*, and in 2011, confirmed cultures of the amoeba were successful in infecting salmon in a controlled environment³².

Despite the many difficulties culturing the true pathogen, research was conducted from the earliest days of AGD detection to study the progression of symptoms in infected fish. In order to better prevent and treat the newly discovered disease, histological examinations of infected gill tissues were conducted from each stage of AGD progression. Study of symptomatic fish resulted in a better understanding of the biological effects of the disease on fish under intensive culture conditions, and systematic scoring methods for gill health^{33,34}. Trial and error of treatment methods by researchers and industry workers led to the development of a patchwork of treatment options, such as short-term freshwater bathing, which is still used extensively today³⁵. Continued monitoring of environmental conditions leading up to outbreaks also contributed to increased preparedness for AGD symptom detection and treatment implementation when necessary³⁵⁻³⁷. Even with the many breakthroughs in understanding AGD treatment and the environmental preferences of its pathogens, little could be done by industry giants to prevent its spread, and throughout the 1990s and early 2000s, *P. perurans* expanded its infectious range to aquaculture operations worldwide.

Detailed analysis of the gross pathological and histological patterns of AGD has left us with a basic understanding of its biological effects on infected fish. After the initial arrival of amoebas to the gill structures, likely due to salmonid's active swimming state during respiration and increased densities of the free-floating pathogen in the water column, pathogens become attached to the gill filaments via the epithelium of the lamellae. The long pseudopodia of actively infectious *P. perurans* embed themselves into the epithelium, while the remainder of the pathogen cell conforms itself to the shape of the epithelial membrane³⁸. The immune response from infected fish results in the excretion of additional mucous, with 94 of ~500 gill and skin mucous proteins being expressed differentially to combat the pathogens³⁹. Additional immune responses include epithelial oedema, or the buildup of excessive fluids within the epithelium, and hyperplasia, the rapid enlargement of gill tissues due to cell reproduction²⁵. Both of these effects are the basis for rapid inflammation of the gill tissues, which decreases water-gill interface for respiration, and leads to further decline of the infected fish's health. Visible symptoms by this time include visible discoloration of the gills due to patches of excessive mucous buildup, the formation of lesions, tissue necrosis, and sloughing (Figure 2). Histological examination of these symptomatic areas will show signs of lamellae fusing together wherever amoebas are attached and the formation of large vesicles within their epithelial cells²⁴. Long term longitudinal studies of gill diseases have shown that gross pathology is effective in diagnosing the outbreak of a gill disease, but that scoring systems

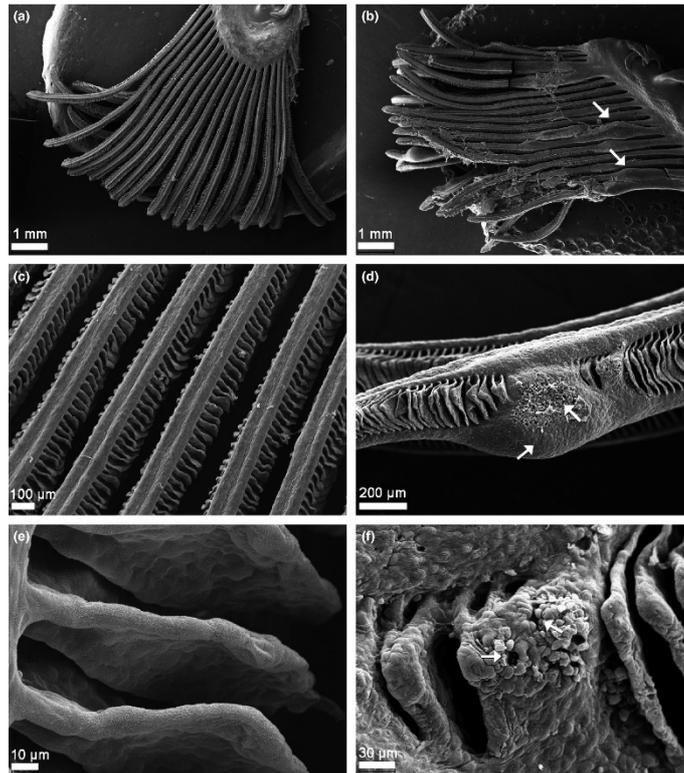


Figure 2 Scanning electron micrographs of gill filaments from healthy Atlantic salmon (images on the left A, C and E) and gill filaments infected by *Paramoeba perurans* on the right (b, d and f).

based on histological examination of infected gill tissues are the most effective way of confirming AGD and monitoring disease progression in areas dealing with systematic outbreaks^{33,34} Therefore, gross pathology has been widely used to indicate the need for treatment in an industrial setting while histology has been essential in confirming the timing and severity of outbreaks where environmental conditions are the variable of interest.

An immediate priority after the discovery of AGD was to develop an understanding of its environmental triggers and the range of conditions at which its pathogens can survive and cause infection. Two methods were used primarily, study of the disease's natural prevalence and intensity through field studies in areas of Tasmania with recurrent outbreaks, and laboratory research experiments within a controlled environment. Study of *Paramoeba* sp. growth rates in culture were undertaken at a wide range of temperatures and salinities. Early cultures of *P. pemaquidensis* were noted for being quite salinity tolerant, being capable of rapid growth in seawater at as low as 15ppt, and survival to 5ppt. Temperature has a stronger effect on the growth rate of the species, with both variables having a positive relationship from 5°C to 20°C³⁰. *P. perurans* cultures responded similarly in their more recent study, reaching its maximum reproduction rate at 15°C and 35ppt, while maintaining positive growth as low as 8°C and 25ppt. Colony survival remained stable for 15 days down to 2°C, but the salinity

threshold for the species appears to be less tolerant than others of the genus, with minimums falling between 20 and 25ppt. Field studies of the disease were also quickly implemented throughout southern Tasmania and the Huon estuary in the 1980s and 90s, where aquaculture facilities were regularly sampled for gill tissue samples and environmental variables such as temperature, salinity, and dissolved oxygen. Results from numerous studies concluded a general range of 13-20°C for possible outbreaks, with very high risk at any temperature >16°C and salinity >30ppt^{28,35,40}. Observations from farms worldwide have challenged these findings as the disease has spread geographically, with incidences now occurring at as low as 6°C in some regions, and infections being sustained in fish undergoing freshwater bathing treatments or relocation to brackish systems⁴¹⁻⁴⁵. It is unknown whether these trends of apparent increased temperature and salinity tolerance are the result of evolution within different strains of the disease, or simply greater resolution in monitoring the disease within a rapidly advancing aquaculture industry. In either case, the lower temperature threshold for disease outbreak has meant spread of the disease throughout the entire geographic area of global salmonid mariculture.

The origin of *Paramoeba perurans* and AGD are not truly known, as the disease was detected in both the NE Pacific and Tasmania within a very short time frame. What is known however, is that the rapid range expansion since its initial detection threatens a rapidly growing percentage of the salmon aquaculture industry⁴⁵. Over the three decades after its first diagnosis, incidences of AGD outbreak have been observed in numerous other cultured marine fish species in the Mediterranean, including *Dicentrarchus labrax* (L), *Diplodus puntazzo* (Cetti), and *Scophthalmus maximus* (L), Salmonids throughout the European Atlantic, Japan, Chile, and South Africa, as well as three cleanerfish species used in European salmonid aquaculture for mitigation of *L. salmonis* in seapens^{45,46}. Other *Paramoeba* species are also documented as opportunistic parasites in marine invertebrates and fishes, most notably, *P. invadens* for its decimating effect on urchin populations in Nova Scotia, Canada in 2011⁴⁷, and *P. pemaquidensis* for its role in AGD coinfection and mass mortalities of American lobster in 1999⁴⁸. Whether or not these other *Paramoeba* spp. outbreaks are the result of similar global range expansion, the mechanism of amoeba derived disease spread is not understood. In the case of each new regional AGD outbreak though, abnormally high sea surface temperatures are always recorded, followed by rapid outbreak and recurrent disease presence in subsequent years^{43,44}. The first recorded outbreak in the Atlantic Ocean occurred at salmon farms along the coast of Ireland in 1995 after weeks of heat wave, record high water temperatures, and a

reduction in rainfall, which led to increased salinity of protected coastal waters where net pens are typically moored⁴³. Since then, salmon aquaculture companies in Ireland, France, and Spain have sustained significant financial losses due to fish mortalities and treatment associated with AGD. 10 years later, during Norway's warmest measured ocean temperatures on record, and prolonged temperatures at 3+°C higher than average, four aquaculture operations spanning a 350km area of western Norway all sustained major AGD outbreaks, with some of the highest mortality rates ever caused by the disease, upwards of 80%^{44,49}. Subsequent years have seen a rise and fall in the threat of *Paramoeba* infection, with warm years such as 2011 and 2012 continuing the geographic spread of the disease northward past the Arctic Circle, as well as in cooler years when only a limited number of AGD cases have been detected nationwide and swiftly resolved through locally implemented treatment measures⁴⁵.

The impact of AGD outbreak on intensive salmon mariculture operations can be devastating without proper implementation of treatment measures. The increased prevalence of *Paramoeba* spp. around the globe, as well as compounding factors such as more frequent harmful algal blooms and coinfection by marine bacteria make rapid detection and treatment of AGD a necessity. Our limited capability of combatting even human parasitic amoebas has prevented the development of any sort of effective vaccine or antimicrobial agent to target *Paramoeba* species, but their intolerance to very low salinities has provided a simple solution to AGD treatment in the form of freshwater submersion baths. First administered to treat gill infection in Tasmania and still widely used today for both AGD and *L. salmonis* outbreaks worldwide, short (2-3hr) periods of freshwater immersion are highly effective at killing amoebas embedded in the fishes' gills and washing excess mucous away from the gill filaments²⁸. The costs associated with this treatment were very high for most facilities that were required to combat AGD, requiring either the transportation of fish to a well boat or shoreside facility for bathing. More recently, large tarps have been created to hold a shallow layer of freshwater on the surface of the net pen, so that groups of salmon may be rotated through it for bathing. Many remote operations, especially in regions new to the outbreak were not able to conduct these sorts of freshwater baths, and were instead forced to attempt treatment within the cages with formalin and hydrogen peroxide⁴³. In many cases, these measures were effective, so the use of hydrogen peroxide has become more common in regions such as Ireland, Scotland, and Chile where large volumes of freshwater may not be easily obtained from the nearest landmass. The implementation of this treatment, unlike freshwater baths, can have hazardous side effects on the fish at temperatures higher than 13°C, but is also effective on sea louse infestations, and is

therefore widely used⁴⁹. Other treatments, including dietary administered immune boosters and vaccines have been tested, but show little to no effect²⁴. The one long term treatment strategy that shows promise is breeding individuals for disease resistance, but the architecture of resistance traits appear to be polygenic, so significant research and administration of the traits into current breeding programs is still necessary⁵⁰.

Overall, the costs associated with treatment of AGD in areas with frequent outbreak are immense, and their toll on both salmon welfare and the surrounding environment must also be considered. The costs in regions such as Tasmania and Scotland can be as high as 20% of the overall production and additional costs of ~41million USD have been estimated for the aquaculture industry in Australia due to AGD treatments. Without these treatment measures, mortalities costing 12 million and 81 million USD in losses in Norway and Scotland have been estimated during outbreak years⁵¹. Additionally, the environmental toll of frequent chemical and freshwater treatments should not be discounted. For example, most delousing treatments performed in an open net pen result in the death of cleanerfish who coinhabit the enclosed area¹⁹. Particularly for chemically enhanced treatment baths, rapid dilution is essential for limiting environmental impact as the chemicals disperse into the surrounding water, but the negative effects on biotia in the immediate area can still be felt. Therefore, prevention of the disease by detection and early response is the utmost priority.

Numerous species of *Paramoeba* have now been detected in the marine environment independent of fish or invertebrate hosts, so it is apparent that they have an alternative life strategy to parasitism. Detection methods for environmental samples began with culturing sediment sample extracts but have evolved to the use of rapid molecular methods, such as real-time polymerase chain reactions (PCR) with species specific primers for samples from water, nets, and sediments. Before the discovery of *P. perurans*, *P. pemaquidensis* was isolated from marine and estuarine sediments collected around Tasmania, and detected in locations both associated and not associated with fish farming and previous AGD outbreak³⁷. Molecular analysis of these cultures by the Institute of Parasitology in Czech, as well as some isolated from nets, water samples, and direct gill swabs, led to the discovery of *P. branchiphila* (Dykova et al. 2005), a new species, and *P. aestuarina* (Page 1970) presence in gill and environment samples alongside the previously known agent *P. pemaquidensis*⁵². Following the characterization of *P. perurans*, a species-specific primer for the amoeba was developed and validated by PCR assay of environmental samples taken at AGD infected farms in Tasmania⁵³. Bridle et al. 2010 were able to detect the pathogen at high abundance in water adjacent to the

farm, and semi-quantitatively assess *P. perurans* loads from gill swab samples with their method, but on a wider scale the detection of the specific pathogen has been unsuccessful when AGD is not present. In Norway, *P. perurans* has been detected in wild fish, filter feeders, biofouling, sediment, and water samples taken from multiple depths during AGD outbreak, but a specific reservoir for the amoeba outside of these events is not clear^{54,55}. Detection of the pathogen in the gills of both wild fish and farmed cleanerfish present a potential vector for reinfection following freshwater bath treatment of farmed salmonids, but the origins of the pathogen for each new outbreak are not understood^{41,55}

As a result of the many studies sequencing infected gill tissues, all the species of *Paramoeba* discussed above, as well as a few other amoebozoans and bacteria have also become associated with AGD. The disease can therefore be labeled cosmopolitan, and an understanding of each associated pathogen should be of interest to prophylactic treatment schemes, in addition to methods for early detection. Previous phylogenetic analysis of cultured, environmental, and tissue derived *Paramoeba* samples has led to a better understanding of the evolutionary pathways that have led to the diversity of species and strains currently known, but research in the field must continue as new samples are obtained and new methods developed.

Phylogenetics

Phylogenetics can be defined as the study and identification of evolutionary patterns and structure in nature. The science originated from the early morphologically based taxonomic assignments of Linnaeus in 1758 and has undergone a continuous development in the centuries since. As with evolution in nature, phylogenetics has been marked by a few explosive discoveries which have led to further radiation of the field. These include, of course, the rise of the theories of phenetics and cladistics, as well as the technical advances in computer algorithms and molecular sequencing^{56,57}. The discovery of DNA and advent of sequencing capability increased the capability for biologists to see evolutionary history at a significantly finer resolution, and gave taxonomists evidence with which to delineate species, subspecies, populations, and beyond⁵⁸. This is possible because molecular data derived from DNA sequencing of different individuals of a given group can be analyzed by specific algorithms to determine their percent relatedness. Specifically, for the study of microbial organisms the implications of these discoveries were vast, as they provided a novel method for distinguishing between large swaths of organisms with complex life histories and immense overlap in visible traits that were previously the only means for characterization⁵⁸. The boom in new species

discoveries and creation of public genome catalogs that followed have opened new doors for a broad range of scientific pursuits. Epidemiology, the study of incidence, distribution, and risk factors of diseases, is one such field. Molecular epidemiologists have benefited substantially from advanced methods insofar as their ability to locate and track evolving strains of harmful pathogens⁵⁹. Use of phylogenetic analysis can help to rapidly identify newly emerging haplotypes that may pose increased resistance or virulence, as well as trace the origins of pathogens earliest forms to understand their previous evolutionary pathways.

The root method for these analyses is the building of phylogenetic trees, which are graphical representations of taxonomic groups whose branching patterns indicate evolutionary relationships⁵⁶. These diagrams are composed of nodes and branches, with nodes representing taxonomic units and branches the evolutionary distance between them. Terminal nodes represent the operational taxonomic units (OTUs), which are the known entities being compared, whereas the internal nodes represent hypothetical individuals who were the common ancestor of the two nodes which branch from them. In trees comparing relationships based on genetic sequences or other molecular data, these are referred to as MOTUs. Branching patterns of phylogenetic trees can be represented in various forms, but they will always emerge in pairs, and their scaled or numerically listed length represents the amount of divergence between the ancestor and descendant⁶⁰. Dendrograms in particular will display clusters of nodes, which represent closely related OTUs. In epidemiological taxonomy, these clusters will often be made up of strains of a particular pathogenic species, who are distinguished by their number of genetic differences. Four primary algorithms have been designed to calculate the difference between individual and clustered OTUs based on their genetic code, but the methods most widely used in computational phylogenetics today are character-based, and are referred to as maximum likelihood and maximum parsimony^{60,61}.

The maximum parsimony method in effect attempts to create a phylogeny that requires the least amount of evolutionary change. To do this, each potential tree is given a "cost", which signifies the sum of all character change lengths over all nodes. The tree(s) with the lowest calculated cost is then presented as the most plausible result. Maximum likelihood, on the other hand, is a statistical approach which calculates the probability that a chosen model predicts the observed outcome. This method requires considerably more computational power, and an appropriate evolutionary model for the taxonomic groups of interest but should produce a phylogenetic tree with the highest probability of occurrence⁶⁰. Each method has its drawbacks; maximum parsimony is susceptible to errors when assessing rapidly evolving lineages, while maximum

likelihood is less reliable when dealing with smaller samples. Maximum likelihood analysis can be significantly enhanced through the use of an outgroup, or more distantly related sequence, which acts as a negative control, and also through the use of bootstrapping. This involves a reanalysis of the data over a set number of iterations, usually 100-1000, and the formation of a "majority rules" final tree which uses the most prevalent results from each individual analysis⁶¹. The combination of many of these techniques creates a phylogenetic analysis with greater accuracy, which in turn allows more in-depth study of the relationships between rapidly evolving pathogens and their hosts.

Specific to AGD and *Paramoeba*, advances in DNA sequencing and phylogenetic assignment algorithms have made possible the characterization of pathogenic species who would otherwise be indistinguishable from their non-etiological relatives. The types of DNA and gene regions used for these identifications have broadened as the collective knowledge related to gene stability and mutation rates has grown. Ribosomal RNA(rRNA)s have been by far the most used for taxonomic assignment of species since Woese and Fox used them to theorize the concept of 3 domains of life⁶². Due to their nonadaptive structure, which is nested in the core of essential cellular functions, rRNAs are highly conserved across species⁶³. Even so, rRNA genes still develop variation in noncoding regions over time, so genes such as 18S have become a reliable method for taxonomic assignment of species⁵⁸. The 18S gene was the primary tool for biologists identifying new species of *Paramoeba* associated with AGD and other marine diseases, as well as establishing relatedness between and in some cases within them^{31,44,47,48,52,64-66}. More recently, the mitochondrial cytochrome oxidase I(COI) gene has become a target for sequencing efforts for the purpose of species identification. The higher rate of variation within the COI gene compared to 18S rRNA provides for greater phylogenetic resolution when studying organisms on the population level⁶⁷. This added resolution when conducting analysis means that divergence between geographically isolated strains of each *Paramoeba* spp. potentially can be more readily detected, and trends in their evolution based on differences in selection pressures can be inferred⁶⁸⁻⁷⁰. Due to the further development of sequencing methods, today, very short fragments of the COI genes, as well as others, can be used to identify thousands of species present in a sample collected from almost any medium⁷¹. In addition, the DNA extracted from these samples, often water or soil, can even be used to discover unidentified new species or strains and make inferences about their geographic distribution⁷².

eDNA metabarcoding

DNA extracted from environmental samples such as water, soil, or air can come from many sources, including whole microbial organisms, cells shed from large eukaryotes, and extracellular DNA present as a result of excretion or cell death. These forms of DNA are referred to as eDNA. Each of these sources have different biological, chemical, and physical forces acting upon them, and can therefore have different rates of production and degradation⁷³. However, they provide an invaluable resource for identifying organisms living within a system and are especially useful for monitoring rare and/or microscopic species assemblages. Microbial organisms which are free living in these environments can be difficult to identify and quantify in small fraction microscopy, and equally troublesome to culture for more in depth morphologically based assessments. Furthermore, taxa with less frequency of occurrence, such as seasonally driven phytoplankton or pathogen populations can go undetected using manual species composition methods while still being present at trace abundances in the water column or sediment. Use of molecular identification techniques however, can characterize microbe presence and relative abundance even at the level of a single cell in a 1L water sample⁵³. Techniques using these rapidly developing methods, namely eDNA metabarcoding, therefore have immense potential in detecting the presence of trace abundances of pathogens, and monitoring their spread and proliferation⁷⁴.

The detection of these organisms first requires a genetic region of known composition, either previously cataloged in a database such as Genbank, or by sequencing the full or partial genome of an independently identified sample⁷⁵. International open access databases provide easy to use search engines and now contain over a billion annotated sequences from 2 million taxa, providing an unequaled resource to scientists worldwide. While most of these sequences relate to other research interests, tens of millions have been submitted specifically for the use of barcoding all the species that inhabit our world, furthering the capability to identify organisms based solely on DNA samples. Barcode sequence regions are numerous, and selection of a target region is heavily dependent on the species groups of interest and the identification resolution needed. For the COI gene briefly described above, which is present in all eukaryotic mitochondrial genomes, coverage has been rapidly expanding since its discovery as the ideal barcode region, but numerous taxa have still not been sequenced⁷⁶. *Paramoeba* sp. are fairly well cataloged, with 50% having complete COI sequences in Genbank, but under 20% of the species within the family, *Paramoebidae*, are represented. The 18S rRNA gene is more widely sequenced across eukaryotic species, including *Paramoeba*, but it is less capable of

distinguishing between taxa at the species and population level. As such, it provides a useful methodology for the calculation of taxa abundance and diversity within a larger system, as well as the potential for identifying unrepresented COI species⁷⁷.

Within each barcoding gene, there are specific regions which are highly conserved and others which are highly variable, providing the structure on which the sequencing chemistry and bioinformatics must take place. For the method to be effective, the leading and trailing fragment must be highly conserved to ensure that all species which contain the gene have the possibility of being bound by the primers and amplified for sequencing. The barcode sequence in between contains the highest variability to ensure maximum differentiation between species. Incased in the COI gene, the 313bp "Leray fragment" has been established as a highly effective region for species identification⁷⁸. Modifying primers with the addition of degenerate bases to ensure the amplification of previously mismatched invertebrate species DNA, the "Leray-XT" primer set maximizes the number of potential target COI sequences for amplification, sequencing, and identification^{79,80}. With next generation sequencing (NGS) platforms, billions of base pairs of DNA can be read and cataloged in a single run, and the chemistry used can be adjusted for specific length of DNA fragments close to the size of the target gene region used for identification^{81,82}. By tailoring the read length to match the small size (~100-300bp) of amplified fragments such those described above, the number of reads for a sequencing run can reach depths into the tens of millions, while the resulting MOTUs are still capable of high resolution taxonomic identification^{83,84}. In order to maximize the number of targeted gene reads during the sequencing process, the specific gene region of interest on the DNA extracted from the environmental samples must be amplified by polymerase chain reaction (PCR). During the PCR process, bulk DNA is split into single strands, then primers designed to amplify the gene region of interest bind to them and begin producing copies tagged with preceding nucleotide sequences differing between each individual sample^{85,86}. Copied DNA fragments are then isolated from the remainder of the bulk DNA, and pooled for multiplex sequencing, where they will be identified and assigned to which sample they originated from by their oligonucleotide tags^{87,88}.

While the advances in phylogenetics and sequencing technology have brought about tremendous capabilities in species detection capabilities, there are still some shortcomings to these methodologies. To start, degradation rates of DNA may differ between samples which are being compared due to abiotic factors such as temperature, salinity, pH, or UV exposure to the sampled medium, which can be affected by seasonal and year to year changes during time

series^{73,89}. Additionally, biological factors, such as increased microbial communication via excreted DNA or consumption by filter feeding zooplankton can potentially impact abundance of reads for extracellular DNA shed from larger marine organisms⁹⁰. Furthermore, differences in gene abundance can vary greatly between species and tissue types, complicating calculations of a single species relative abundance or measures of abundance changes between species. Hence, metabarcoding is limited in its ability to ascertain absolute abundance or biomass measures for species detected, but instead can be used to estimate semiquantitative relative abundances within closely related species for analysis over time and/or space.

Objectives

The primary objective of this master thesis was to use metabarcoding methods to detect whether *Paramoeba perurans* is present in the water and sediments surrounding an aquaculture facility in Arctic Norway. Dependent on metabarcoding results and fish health status over the course of the study period, further aims of this research included the characterization of potential reservoirs and transmission vectors for the amoebic gill disease agent. Additionally, establishing an understanding of the diversity of *Paramoeba* present within an aquaculture environment and its surrounding area was also a priority, with the purpose of identifying predominant and potentially emerging strains.

Hypotheses

Based on unofficial reports of AGD outbreak from other farms in the region during previous years and the findings of Peters *et al.* 2018 on eDNA metabarcoding uses, I hypothesize that *P. perurans* and other *Paramoeba* pathogens will be detected in the environment samples collected for this project. I further hypothesize that a semiquantitative analysis of the distribution of these pathogens will support previous findings that these organisms are present in the sediments in greater abundance than the water, and that the benthos may act as a reservoir. I also theorize the analysis will show that that physical variables such as temperature may affect the temporal patterns of pathogen abundance. Lastly, phylogenetic analysis of COI Leray-XT metabarcoding MOTUs will reveal inter and intraspecific genetic diversity of *Paramoeba* spp. present in the Arctic.

METHODS

Development of COI/18S reference sequences

Two cultures of *Paramoeba perurans* were obtained from the University of Glasgow, Scotland, which originated from local outbreaks of AGD at aquaculture facilities in the North Sea. These cell cultures underwent DNA extraction and library prep within the UiT Genetics Group labs and were sent off for shotgun sequencing at NOVOGENE, China using an Illumina HiSeq4000 platform and 150 bp paired-end chemistry. Sequence data from the culture samples was filtered from bacterial and algal feed DNA, and full mitogenome assembly was attempted using *GetOrganelle*⁹¹ to map reads against reference genome seeds from *P. pemaquidensis* and *P. aparasomata*. Unsuccessful trials using this method led to the adoption of a more manual approach.

Raw reads from shotgun sequencing of both *P. perurans* cultures mentioned above were independently matched against COI and 18S reference libraries using Magic-BLAST (NCBI) software⁹². Reference libraries (Appendix A) were assembled from sequences obtained through literature, NCBI Genbank and BLAST searches. 18S gene libraries consisted of numerous strains of *P. invadens*, *P. karteshi*, *P. perurans*, *P. pemaquidensis*, *P. aestruarina*, *P. atlantica*, and *P. branchiphila*. Queries for *Paramoeba* spp. COI sequences were less forthcoming, with only limited strains of *P. aparasomata*, *P. branchiphila*, *P. eilhardi*, *P. karteshi*, *P. perurans*, and *P. pemaquidensis* available. Matched reads obtained from magic-BLAST were used to assemble gene sequences for phylogenetic analysis and for addition to the local metabarcoding reference libraries.

Description of aquaculture site

Sampling took place at the Havbruksstasjonen i Tromsø's site for pen raising sea phase salmonids, which is located adjacent to the shoreline of Skogshamn, Dyrøya (Figure 3). The aquaculture research station site, hereby referred to as Skogshamn, operates under normal commercial practices, housing up to 5200t of Atlantic Salmon in 6 net pens for their entire saltwater phase. The net pen construction which began in late August was performed by Salmar AS, which also is responsible for the everyday operation of the farm. Post-smolts were purchased from a commercial smolt provider by Havbruksstasjonen i Tromsø and transported to the farm in 3 shipments between 25/09/2020 and 03/10/2020.

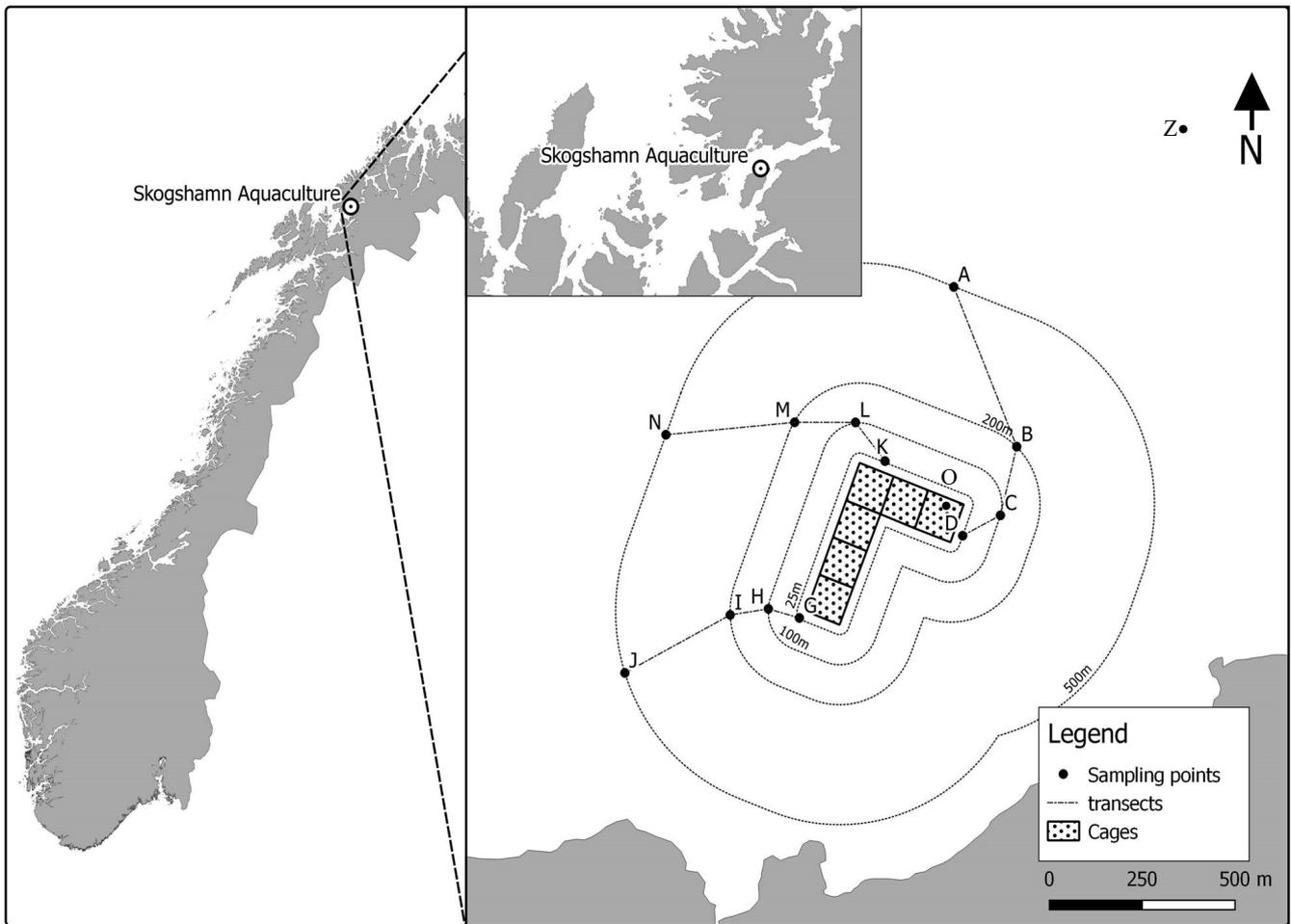


Figure 3 Location of the Skogshamn aquaculture research station and sampling points as they relate to the greater Norwegian coastline. Transects connecting sampling points radiate west (G-J), north (K-N), and east (D-A) from distances of 25m to 500m, as shown. Point O, taken from within the net pen near D, and point Z, taken from 1Km Northeast of the farm are also displayed.

The placement of the facility is in line with normal commercial salmon farm environments, being among five other permitted aquaculture sites in the immediate fjord and 120 in the county of Troms (Figure 4). The bathymetry of Skogshamn, ranging from approximately 200m-350m depth is similar to the other farms in the locality, as well as most others in the northern Norway. Water currents at 5m, 15m, and the bottom of the nets have a general southwest->northeast and northeast->southwest flow, ranging from 15cm/s at ebb and flow tides, to a mild NE flowing current at neap tide (Figure 5). The surface water of the fjord is highly influenced by the strong winds which blow seasonally in the region. Overall, this aquaculture location reflects the conditions of most open net pen salmon farms in the three northern Norwegian regions of Nordland, Troms, and Finnmark and is a model site for the study of disease interactions within them.

This location on Solbergfjord has been used to farm salmon for research purposes in rotation with 2 others in Troms county since 1986 and undergoes the same disease monitoring programs mandated for all other Norwegian aquaculture operations pursuant to *Regulations on the marketing of aquaculture animals and products of aquaculture animals, prevention and control of communicable diseases in aquatic animals* (FOR-2008-06-17-819)⁹⁴. Information regarding results of past and present monitoring were obtained from public resources (Barrentswatch.no/fiskehelse, Yggdrasil.fiskeridir.no)^{93,95}.

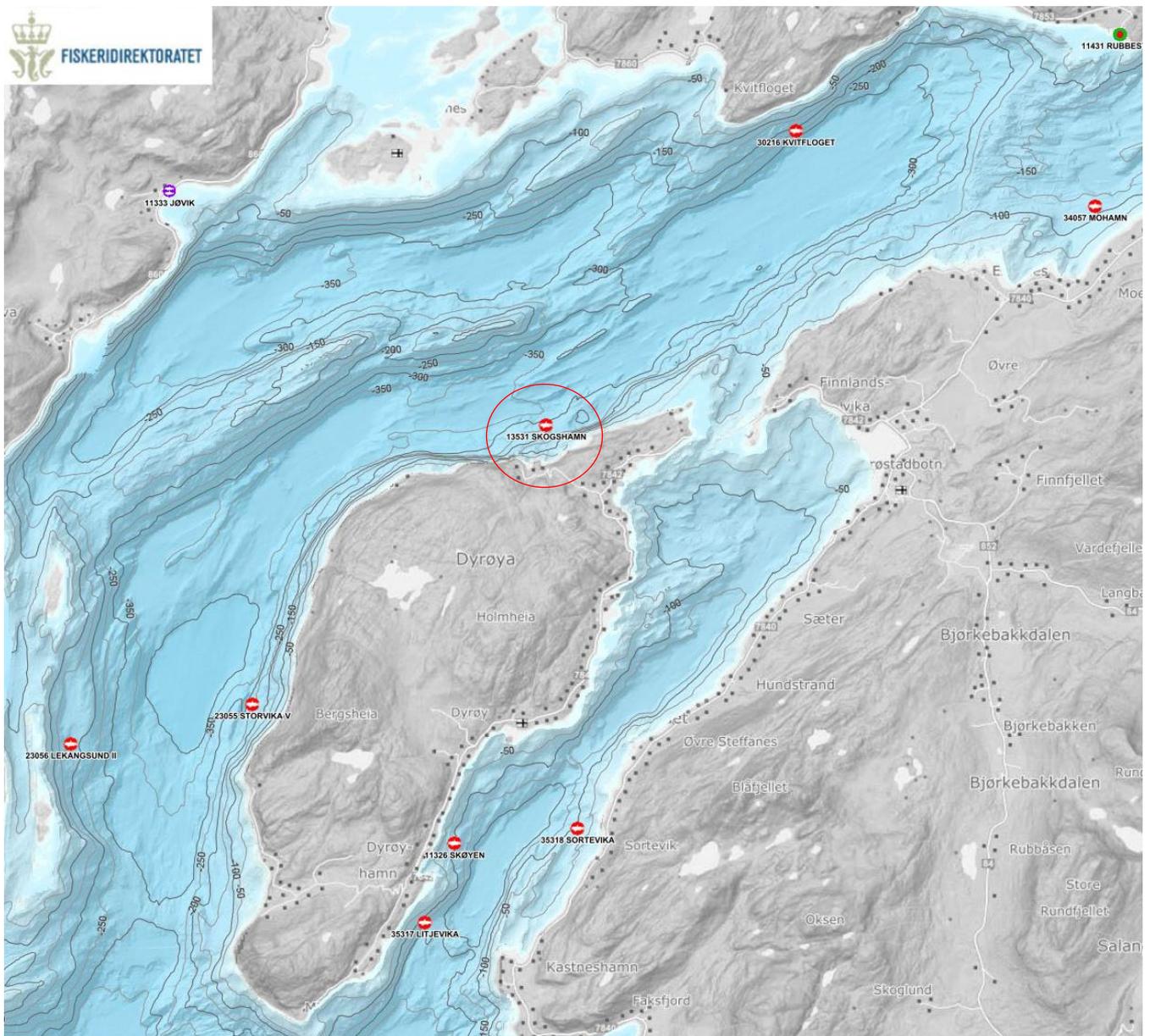


Figure 4 Solbergfjord bathymetry and commercial scale salmon aquaculture farm map⁹³. Red dots indicate permitted open sea cage aquaculture facilities and Skogshamn is circled in red. Grey and black lines mark 50m bathymetric zones overlaid on blue shading displaying bottom aspect.

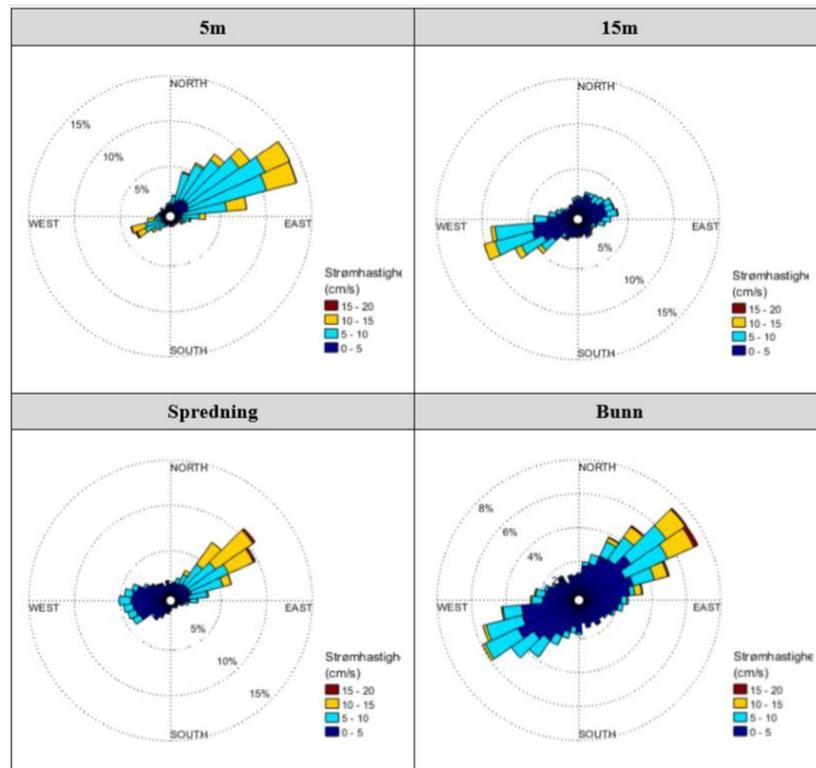


Figure 5 Circular barplot of Skogshamn current velocities and directions by depth. Current velocities (cm/s) are represented by color, and their % occurrence in each direction at those velocities is indicated by each bar length extending from the focal point. Four plots are displayed for four depths where this data was collected at the farm (5m, 15m, net spreading, and fjord bottom)

Sampling design

A rigorous spatial and temporal sampling strategy was designed to measure seasonal effects on pathogen abundance within the area of the farm, as well as the potential for disease spread. 12 sampling locations were selected at four distances from the farm: 25m, 100m, 200m and 500m in three directions: southwest, northwest, and northeast (Figure 3). The directions and distances between these points were chosen to establish a spatial system for measuring the effects of current and source pathogen abundance on disease transmission within the environment. The samples collected at each of these sites included surface seawater and sediment grabs.

Water sample collection was conducted twice a month for the six 25m and 500m distance sampling points (Figure 3) during the setup of the salmon pens and during the initial production phase (Sept 2019-Apr 2020) of the smoltified salmon. A bimonthly frequency was chosen for these samples to ensure finer resolution in abundance changes during the onset of a potential disease outbreak. The six 100m and 200m sample points (Figure 3) were collected from monthly during the same period. All 12 points were collected from before any commercial activities took place at the site in June 2019. Additional samples were collected from a point (Z) 1000m away as a form of seasonal control on background abundance of pathogens present

in the surface water of the fjord. Samples (O) were also collected from within a cage during the season of peak water temperatures, which is usually the highest risk of AGD outbreak. An additional sampling of point O was also taken following notification from farm technicians of a potential bacterial disease outbreak in Jan/Feb 2020. All environmental samples obtained for this project are listed by date of collection in Table 1.

Sediment samples were collected from the 12 primary sampling points (Figure 3). Collections took place immediately before the start of production in fall 2019 and again during sampling trips in February and March in order to detect long term sedimentation of pathogens. The collection of the 12 sediments were split between two sampling trips on each occasion due to

time limitations on boat use, and the occasional malfunctioning of the sediment grab. A table outlining the full water and sediment sampling regime can be found in Appendix B.

Table 1 Skogshamn sampling schedule. Categorization of each date (Before, After). Number of samples collected on each date for either water or sediment. Distances and alpha characters representing the points collected on each date.

Sampling Date	Water	Sediments	Distance (m)	Sampling Points	
Before	30 th June	12	25, 100, 200, 500	A,B,C,D,E,F,G,H,I,J,K,L	
	<i>Mooring blocks, signal buoys, and net pens were installed at the facility</i>				
	12 th September	12	0, 25, 100, 200, 500	A,B,C,D,G,H,I,J,K,L,M,N,O	
	12 th September		10	25, 100, 200, 500	A,B,C,D,G,H,I,J,K,L
	20 th September	6		0, 25, 500	A,D,G,J,K,N,O
21 st October		2	200, 500	M,N *	
<i>Salmon smolts were introduced at the facility</i>					
After	3 rd October	12	25, 100, 200, 500	A,B,C,D,G,H,I,J,K,L,M,N	
	21 st October	6	25, 500	A,D,G,J,K,N	
	5 th November	12	25, 100, 200, 500, 1000	A,B,C,D,G,H,I,J,K,L,M,N,Z	
	19 th November	6	25, 500, 1000	A,D,G,J,K,N,Z	
	12 th December	12	25, 100, 200, 500	A,B,C,D,G,H,I,J,K,L,M,N	
	24 th January	6	25, 500	A,D,G,J,K,N	
	12 th February	12	0, 25, 100, 200, 500	A,B,C,D,G,H,I,J,K,L,M,N,O	
	12 th February		2	500	A,N
	12 th March		10	25, 100, 200, 500	B,C,D,G,H,I,J,K,L,M

*These samples had to be collected on a later date, but were considered in the temporal period “Before” because of the relatively close proximity to the original collection date.

Sample collection

Seawater collection & filtration

Seawater samples were collected manually by a 2.5L Model 1010 Niskin Water Sampler (General Oceanics, Miami, FL, USA) from ~2m depth. The water was transferred directly from the Niskin bottle to a sterilized 2.04L Whirl-Pak™ Stand Up Bag (Nasco, Ft Atkinson, WI, USA), then sealed for short term storage (1-2hrs) on board the zodiac that was used for sampling. Sampling equipment was sterilized with 5 % bleach solution before each sampling event and thoroughly rinsed with seawater from the sampling point area before each use. All samples were collected and processed while wearing newly donned protective equipment such as nitrile gloves to prevent risk of contamination between samples or from outside sources.

Upon collection of the 12 seawater samples, a filtering station was set up on site, and each sample bag was filtered through three 0.22µm Sterivex™ filter units (Merck KGaA, Darmstadt, Germany) using a multichannel peristaltic pump. The output volume of each filter was monitored, and filters removed from the pump at 0.5L to ensure a standard volume between the pseudoreplicates of each sample. After drying the filters by pumping air through them, the filters were placed in pre-labeled sterile 50ml Falcon tubes (Thermo Fisher Scientific, Waltham, MA, USA), and pre-labeled bags for transport to UiT and long-term storage at -80°C in an eDNA only freezer. The filtering station, pump tubes, and operators' hands were meticulously sterilized between each sample using 5% bleach solution and a MilliQ Ultrapure deionized water rinse to limit contamination. A control blank was run on each sampling day to quantify contamination during the filtering process by pumping 0.5L of the remaining MilliQ rinse water through a filter and drying the filter in the same manner as the previous samples. A detailed protocol of the sampling and filtering process described above can be found in Appendix B.

Sediment collections

Sediments were collected from the seabed using a 250cm² Van Veen Grab (KC Denmark, Silkeborg, Denmark) and a gas-powered winch mounted to the zodiac. The grab usually produced ~2L of dense sediment, which came aboard the vessel sealed within the closed grab, in an intact layered state. Three replicates were taken from each sediment collection through 3 of 4 of the top opening doors of the Van Veen Grab. Three coring tubes were created during sediment retrieval by cutting the bottom cone from pre-labeled 50ml Falcon tubes and slowly twisting them cap end downwards into the sediment. Sealing the hole with a nitrile gloved finger created adequate suction to withdraw the sample from the sediment, then the hole was

sealed permanently with duct tape, and the triplicate samples were bagged for transport to UiT for long-term storage at -80°C in a specimen archive. All sediment sampling equipment was sterilized with 5% bleach spray before each sampling event and rinsed thoroughly with seawater between each sampling point.

Other samples and data

Feed samples were collected and saved by technicians during each change in formula or pellet size. Changes in feed formula were expected to occur on a monthly basis for both the main culture species, *S. salar*, and twice for the cleanerfish, *C. lumpus*, over the six month grow-out period when sampling occurred. Schedules were allotted so that any sign of diseased fish at the site would trigger a visit from one of the project participants for additional water sample collections, gill or other tissue swabs, and any specific specimen collections of interest. Mortalities related to any disease events were to be recorded and frozen by on site technicians for later necropsy and molecular analysis at UiT.

Factors such as salmon biomass within each pen, feed usage, current velocities, and other water properties were also collected from Salmar technicians who were on-site daily, managing the day to day growth and health of the fish.

Laboratory practices

Sediment and feed sample extractions

Prior to extraction, laboratory workspace was cleaned and rinsed with 5% bleach solution, MilliQ water, and 70% ethanol, then further sterilized with UV light for 10min. Following thawing of samples and the above cleaning steps, a 0.3g subsample was taken from each of the three replicates from the 12 sampling points which were collected once during the summer/fall and once during the winter/spring seasons. Community DNA and settled eDNA from sediment subsamples was isolated and extracted using DNEasy Powersoil® (Qiagen, Hilden, Germany) kits with a modified protocol (Appendix C). The final DNA product from each subsample was stored with 100µl of elution buffer solution in a pre-labeled 2ml Eppendorf tube in a cryobox at -40°C. An extraction blank was also created during each day extractions took place to control for any residual contamination which may have been introduced during the DNA extraction process. The same kits and protocols were also used to extract DNA from the various formulated pellets used to feed the salmon and cleaner fish during the 6 months between sediment sampling events. In all, 72 sediment samples and four feed types were extracted and sequenced, as well as three blanks.

Water sample extractions

The Sterivex filters used for water sampling underwent DNA extraction in over-pressured eDNA clean-labs using trace eDNA extraction protocols (Appendix D) specifically designed to prevent contamination from all airborne DNA present within university facilities or present on the lab user's skin, hair, or breath. These protocols relied on vigilant care for cleanliness within and outside of the eDNA laboratory and avoidance of potential contaminant sources at the university and personal life during the weeks of eDNA extraction lab use. Personnel were restricted to certain areas in the NFH building and followed a marine organism restricted diet on days of eDNA extraction to prevent the collection of any contaminant DNA on one's person. Airborne DNA contamination risks were mitigated through use of a pressure positive eDNA extraction room and airlocked changing and sample preparation room. After abiding by strict airlock entrance rules, eDNA extraction protocols were meticulously followed for the modified use of DNEasy Blood and Tissue® (Qiagen, Hilden, Germany) kits. Due to the enclosed state of the Sterivex filters, an extended incubation time (24hr) was used for full lysis of the particulates captured within the filter membrane. The lysed solution was then centrifuged out of the filter casing and into 2ml Eppendorf tubes following the standard protocol recommended by the extraction kit handbook (Qiagen 2006). Each sample was eluted in 75 µl elution buffer, of which 20 µl was aliquoted for library preparation and sequencing. The remaining 55 µl was labeled in detail and stored at -80°C as stock for future sequencing runs. 363 water samples, filter blanks, and eDNA lab blanks were extracted, but due to coronavirus pandemic limits on university access and maximum staffing densities in laboratories, only 305 samples went on to library preparation and sequencing. Samples taken from June – Feb deemed essential to the surface water spatial analysis were assigned first priority for sequencing as well as both complete sets of sediment extractions, coming to a total of 382. Samples assigned for later sequencing included water samples collected at alternative depths, sampling points, and from March onwards.

PCR amplification, library preparation, sequencing, and bioinformatics

A multiplexing approach was used for sequencing the 382 samples on an Illumina MiSeq next-generation sequencer (Illumina, San Diego, CA, USA). Prior to sequencing, extraction aliquots for each sample were pipetted into separate PCR well plates for 1 step amplification of the target COI gene region, Leray. PCR amplifications for COI samples were conducted in 20 µl reactions containing 2 µl of DNA template, 10 µl of AmpliTaq Gold Master mix, 0.16 µl of Bovine Serum Albumin (20 µg/µl), 1 µl of each forward and reverse primer (5 µM) and 5.84

μl of H₂O. The PCR temperature profile is further described in the protocol for COI metabarcoding (Appendix E). PCR products were added to multiplex sample pools. MinElute PCR purification columns (Qiagen) were used to remove fragments <70 bp and concentrate the pooled DNA. Library preparation was performed with the NEXT flex PCR-free library preparation kit (BIOO Scientific) and exact concentrations were measured by qPCR using the NEBNext Library Quant Kit (New England BioLabs). Finally, pools were sequenced along with 1% PhiX on Illumina MiSeq platform using v3 chemistry (2x250 bp)⁷⁹.

Bioinformatics pipelines were conducted using OBITools v1.01.22 software suite⁹⁶. Quality control measures used included *illuminapairedend* for alignment of paired end reads, *ngsfilter* for demultiplexing and removal of primer sequences, *obigrep* and *obiuniq* for selection of specified length sequences and dereplication. *Uchime denovo (vsearch)* was then used to remove chimeric sequences and *SWARM 2.0* clustered sequences into Molecular Operational Taxonomic Units (MOTUs) with specific d value 13 for COI. Taxonomic assignment of MOTUs took place using *Ecotag* against a local database of Leray fragment sequences.

Each COI MOTU of interest was manually checked for better match by BLAST search of the recorded Leray sequence, and best IDs were changed to reflect a higher percent match if one was found. MOTU best IDs were then reassigned to an appropriate taxonomic level based on percent match to the assigned species. A threshold of 97% for species level assignment was established based on methods from Ammon *et al.* 2018⁹⁷. Anything below that threshold was assigned to the last common ancestor of the top matched results from the BLAST search described above. The entire MOTU data table for Skogshamn COI reads was summed for total reads from each sample for later use in normalization, then MOTUs for members of the genus *Paramoeba* were isolated and transferred to separate excel files for analysis.

Statistical analysis

Data manipulation and statistical analysis took place in R software v4.0.0, beginning with summation and normalization of raw sequencing reads. Using base R functions, MOTU rows were summed by their best ID, then the number of reads for each was converted to relative abundance. The following formula was used to calculate relative read abundance:

$$RA_R = \frac{n_{i,k}}{\sum_{i=1}^T n_{i,k}} \times 10,000$$

where n is the number of reads for each MOTU, i , within a particular sample, k ⁹⁸. A constant of 10,000 was multiplied against the result to adjust relative read abundance to improve

readability and simplify downstream presentation. An analysis of variance was conducted for each subset of the overall data table by variable, for example transect direction, to test the significance of its effect. Tukey's "Honest Significant Difference" method was then used to test for confidence intervals within levels of each variable if significance was established and more than 2 levels (e.g. west, north, and east) occurred. The data values were then visualized using *ggplot()* and a combination of violin, jitter, and point layers using the *ggplot2* library. This analysis was conducted for each of the designed spatial and temporal variables Distance, Direction, and Date for water and sediment samples. The environmental variables Bottom Depth and Water Temperature were used for analysis of reads from sediment and water samples, respectively. The analysis described above was run on the whole *Paramoeba* MOTU dataset as a single summed Best ID of *Paramoeba* sp. A full script for the statistical analysis and generation of results figures can be found in Appendix F. Specific MOTUs of interest were determined based on the initial statistical and phylogenetic analysis, then exploration of those individual MOTUs' spatial and temporal distribution within the sample types was attempted to investigate potentially divergent trends. Due to limited reads and similar trends observed within the *Paramoeba* MOTUs, these exploratory analyses were not used for further data presentation and coding removed from the final script.

Phylogenetic analysis

The Skogshamn COI Leray-XT MOTUs assigned to *Paramoeba* spp. were uploaded to MEGAX⁹⁹ and aligned with the *Paramoeba* spp. reference library COI genes obtained and curated as previously described. Alignments were conducted with the Muscle algorithm for nucleotides using the UPGMA clustering method under standard settings. Phylogenetic analysis was also conducted using MEGAX data explorer tools. 500 replicate bootstraps¹⁰⁰ were run using the maximum likelihood method to test for robustness of trees. The Tamura-Nei model¹⁰¹ was used to assess nucleotide substitutions at a uniform rate between all sites and the nearest neighbor interchange strategy was used to simplify tree rearrangement for each branch during each bootstrap replicate.

RESULTS

Culture mitogenome assembly

Initial efforts to assemble mitogenomes from the two shotgun-sequenced *P. perurans* cultures against existing *Paramoeba* spp. mitogenomes returned promising contigs of ~45000bp, but MEGAX alignment of the contigs with COI sequences from the gathered reference libraries of

Paramoeba spp. was unsuccessful. External BLAST searches of both contigs returned negative results, with the closest matches registering as ~80% identical to a variety of proteobacteria, indicating potential additional contamination of the extracted sample with bacteria present within the cultured organisms.

Results from the magicBLASTing of raw shotgun sequence reads were more successful. Though no reads were matched to COI genes for either of the species groups searched, 146 paired end matching reads were obtained with identity values over 80%, of which 45 were matched with *Paramoeba* spp. and one with a *Perkinsela* spp. for a fragment of 18S gene. The remaining 100 reads were matched to the full mitogenome seeds of *P. aparasomta* and *P. pemaquidensis*, which were also included in the Magicblast reference libraries. The single *P. pemaquidensis* read and 99 *P. aparasomata* mitogenome matched reads were insufficient for a COI gene assembly and therefore discarded, as was the one *Perkinsela* sp. match. These results will therefore not be used or discussed further within this study.

COI barcode assignments

Querying raw COI Leray reads against the local reference database lead to the assignment of 12 MOTUs to the two divergent *Paramoeba pemaquidensis* reference sequences, with matches ranging from 89-95%. After individually BLASTing each MOTU for assurance of the best possible ID, only one was assigned to a better match, increasing to 90.7% identity for *Paramoeba longipdia* (Volkova & Kudryavtsev 2017), a newly described species discovered from sediments in 5.1km deep water in the western Atlantic ocean (Table 2). All MOTUs identifications fell below the 97% identity match threshold for species assignment. However, this threshold, which is commonly used for metazoans, is far from being established for amoebozoans. Therefore, all MOTUs were assigned to the genus level, *Paramoeba* sp., for downstream analysis.

Total reads for each MOTU ranged from 5 to 352, with the maximum for a single MOTU in a single replicate reaching 50 reads in Sediment D1 12/2/20 (total 102 reads), and 25 in water sample C1 12/12/19. All control blanks and feed pellets returned 0 reads for *Paramoeba*, indicating no contamination of samples with pathogenic material during filtering or extraction steps. Of the remaining samples, 47 of 72 sediments contained reads from *Paramoeba* MOTUs, but the vast majority of water samples, 276 of 295, did not.

Table 2. Skogshamn COI Leray-XT MOTUs assigned to *Paramoeba* spp. by highest percent match against NCBI BLAST database. All reads displayed are not yet normalized and are shown as a sum of all samples, by sediment collection date (Before & After), and the total from all water samples for comparison.

MOTU ID	Percent Match	Species Name	Ascension Number	Total Reads	Sediment Before	Sediment After	Water
SKOC_001212579	89.17%	<i>Paramoeba pemaquidensis</i>	MN025475	43	32	11	0
SKOC_000217484	94.57%	<i>Paramoeba pemaquidensis</i>	MK990593	6	0	0	6
SKOC_002580467	92.68%	<i>Paramoeba pemaquidensis</i>	MN025475	85	0	84	1
SKOC_002558372	90.79%	<i>Paramoeba pemaquidensis</i>	MN025475	12	2	10	0
SKOC_002461362	90.45%	<i>Paramoeba pemaquidensis</i>	MN025475	7	3	4	0
SKOC_002439258	89.81%	<i>Paramoeba pemaquidensis</i>	MN025475	250	44	159	47
SKOC_002345164	91.40%	<i>Paramoeba pemaquidensis</i>	MN025475	5	2	3	0
SKOC_002344563	90.76%	<i>Paramoeba pemaquidensis</i>	MN025475	115	34	81	0
SKOC_002332487	91.08%	<i>Paramoeba pemaquidensis</i>	MN025475	10	8	2	0
SKOC_001647717	92.35%	<i>Paramoeba pemaquidensis</i>	MN025475	31	30	0	1
SKOC_001226214	90.71%	<i>Paramoeba longipodia</i>	MF140256	6	0	5	1
SKOC_001212193	91.40%	<i>Paramoeba pemaquidensis</i>	MN025475	352	62	287	3

Distribution of reads between sediments and water

Average reads per sample for all MOTUs detected were approximately the same between sediment samples (mean 17,987 SD 13,106) and water (mean 18,395, SD 15,320), but raw reads of the 12 *Paramoeba* sp. MOTUs were detected at significantly greater absolute abundance in sediment (mean 12.0, SD 21.0) than in water (mean 0.2, SD 1.6) samples ($p < 0.001$). After read values were normalized for each sample replicate, the mean relative read abundance in sediments was even greater by comparison, namely, two orders of magnitude higher than that of the water sampled, 7.2 and 0.072 respectively (Figure 6) ($p < 0.001$).

All MOTUs except SKOC_000217484 were found in the sediments, which was only detected in water samples on Sept 12th. Only six of the 12 MOTUs were detected in the water samples, of which three had only one read (Table 2). By far the most prevalent MOTU in the water was SKOC_002439258, which accounted for 80% of all *Paramoeba* sp. reads in water samples and 23% in sediment. Five MOTUs (SKOC_002580467, SKOC_002439258, SKOC_001647717, SKOC_001226214, SKOC_001212193) were found in both sample types, with three of the four overall most abundant MOTUs contributing.

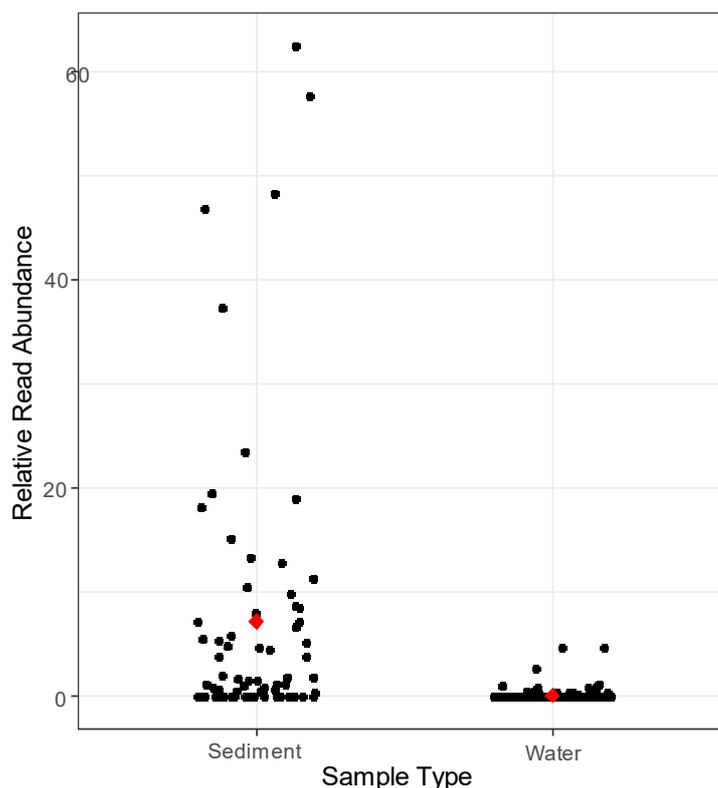


Figure 6 Relative read abundance of *Paramoeba* sp. by sample type. Each black dot represents a relative read abundance value for each replicate of each sampling point collected over the course of the study. Red diamonds indicate the calculated mean relative read abundance for all sampling replicates of each sample type.

Of the five MOTUs present in both sediment and water samples, four MOTUs displayed an increase in their number of reads in the sediment samples after the 5 months of aquaculture activities (Table 2). SKOC_002439258 and SKOC_001212193, by far the most frequently detected MOTUs overall, were both detected approximately four times as often in sediments after, though neither reached the statistical significance threshold for a temporal effect ($p=0.09$, $p=0.12$). The SKOC_001226214 and SKOC_002580467 MOTUs were only detected in the “After” sediment samples. The last MOTU, SKOC_001647717 was the one MOTU of the five from which all sediment reads came from before aquaculture activities began, and none after.

Distribution of sediment COI reads over space and time

A visible decrease in relative read abundance of *Paramoeba* sp. can be seen as distance from the farm increases (Figure 7). Due to the frequency (25/72) of 0 read values across all distances, the analysis of variance within this variable did not yield statistical significance ($p=0.34$). 50-83% of samples within the four distances registered *Paramoeba* sp. reads, but most contained <10 relative reads. Of the 14 sediment sample replicates that had relative read abundances above this threshold, 13 were detected within the 3 closest distances, 25m, 100m, and 200m (Figure 7). Similar trends were observed for distance effect on sediment reads when

the data from the two collection periods were explored independently, but both before ($p=0.71$) and after ($p=0.53$) produced results lacking statistical significance.

Directional effects on relative read abundance of *Paramoeba* within the sediment was observed between the three transects ($p=0.06$), with the northern transect samples reporting the lowest relative read abundances, followed by samples from the eastern transect, while sediment grab samples from the western transect displayed the highest relative read abundances (Figure 8). Direction was considered closely associated with approximate bottom depth at the point of collection, as all points on the western transect were shallower (192-271m) than 300m, all northern points were deeper (353-412m), and collection points on the eastern transect were split (291-315m). When relative abundance was tested by binomial depth category, shallow samples showed a significantly ($p=0.004$) higher relative abundance of *Paramoeba* reads, compared to samples collected from >300m depth (Figure 9).

The relative read abundance of *Paramoeba* within the sediments showed an increase of ~50% increase over the five months of aquaculture activities (Figure 10), though this trend did not have statistical significance ($p=0.15$).

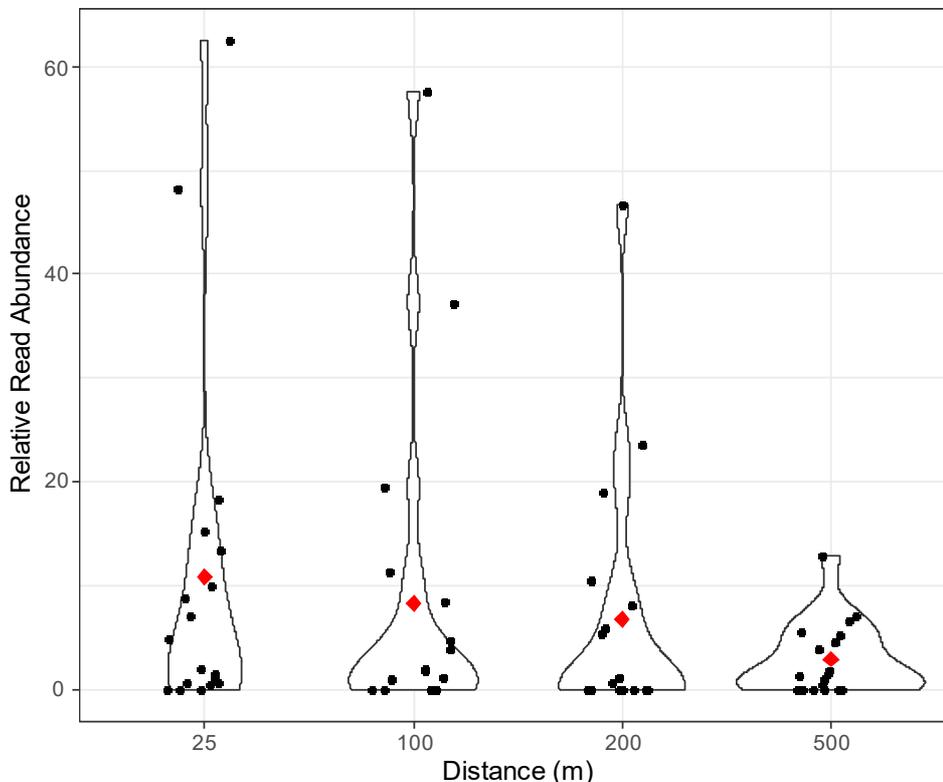


Figure 7 Relative read abundance of *Paramoeba* sp. in sediments by distance from aquaculture pens. Each black dot represents a relative read abundance value for each replicate of the 3 primary sampling points collected at each distance over both collection periods. Red diamonds indicate the calculated mean relative read abundance for each distance.

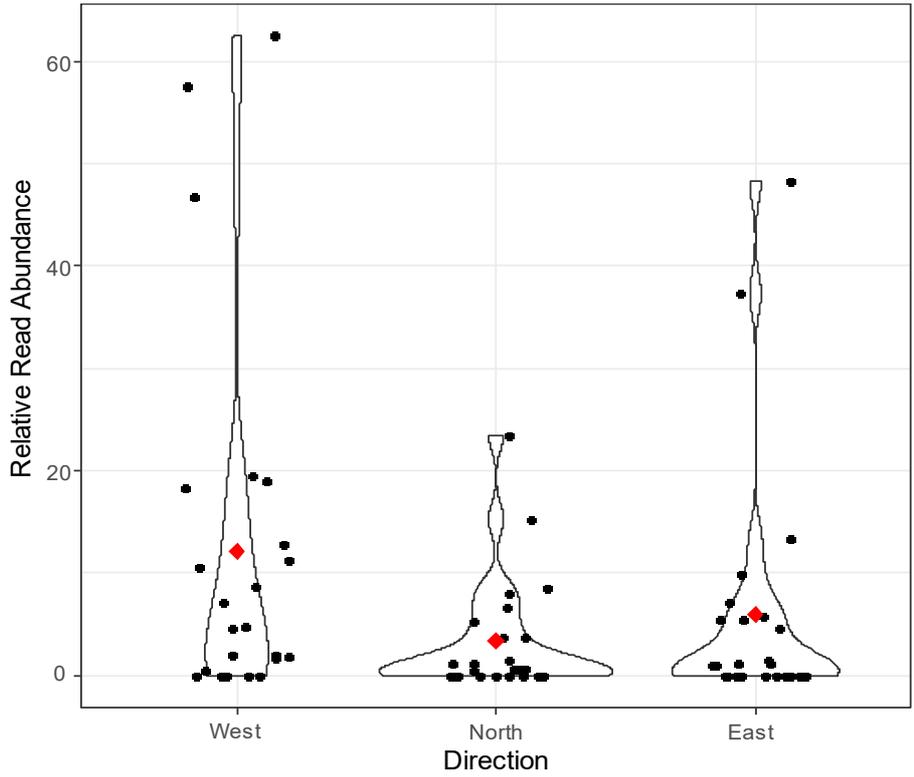


Figure 8 Relative read abundance of *Paramoeba* sp. in sediments by directional transect. Each black dot represents a relative read abundance value for each replicate of the 4 primary sampling points collected on each transect over both collection periods. Red diamonds indicate the calculated mean relative read abundance for each directional transect.

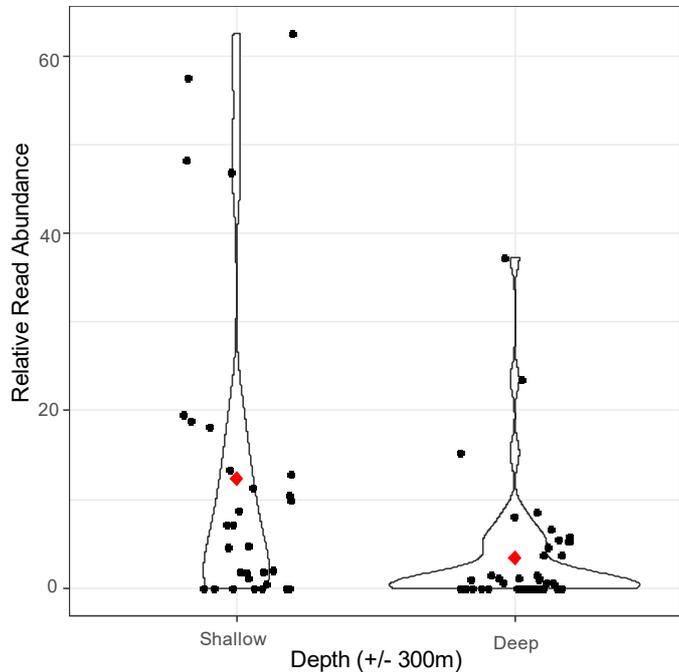


Figure 9 Relative read abundance of *Paramoeba* sp. in sediments by depth. Shallow<300m<Deep. Each black dot represents a relative read abundance value for each replicate of the primary sampling points collected for each of the depth zones over both collection periods. Red diamonds indicate the calculated mean relative read abundance for the two depth zones.

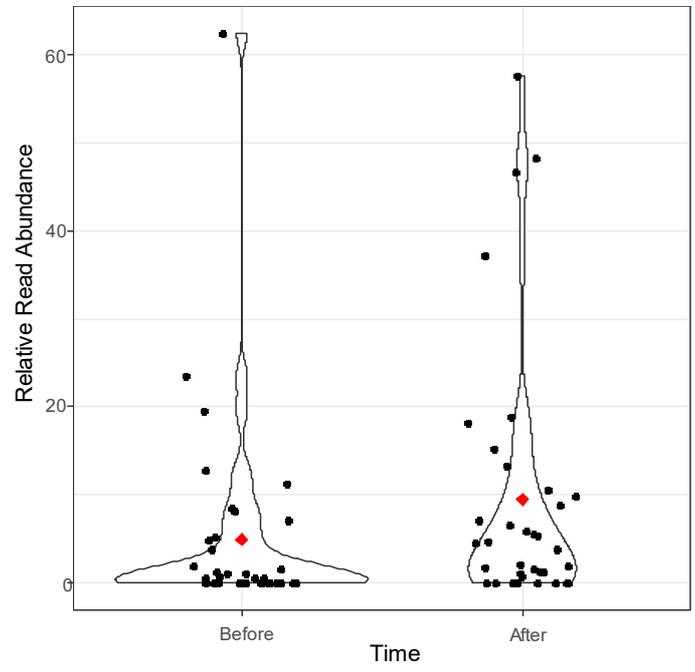


Figure 10 Relative read abundance of *Paramoeba* sp. in sediments by temporal relation to aquaculture practices. Each black dot represents a relative read abundance value for each replicate of the 12 primary sampling points collected during each collection period (Before vs After). Red diamonds indicate the calculated mean relative read abundance for the two collection periods.

Distribution of Paramoeba reads in water samples

Variation in relative read abundance for *Paramoeba* assigned MOTUs was also seen across spatial and temporal factors in the water samples, though the lesser frequency of detection and lower number of reads per detection caused decreased effect strength across all variables.

Mean relative read abundance of *Paramoeba* and the frequency of its detection did not show a clear effect pattern between 25m, 100m, and 200m distance samples when looking across the entire sampling period, but intensity of detection was limited to 1.12 relative reads for 200m compared to 4.63 and 4.71 for 25m and 100m respectively (Figure 11). These three distances all had similarly greater mean relative read abundance than the 500m samples, which had a moderate number of detections, but low intensity (<1 relative read) for each of those occurrences. The 1000m distance sampling point Z did not produce any *Paramoeba* sp. reads, though it was only collected on two sampling dates. Overall, detections were infrequent and therefore effects of distance on relative read abundance were not significant ($p=0.32$) when

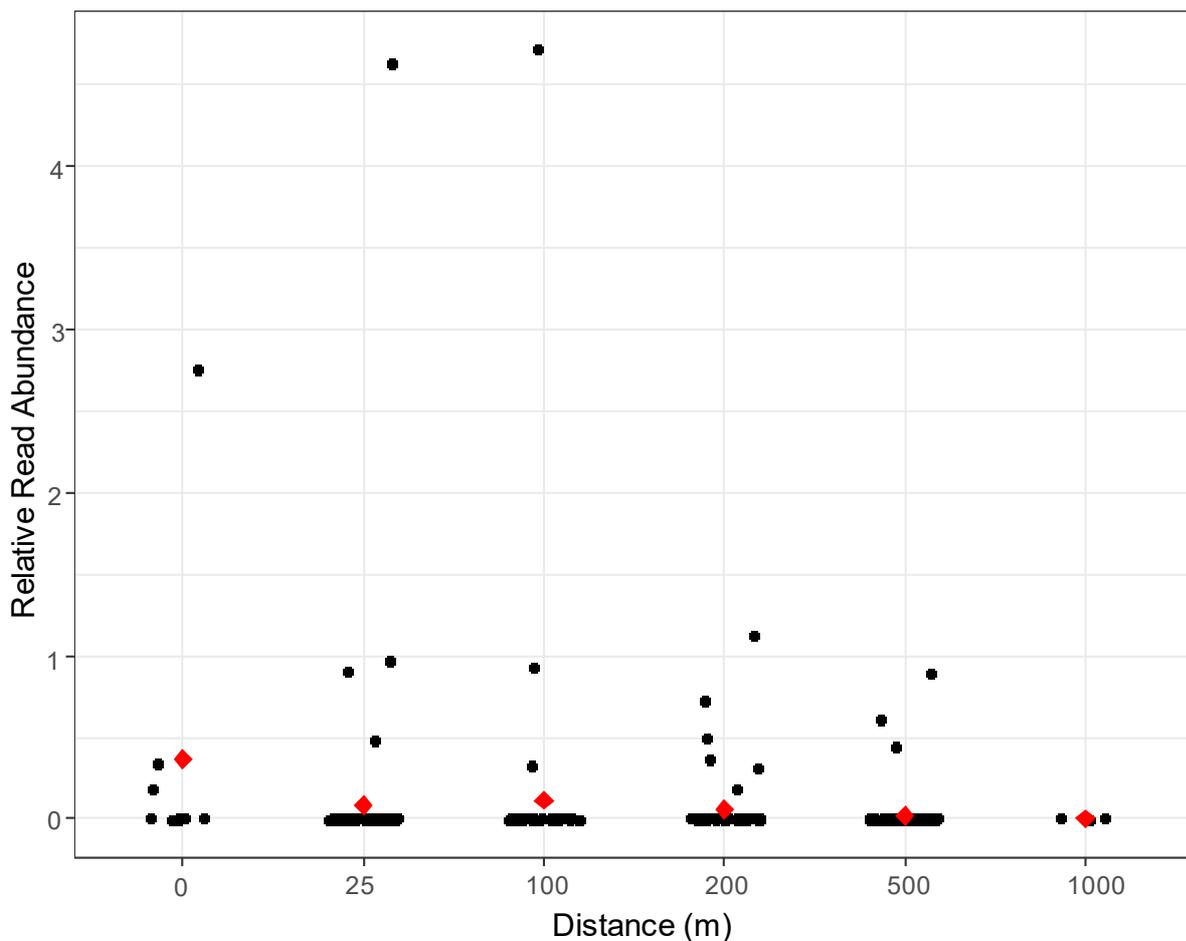


Figure 11 Relative read abundance of *Paramoeba* sp. in surface water by distance from aquaculture pens over all sampling dates. Each black dot represents a relative read abundance value for each replicate of the 3 primary sampling points collected at each distance for each date and the O and Z sampling replicates for the dates they were collected. Red diamonds indicate the calculated mean relative read abundance for each distance over all dates.

looking across all dates. One date, the 12th of September, had a statistically significant distance effect ($p=0.008$), as a result of eight sample replicates registering *Paramoeba* reads across four distances (Figure 12). All three of the replicates from point O on this date contained *Paramoeba* sp. reads, as well as three replicates from two samples at a 200m distance, and one replicate each from the 25m and 100m distances. Analysis of the differences between each distance level by TukeyHSD test produced confidence levels confirming the statistically significant difference between the sample replicates from 0m vs samples from each of the other four distances measured on that date: 25m ($p=0.012$) 100m ($p=0.007$) 200m ($p=0.037$) and 500m ($p=0.005$).

Analysis of the temporal impact on relative read abundance of *Paramoeba* produced no statistically significant effect ($p=0.45$), but patterns of zero *Paramoeba* read dates vs various intensities of detection were noted (Figure 13). An initial peak in reads took place on September 12th, the second water sampling date, shortly after the pens had been placed, but before the smoltified salmon were brought to the station. The second peak in reads occurred over the

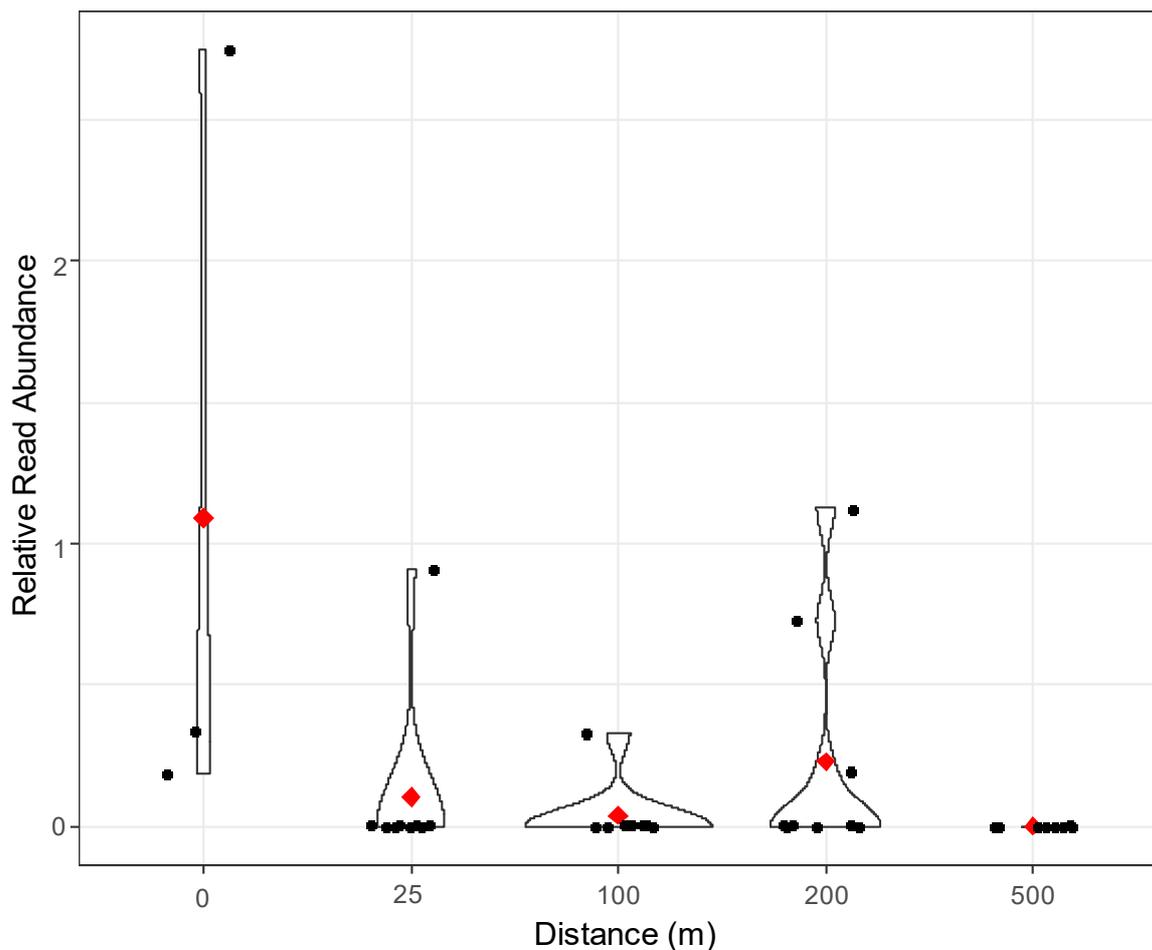


Figure 12 Relative read abundance of *Paramoeba* sp. in surface water on Sept 12th by distance from aquaculture pens. Each black dot represents a relative read abundance value for each replicate of the 3 primary sampling points collected at each distance on Sept 12, plus the 3 O replicates from that date. Red diamonds indicate the calculated mean relative read abundance for each distance over all dates.

course of two sampling dates, Nov 19th and Dec 12th, after the fish had been present in the cages for ~6 to 8 weeks (Figure 14). The September peak was heavily influenced by *Paramoeba* reads from SKOC_000217484, which was detected 6 times in 3 sample replicates on this date, but never during the remaining dates. The second peak is exaggerated by a single high intensity read that occurred during the sample collection date on Nov 19th, when only 25m and 500m samples were taken, and an equally intense read during the complete 12-point sampling event on Dec 12th. No reads were produced from the DNA extracted seawater from the June 30th, Sept 20th, or either of the October sampling events despite these being four of the five dates with the warmest recorded water temperatures (Figure 13). When sample points O and Z are removed, Sept 12th, Nov 5th, Jan 24th, and Feb 12th each contain trace relative read abundances (Mean<0.1) of *Paramoeba* reads.

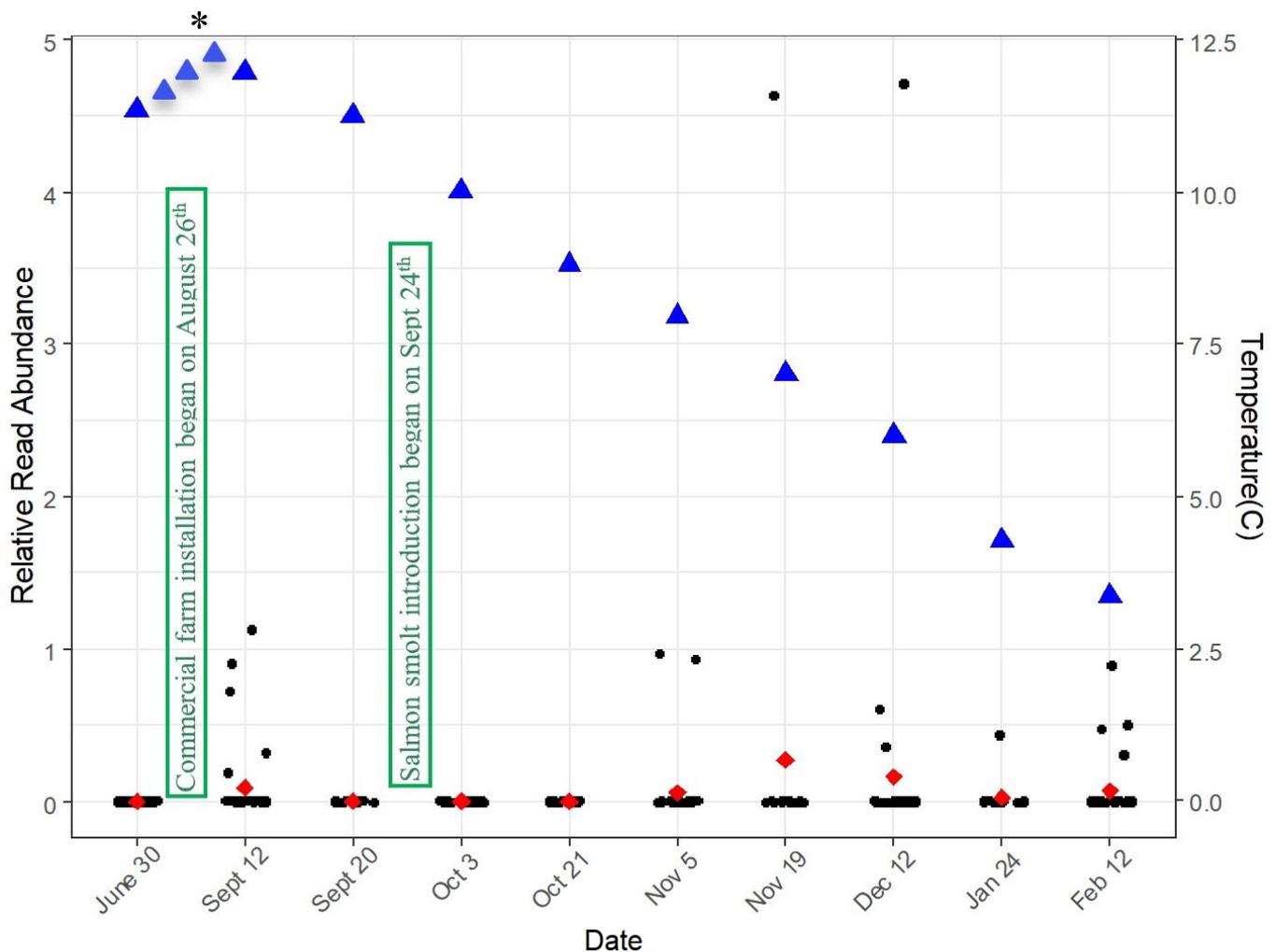


Figure 13 Relative read abundance of *Paramoeba* sp. and sea surface temperature by sampling date. Each black dot represents a relative read abundance value for each replicate of the 6 or 12 primary sampling points collected on each date. Sample points O and Z were excluded due to their inconsistent collection over the entire time series.. Red diamonds indicate the calculated mean relative read abundance for each date. Green text boxes mark significant change in human activity at the site. Blue triangles indicate sea surface temperature readings at Skogshamn or one of the nearby farms from Fig. 3

*Three temperature readings were added for July-August to provide continuity to seasonal changes.

Directional effects on relative read abundance were also observed for water samples, though the trend was opposite the effects in the sediments and not statistically significant ($p=0.45$). Water samples from the eastern and northern transects had greater intensity detection events and resulting higher mean relative read abundances while the western transect samples had only low intensity detections from the few reads that were produced (Figure 14).

Phylogenetic analysis

Alignment of sequenced Leray-XT MOTUs which matched *Paramoeba* spp. reference sequences resulted in the identification of 12 novel haplotypes within the genus (Figure 15), as none of these haplotypes have been previously recorded in published literature or online reference databases. These 12 haplotypes originated from 75 single nucleotide polymorphisms (SNPs) (Appendix G). After being classified by 500 bootstrap replicates of MEGAX maximum likelihood analysis, all Skogshamn *Paramoeba* MOTUs and both *P. pemaquidensis* references were assigned to a single clade, while all other reference species of the genus were assigned to another clade (Figure 15). 99% of replicated analysis shared this finding, confirming the robustness of this assignment. Branch lengths between detected individuals and their immediate

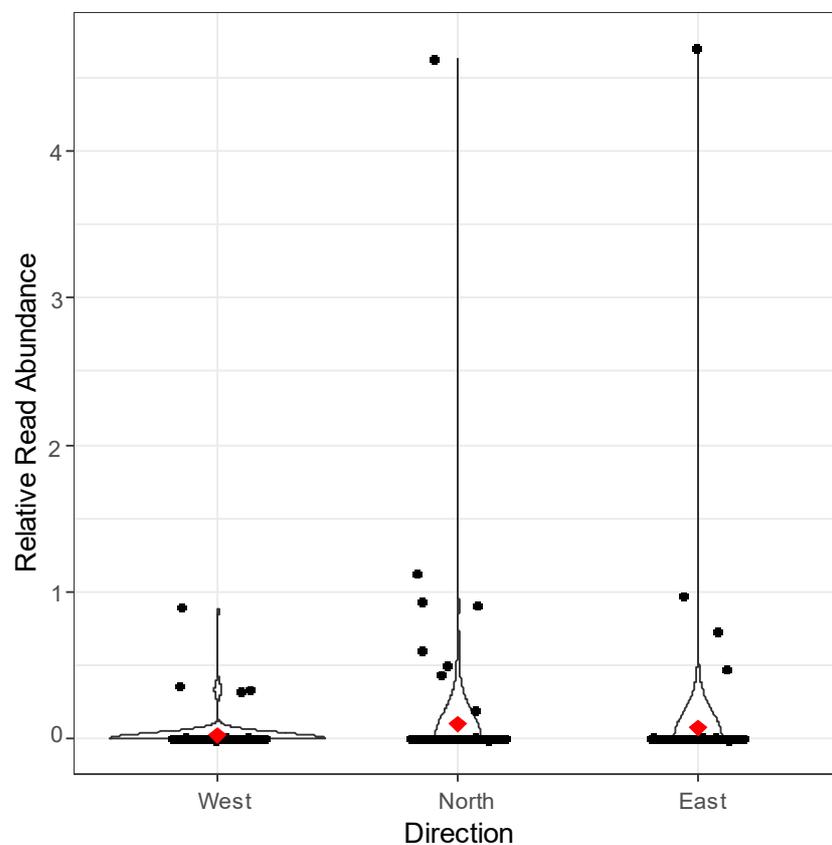


Figure 14 Relative read abundance of *Paramoeba* sp. in surface water by directional transect. Each black dot represents a relative read abundance value for each replicate of the 2 or 4 primary sampling points collected on each transect for each date. Red diamonds indicate the calculated mean relative read abundance for each directional transect over all dates.

calculated ancestral nodes estimate divergences between 1.5 and 5%, compared to <1% between known *P. perurans* haplotypes, 9% in *P. pemaquidensis* strains, and 5 to 15% within each of the three genera represented. Individual MOTU divergence from other *Paramoeba* species similarly ranged from 5-15%.

Bootstrap replication of the ML analysis resulted in maximum confidence for an immediate relationship between *P. pemaquidensis* (MK990593) and SKOC_000217484 with 99% of replicated analysis producing the same result. The second strain of *P. pemaquidensis* (MN025475) used as a reference sequence, which was the Best ID for 10 of 12 MOTUs based strictly on percent nucleotide similarity, was deemed the next closest relative with an also significant bootstrap value of 78. Another 78% replicate agreement in the coupling of MOTUs SKOC_002439258 and SKOC_002558372 also indicates a strong ancestral connection between these two haplotypes, though the relationship with SKOC_002345164, the above-mentioned three MOTUs, and the remaining Skogshamn MOTU clades is unclear (Figure 15).

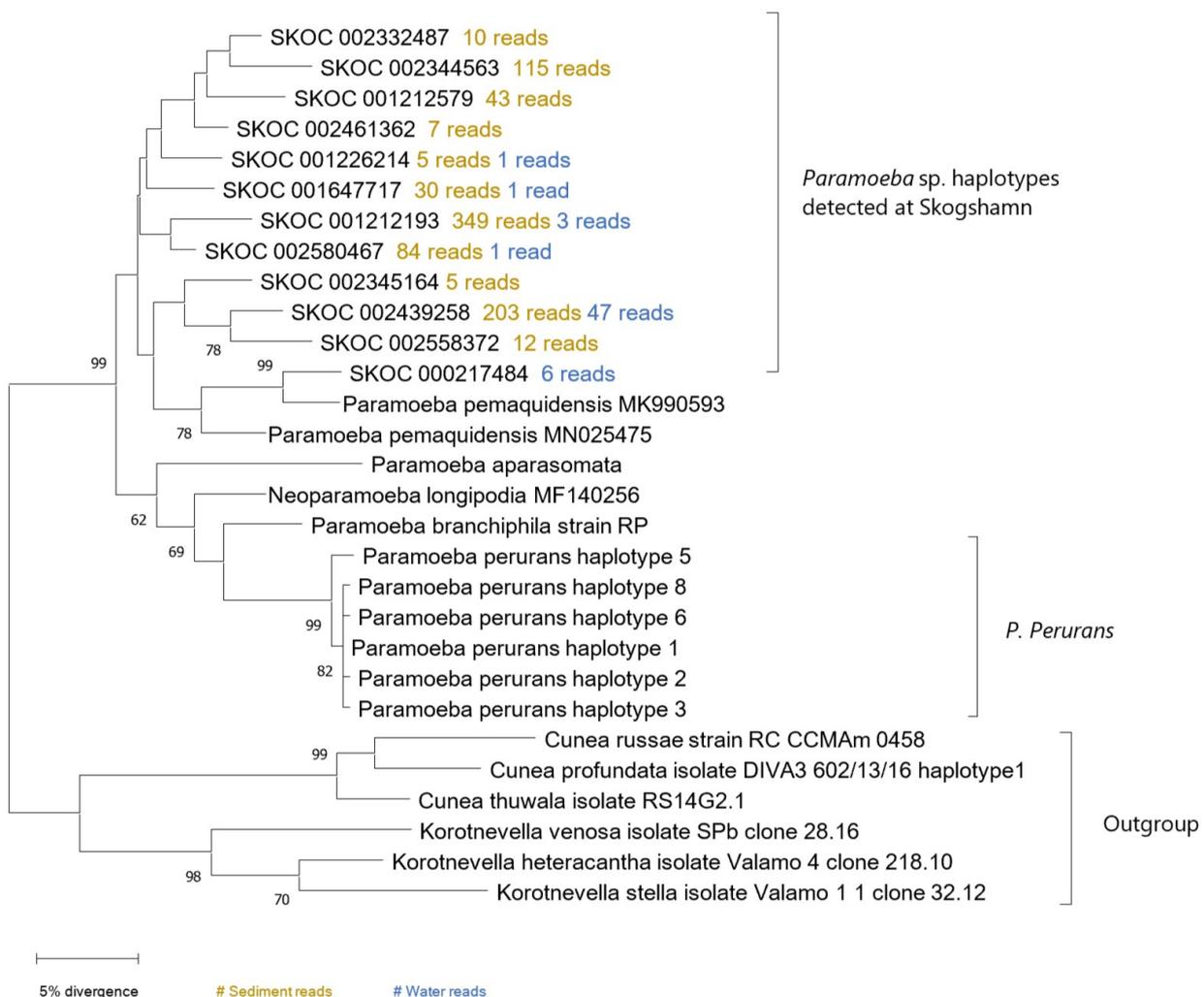


Figure 15 Dendrogram of detected *Paramoeba* COI Leray-XT MOTUs and reference sequences of related *Paramoebidae* species. Nodes with >50% bootstrap support display replicate agreement percentage. Total number of sediment and water reads detected for each MOTU are indicated by brown and blue colored text.

Replicate agreement for the remaining two immediate MOTU-MOTU nodes, SKOC 002344563 - SKOC 002344563 and SKOC 001212193 - SKOC 002580467 dips below 50% significance thresholds to 48% and 44% respectively, but still represents evidence of close genetic relatedness based on reference sequences currently available. Low replicate agreement (2-31%) between many of the ancestral nodes and remaining MOTUs is reflective of the evidence that all 12 detected MOTUs share their closest common ancestry with each other and only 2 of the 11 reference sequences included in the assessment's ingroup. Furthermore, the phylogenetic analysis used only one evolutionary model, was conducted on a very short region of genetic material with a high rate of mutation, and included larger than normal outgroups, all which contributed to an increased variability in the accumulated 500 bootstrap results.

DISCUSSION

This study demonstrates that environmental samples such as surface seawater and benthic sediments can be analyzed to detect the presence of *Paramoeba* pathogens in trace abundances using metabarcoding methods. It also finds that the Leray fragment can be an effective gene region for distinguishing between the specific *Paramoeba* species and their intraspecific genetic variation. While the target pathogen, *P. perurans*, was not detected, there were no clinical signs of AGD outbreak observed at the farm by onsite personnel. However, the MOTUs that were detected and assigned to *Paramoeba* spp. followed previously observed patterns of spatial and temporal distribution for *P. perurans* and other known AGD related pathogens that have been studied in the water column^{54,55}. Additional patterns in spatial and temporal distribution of the species on the benthos were discovered, relating to changes in direction and distance from the farm, bottom depth, and aquaculture activities. The relative abundance of reads for the MOTUs detected herein, as well as the catalog of physical factors and commercial activities that were documented at the farm provide further insight on the potential reservoirs and vectors for AGD related pathogens.

The absence of *P. perurans* detections at Skogshamn failed to fulfill the primary hypothesis of this project. Therefore, we are unable to validate metabarcoding as a method for detecting this specific AGD causing pathogen in environmental samples collected from a salmon farm. Thus far, in studies of *P. perurans* the most successful method for detecting the pathogen in water, wild fish gills, and farmed salmon tissues has been qPCR assay with 18S primers^{42,54,55}. While some previous attempts to detect this pathogen using short DNA fragments have proven unsuccessful due to ineffective primers, phylogenetic analysis of the COI Leray fragment in

Paramoeba spp. conducted for this experiment shows that all available reference sequences for the genus show compatibility, including all known COI haplotypes of *P. perurans*, and that there is suitable inter-species variability within the gene fragment for detection to the species level. These findings, in combination with the lack of *P. perurans* reads, lead to the deduction that this specific pathogen may have been in low enough abundance to evade our limited water sample size or could even be absent from the study site altogether. Though some aquaculture professionals have claimed that this disease has struck northern Norway in recent years, conclusive evidence of an outbreak at this particular farm or even within the Solbergfjord waterway has not yet come forward. Despite the nonappearance of *P. perurans* in detected MOTUs, the presence of related *Paramoeba* spp. is unequivocal.

MOTU prevalence between sediment and water

The detected MOTUs were of closest resemblance to *P. pemaquidensis* – a known pathogen involved in AGD infections in salmon aquaculture worldwide – and were found in both the sediments and water surface over the course of the sampling period. Detections ranged from moderately high intensity in many of the sediment replicates, to single raw reads in most of the water samples. The significantly greater relative read abundance and frequency of occurrence of *Paramoeba* reads in sediments compared to water samples was hypothesized, as *Paramoeba* spp. are known to attach to surfaces and accumulate in sediments and detritus both at open net pen sites and in recirculating aquaculture systems where AGD is present^{52,66}. The low detection frequency of MOTUs in water samples may be explained by the relatively small volume of water collected for each sampling replicate, and/or by the relatively low read depth obtained by multiplexing so many samples. However, these methodical shortcomings must be tested in another study to determine the appropriate size of sample for filtration and optimal read depth for detection of *Paramoeba* or other trace pathogens. Likewise, the homogeneity of amoeba density across the water surface cannot be assumed, due to the effects of wind, eddies, and other turbulent surface currents which may aggregate microbial cells both horizontally and vertically^{102,103}. And this unpredictability was not resolved by sampling replication on each point in this case as sample replicates were generated during the filtration and extraction step. Skogshamn water replicates did not often share detections of *Paramoeba* over the sampling period, but the collection of 12 independent samples on each date was an effected form of replication on the temporal scale, which enhanced the capability of the method to detect very minute changes in already trace abundances over the sampling period. Despite potential detection limitations, the patterns of relative read abundance displayed by *Paramoeba* MOTUs

support the second hypothesis of this study, that these organisms are more abundant in the sediments than in the water column.

Numerous *Paramoeba* spp. have been deemed free-living opportunistic parasites after being found in the environment separate of their associated host, such as *P. aestuarina*, *longipodia*, and *pemaquidensis*, most of which have been isolated from sediments, leading many to characterize the entire genus as such^{37,69}. On the contrary, conflicting evidence has been presented for the remaining members of the genus, who have only been found in environmental samples in the immediate vicinity of a host species or shortly following that host's removal^{55,104}. *P. invadens* and *P. perurans* for example, the two species most associated with diseases outbreaks in marine organisms, have only ever been detected in the water surrounding diseased hosts, or attached to surfaces such as net pens or nearby substrates. Our detection of *Paramoeba* related MOTUs in the water column and sediments before aquaculture practices began support claims that some of the genus, including *P. pemaquidensis*, may remain free-living in the environment for long periods after aquaculture practices end. However, these findings from pre-salmon placement eDNA samples do not support that *Paramoeba* sp. be considered endemic organisms of this fjord. Decreasing read occurrence by distance from the farm site and knowledge of its extensive commercial use over the last 30 years allow for speculation that these pathogens may have been introduced to the site previously. Furthermore, MOTUs detected in the water samples from Sept 12th should not be considered evidence of long-term persistence of *Paramoeba* sp. in the water column because salmon farm vessel activity and placement of net pens had begun taking place by that time. Hence, further study of sediments in fjords with limited or no aquaculture presence are necessary to determine these organism's biological success in the Norwegian coastal marine environment, and their potential for being either native or introduced.

Effects dictating spatial distribution of Paramoeba

Patterns in the spatial distribution of *Paramoeba* sp. detected by our molecular methods were in agreement with the limited spatial analysis conducted previously at AGD-infected aquaculture sites in Tasmania¹⁰⁵. While the effect of distance was not considered statistically significant within our analysis, the decreasing mean relative read abundance of *Paramoeba* sp. observed in water samples by distance from the farm agrees with the findings of Douglas-Helders *et al.* (2002), who used immune-dot botting to measure amoeba density up to 1.1km from an actively AGD-infected farm. Presumably, no previous studies have addressed the

relationship between horizontal distance and *Paramoeba* sp. density in sediments below salmon farms, but the trends visible within the present analysis further support that distance from salmon farms is a factor in pathogenic amoeba abundance.

Direction and bottom depth, the remaining two spatial variables measured for sediment collections, both significantly influenced abundance of *Paramoeba* sp. MOTUs present in sediments. These variables were tightly coupled as the western transect points towards shallow water near shore, while the northern transect ventures off to deeper waters towards the center of the fjord. The significant differences in relative read abundance between categories of each variable give a clear picture of *Paramoeba* presence on the benthos. While these results could be due to specific habitat preferences by these unidentified *Paramoeba* haplotypes, it is also likely that they are due to differential sedimentation of dead or cyst-like *Paramoeba* cells from the farm above. Sedimentation rates in fjords are often highly variable, due to strong current velocities in the upper water column and rapid fluctuations in bottom depth which result from the unique geologic history of these landforms¹⁰⁶. Current velocities measured for the Skogshamn salmon farm (Figure 5) show that the strongest surface currents experienced at the net pens generally move in the NE direction, supporting the minor trends observed in directional distribution of *Paramoeba* sp. reads in surface water samples. As depth increases, the measured currents change to more SW driven and eventually become less unidirectional, with the final measurement on the bottom indicating two nearly equal currents which fluctuate with tidal changes. These two current measurements from 15m depth and the bottom indicate that *Paramoeba* which have become detached from their salmonid hosts or other structures could sink towards the bottom while being pushed either southwest towards the collection points on the western transect, or eastwards towards the two shallower points on the eastern transect. The points deeper than 300m, which contained significantly fewer *Paramoeba* sp. reads and were generally not within the path of these currents, were also less likely to undergo sedimentation from particulates near the surface because of the separation of water masses that typically takes place in the deep regions of fjords¹⁰⁶. As seen in Figure 4, the sill in Solbergfjord comes to approximately 300m, making this depth the most likely zone for differentiation of water masses in the central parts of the fjord where the bottom is deeper. Our findings on benthic and surface water *Paramoeba* relative read abundance corroborate the theory that once established in a large aggregation of farmed salmon, these amoebas are potentially dispersed from that focal point to other farms or populations of wild fish by water currents and sedimentation until the salmon are harvested and net pen structures are removed from the area^{55,107}.

Temporal clues to potential vectors and environmental tolerances

Assuming that these resilient organisms may accumulate and survive for an extended period on the sediment after sinking from the surface waters above^{66,108}, the benthos may provide a reservoir for AGD-associated pathogens in areas such as Skogshamn, where aquaculture facilities are frequently placed. The chronology of *Paramoeba* MOTUs occurrence at Skogshamn provide further evidence for speculation on this matter, as the first detection of the pathogens was not during the first sampling event, but shortly after installation of large mooring anchors in the sediment around the farm and the deployment of net pen structures took place. Such activities have the potential to resuspend a tremendous amount of sediment from soft bottom sediments into the warmer surface waters and reintroduce pathogens such as *Paramoeba* to structures like nets or mooring lines where they may attach and feed or reproduce. Wind driven mixing is also a seasonal source of sediment resuspension but is not be suspected to have occurred at Skogshamn until the onset of storms in early October. The presence of MOTU SKOC_000217484 in the water samples during only this peak in abundance encourage postulations that perhaps the nets, lines, feeding tubes, and other installed equipment could have been a vector for this MOTU from the facility where the equipment was previously used. Further, MOTUs SKOC_001226214 and SKOC_002580467, which were each detected in the water once, and then only in the sediments from the Feb/March collection also push notions that MOTUs may have been introduced to this small geographic area by the sharing of culturing equipment between sites.

September 12th was the only date during which a statistically significant distance effect was observed on the relative read abundance of *Paramoeba* in the water samples. Moreover, confidence levels calculated between distance levels produced significance only between the sample replicates taken from within the nets and the samples taken from each of the other distances (0-25, 0-100, 0-200, and 0-500m), not between any of the other distances, providing further evidence that the empty nets may have been the focal point of initial *Paramoeba* sp. dispersion. According to national regulation, industry-wide sanitation measures are only required when an ISA or similar disease class outbreak has been detected⁹⁴. This leaves the possibility that normal biofouling removal practices at farms may not be sufficient for cleaning the nets of pathogens like *Paramoeba* between uses.

The long absence of *Paramoeba* reads from the water samples between September 12th and November 5th runs contrary to the hypothesis that warmer water temperatures in early fall would

drive rapid increases in *Paramoeba* sp. abundance. Previous studies of infection rates in laboratory inoculated salmon have shown that *Paramoeba* sp. fresh isolate and *P. perurans* clonal cultures have infected fish within 7 to 38 days^{32,109}, though these tests occurred with amoeba cells at higher densities and water temperatures than are experienced here. Cultures of *P. perurans* were grown by Collins *et al.* in 2019 at temperatures similar to the Sept-Nov Skogshamn measurements, and resulted in cell counts doubling every 40-60hrs¹¹⁰. Taking amoeba reproductive rates into account, an 8-week period between potential resuspension in early September and the next detection event in November are not so farfetched for a pathogen present at only trace abundances. An additional consideration must also be made for the timeline of salmon placement in the pens, which began during the final week of September, leaving approximately 5 weeks for amoeba population growth if that was reliant on parasitizing the salmon smolts. From the second detection date in early November through the peak abundance of *Paramoeba* sp. at ~6-8 weeks of salmon presence, *Paramoeba* MOTUs were consistently detected until the final sampling date. Read intensities dropped to minimum values during the subsequent samplings in January and February, but successive positive detections throughout the winter months suggest that a reservoir for these pathogens may have been established, and that we encountered trace abundances of dead or motile forms of these *Paramoeba* in our water samples.

The maintenance of *Paramoeba* sp. populations during the winter months further challenges established environmental tolerances of these organisms, which were considered to only infect salmon farms during periods with temperatures of 10°C or higher³⁰. Sustained surface temperature minimums in this fjord approach 2.5°C during the coldest weeks of the year and measurements taken for the final sampling dates during week six of 2020 were 3.2°C. Evidence presented by Buchwald *et al.* (2015) suggests *P. invadens*, which was formerly considered an exotic parasite in Nova Scotia, has potentially evolved enough cold tolerance to allow for overwintering in subtidal waters with temperatures as low as 2°C¹¹¹. If similar physiological changes are occurring in AGD associated *Paramoeba* sp. as a result of increase selection pressures on populations now present in the Arctic, further threat of fish mortalities should be expected by the aquaculture industry throughout Norway and the other high latitude salmon producing regions.

Diversity and genetic origins of Arctic Paramoebae

The results validate metabarcoding as an effective method for detecting specific pathogens in trace volumes and demonstrate its capability to distinguish clearly between genetic variants within some genera. Further consideration of this inter-MOTU genetic divergence also allows us to delve more precisely into the genetic origins of these amoebas. The phylogenetic analysis conducted herein was limited by the absence of reference sequences for multiple members of the *Paramoeba* genus but was successful in concluding that none of the individuals we detected were the primary agent of AGD, *P. perurans*, or two other species associated with the disease, *P. aparasomata* or *P. branchiphila*. Significant bootstrap replicates of the analysis were in agreement that SKOC 000217484 was the most closely related MOTU to any of the reference strains. Most specifically, it shared 95% genetic similarity with a *P. pemaquidensis* strain (Asc: MK990593) derived from farmed coho salmon, *Oncorhynchus kisutch* (Walbaum 1792), gills at farms in US waters of the Pacific Ocean⁶⁸. Because it was more closely related to this strain than the two *P. pemaquidensis* reference sequences were to each other, we can infer that its initial assignment to this species by *Ecotag* during bioinformatic steps was correct, and that a lower percent match threshold may be appropriate for species level MOTU assignment of *P. pemaquidensis*. While the exact origins of this haplotype cannot be deduced by the limited analysis conducted here, it is important to consider if this may be an exotic strain of *P. pemaquidensis* that has been accidentally introduced to Norway and the North Atlantic. Alternatively, these strains could perhaps have always had a circumpolar or near-global distribution.

The 11 other Skogshamn *Paramoeba* MOTUs can be considered most closely related to the second strain of *P. pemaquidensis*, which was cultured from environmental samples taken in Wales⁶⁸. The replicate agreement for the ancestral nodes between those 11 MOTUs and the *P. pemaquidensis* clade was not strong enough to compare each's individual relationship with the Welsh strain due to the diversity within them and the short size of the Leray fragment used for comparison. Furthermore, the lack of reference sequences for *P. aestuarina*, *atlantica*, *eilhardi*, *karteshi*, and *invadens*, leaves them all as possible true identifications of these MOTUs.

Between individual MOTUs, SKOC 002439258 and SKOC 002558372 were concluded to be significantly more similar to each other than any of the other MOTUs and were both detected at considerably greater abundance in the samples collected after fish farming had been underway. Unlike SKOC 002439258, SKOC 002558372 was not found in the water column during

sampling events, but it may have gone undetected due to technical shortcomings discussed above, and the already trace abundances of this MOTU encountered in the sediments. Two remaining clades exhibited bootstrap replicate frequency of note, with 48 and 44 percent of the analyses producing them. The first is made up of four MOTUs found only in the sediments, which do not display strong trends of increasing or decreasing abundance over time individually, or as a whole. The second contains the 1st and 4th most abundant MOTUs in terms of raw reads, which both were detected in the water and displayed a substantial increase in the sediments over the study period. One of these, SKOC 002580467, was also the least divergent from the calculated ancestral node for all *Paramoeba* spp., with a difference of less than 10 nucleotides.

Based upon the percent divergence within and between species of the 3 genera represented, the 11 MOTUs identified to *Paramoeba* sp. have potential to represent new strains of existing species, or perhaps even new species within the genus. These MOTUs diverge on between 10 and 35 SNPs (3-11.2%) from the furthest *Paramoeba* ancestral node, compared to four SNPs (1%) between strains of *P. perurans*, 27 (8.6%) between the two divergent strains of *P. pemaquidensis*, and up to 50 (16%) between some of the more distant *Paramoeba* sp. and *Korotnevella* sp.. These relationships display a high level of intraspecific genetic variability for this gene fragment within the species available for analysis, further supporting use of a lower threshold for species level assignment of *Paramoebidae* taxa in future COI Leray metabarcoding. A more detailed analysis using larger gene fragments and a more comprehensive reference database would be a prerequisite to establishing precise lineages for these new *Paramoeba* haplotypes, but the findings expose a yet unknown diversity of the genus in Arctic latitudes.

Collectively, these results provide support for the final hypothesized use of Skogshamn COI metabarcoding data to reveal inter and intraspecific genetic diversity of *Paramoeba* sp. within this Arctic fjord and their known relatives worldwide.

CONCLUSION

The findings of this study do not support the primary hypothesis that *P. perurans* would be detected in trace abundances at this aquaculture facility, but do confirm the remainder of the hypotheses, and further encourage the development of eDNA metabarcoding methods for pathogen monitoring. Detections of *Paramoeba* sp. most closely related to *P. pemaquidensis*

were shown to be more abundant in sediments than in the water column when analyzed based upon relative read abundance across samples. Relative abundances of these MOTUs in both water and sediment samples were also determined to be linked to spatial traits of sampling points, physical environmental factors measured at the sample site, and temporal changes in aquaculture facility use. Finally, inter and intraspecies genetic variability for 12 novel haplotypes of genus *Paramoeba* detected in this study were assessed, and both genetic and potential ecological relationships were inferred.

PERSPECTIVES

Further assessment of pathogens using markers with greater reference availability such as 18S rRNA could be a useful alternate perspective for identifying strains and detecting them at greater read abundances. This thesis project was of limited duration but was successful in confirming and expanding upon established theories of *Paramoeba* interactions with salmon farms and their surrounding environment. Continued study of this genus and other specific pathogens, such as sea louse and toxic algae over multi-year time series could greatly expand the collective knowledge of these organisms' behavior in the natural environment, and one's modified by aquaculture practice and other anthropogenic impacts. Therefore, promotion of epidemiological research from this perspective is vital, and should be pursued down countless avenues of the aquatic and health sciences.

Discovery of such a broad diversity of unrecorded pathogen haplotypes around a Northern Norwegian aquaculture system warrants further investigation by Leray-XT metabarcoding. The method is commonly used to establish and monitor biodiversity indices for ecosystem health assessment and could be adapted during that purpose to study more specifically the biodiversity of natural pathogens in ecosystems with limited or no influence of aquaculture. Results of such an analysis would be crucial to determining which haplotypes may or may not be reliant on salmon or other densely gathered fish for proliferation.

Specific to *Paramoeba*, this work outlines several trends in relative abundance that should be further monitored, and potential sampling strategies to further elucidate relationships between the potential parasites and their hosts. In particular, frequent collection of gill and other tissue swab specimens from farmed fish at the site and any wild fish aggregating nearby, as well as biofouling samples prior to and after net cleanings. Consistent and higher replicate sampling of the water from within cages should also be prioritized to calculate pathogen density in the

immediate vicinity of the salmon. Lastly, culturing and sequencing *Paramoeba* isolates from any gill, sediment, and/or water samples would be beneficial in characterizing the genetic and geographic origins of one or many of these pathogens, and lead to a greater understanding of the evolution of these organisms as they become more common in areas with extensive ocean aquaculture around the world.

Acknowledgements

This research project and thesis manuscript has been the greatest test of my academic capabilities thus far in life, and I cannot do enough to thank the institutions that made it possible or the people who have helped me along the way.

First and foremost, I am thankful of the Kingdom of Norway and the University of Tromsø for welcoming me from half the world away. Providing a state funded education to a foreign citizen like myself is a kind gesture of global cooperation, and one I will attempt to replicate over the course of my career and life in general.

My greatest appreciation also goes out to my supervisors Kim and Owen for all their guidance over this tumultuous year. Through difficulties with sampling and extraction timelines, to a worldwide shutdown for COVID 19, you have provided the support needed to complete the immediate task at hand, as well my long-term goals with this project.

Enough cannot be said in thanks of the assistance provided by my fellow RGG members. From helping with field and lab work to giving feedback on my data analysis and manuscript drafts; even providing a social safety net via zoom during home isolation in the pandemic. Many thanks to Gledis, Marta, Julie, Shripathi, Enrique, Anna, Griffin, Kamilla, and Carlos.

I would like to acknowledge the efforts of the faculty administration, particularly Camilla, Siv, and Tone, for their aid in navigating university policies and paperwork of all kinds.

I cannot begin to express my gratitude for my family, Mom, Erik, and Leah, for your support in this whole endeavor. From the moment I chose Tromsø as the place to apply for grad school to the last week before my thesis submission, you have been standing by to help me achieve my goals and make the decisions that will define my future.

The completion of this thesis would not have been possible without my partner Mia, a bastion of love and encouragement throughout the process. Your sharing (or at least listening to) my overwhelming excitement and bitter frustrations for this subject has helped me stay the course and overcome many of the hurdles I have encountered.

And finally, in memory of my late father, who inspired me to pursue science in its many forms and explore the world along the way. Thank you for all the wisdom you shared.

References

1. FAO Fisheries & Aquaculture - National Aquaculture Sector Overview - Norway. http://www.fao.org/fishery/countrysector/naso_norway/en#tcN701A4. Accessed March 31, 2020.
2. Schwach V. Havet, Fisken Og Vitenskapen. Fra Fiskeriundersøkelser Til Havforskningsinstitutt 1860-2000. Bergen: Havforskningsinstituttet; 2000.
3. Aarset B. Norwegian salmon-farming industry in transition: Dislocation of decision control. *Ocean Coast Manag.* 1998;38(3):187-206. doi:10.1016/S0964-5691(97)00037-9
4. Ørstavik F. Knowledge spillovers, innovation and cluster formation: the case of Norwegian aquaculture. In: Karlsson C, Flensburg P, Hörte S-Å, eds. *Knowledge Spillovers and Knowledge Management.* ; 2004:204-226.
5. Aslesen HW. The innovation system of Norwegian aquacultured salmonids. 2007.
6. Langdon JS. Smoltification physiology in the culture of salmonids. In: Muir JF, Roberts RJ, eds. *Recent Advances in Aquaculture. Vol 2.* Springer US; 1985:79-118. doi:10.1007/978-1-4684-8736-7_3
7. Gjedrem T, Aulstad D. Selection experiments with salmon I. Differences in resistance to vibrio disease of salmon parr (*Salmo salar*). *Aquaculture.* 1974;3(1):51-59. doi:10.1016/0044-8486(74)90098-2
8. Gjedrem T, GjØen HM, Gjerde B. Genetic origin of Norwegian farmed Atlantic salmon. *Aquaculture.* 1991;98(1-3):41-50. doi:10.1016/0044-8486(91)90369-I
9. Robb DHF, Crampton VO. On-farm feeding and feed management: perspectives from the fish feed industry. In: Hasan MR, New MB, eds. *On-Farm Feeding and Feed Management in Aquaculture.* FAO Fisheries and Aquaculture Technical Paper No. 583. Rome: FAO; 2013:489-518.
10. Figueroa J, Cárcamo J, Yañez A, et al. Addressing viral and bacterial threats to salmon farming in Chile: historical contexts and perspectives for management and control. *Rev Aquac.* 2019;11(2):299-324. doi:10.1111/raq.12333
11. Ministry of Fisheries and Coastal Affairs. *Strategy for a Competitive Norwegian Aquaculture Industry.* Norwegian Ministry of Fisheries and Coastal Affairs; 2008.
12. Benfey TJ. Producing sterile and single-sex populations of fish for aquaculture. In: *New Technologies in Aquaculture: Improving Production Efficiency, Quality and Environmental Management.* Elsevier Inc.; 2009:143-164. doi:10.1533/9781845696474.1.143
13. Norway's market share shrinking. *Salmon Business.* <https://salmonbusiness.com/norways-market-share-shrinking/>. Published 2018.
14. Asche F, Bjørndal T. *The Economics of Salmon Aquaculture.* 2nd ed. Wiley-Blackwell; 2011.

15. Kent M, Poppe T. Infectious diseases of coldwater fish in marine and brackish water. In: Woo P, Bruno D, Lim S, eds. Diseases and Disorders of Finfish in Cage Culture. New York: CABI; 2002:61-106.
16. Hektoen H, Berge JA, Hormazabal V, Yndestad M. Persistence of antibacterial agents in marine sediments. *Aquaculture*. 1995;133(3-4):175-184. doi:10.1016/0044-8486(94)00310-K
17. Grave K, Lingaas E, Bangen M, Rønning M. Surveillance of the overall consumption of antibacterial drugs in humans, domestic animals and farmed fish in Norway in 1992 and 1996. *J Antimicrob Chemother*. 1999;43:243-252.
18. Rodgers CJ, Furones MD. Antimicrobial agents in aquaculture: practice, needs and issues. In: Rogers C, Basurco B, eds. The Use of Veterinary Drugs and Vaccines in Mediterranean Aquaculture. Zaragoza : CIHEAM. ; 2009:41-59. doi:10.13140/2.1.4697.0560
19. Hjeltnes B, Bang-Jensen B, Bornø G, Haukaas A, Walde C, eds. The Health Situation in Norwegian Aquaculture 2017. *Nor Vet Inst Rep Ser*. 2018.
20. Novak CW, Lewis DL, Collicutt B, Verkaik K, Barker DE. Investigations on the role of the salmon louse, *Lepeophtheirus salmonis*, as a vector in the transmission of *Aeromonas salmonicida* subsp. *salmonicida*. *J Fish Dis*. 2016;39(10):1165-1178. doi:10.1111/jfd.12449
21. Nylund A, Hovland T, Hodneland K, Nilsen F, Levik P. Mechanisms for transmission of infectious salmon anaemia (ISA). *Dis Aquat Organ*. 1994;19:95-100.
22. Overton K, Dempster T, Oppedal F, Kristiansen TS, Gismervik K, Stien LH. Salmon lice treatments and salmon mortality in Norwegian aquaculture: a review. *Rev Aquac*. 2019;11(4):1398-1417. doi:10.1111/raq.12299
23. Wagner GN, Fast MD, Johnson SC. Physiology and immunology of *Lepeophtheirus salmonis* infections of salmonids. *Trends Parasitol*. 2008;24(4):176-183. doi:10.1016/j.pt.2007.12.010
24. Rozas-serri M. Gill diseases in marine salmon aquaculture with an emphasis on amoebic gill disease. *CAB Rev*. 2019;14(June). doi:10.1079/PAVSNNR201914032
25. Downes JK, Yatabe T, Marcos-Lopez M, et al. Investigation of co-infections with pathogens associated with gill disease in Atlantic salmon during an amoebic gill disease outbreak. *J Fish Dis*. 2018;41(8):1217-1227. doi:10.1111/jfd.12814
26. León-Munõz J, Urbina MA, Garreaud R, Iriarte JL. Hydroclimatic conditions trigger record harmful algal bloom in western Patagonia (summer 2016). *Sci Rep*. 2018;8(1):1-10. doi:10.1038/s41598-018-19461-4
27. Karlsen KM, Robertsen R, Hersoug B. Kartlegging Av Hendelsesforløp Og Beredskap under Giftalgeangrepet Våren 2019 - Astafjorden, Ofotfjorden, Vestfjorden Og Tysfjorden. Nofima; 2019.

28. Munday BL, Foster CK, Roubal FR, Lester RJG. Paramoebic gill infection and associated pathology of Atlantic salmon, *Salmo salar*, and rainbow trout, *Salmo gairdneri*, in Tasmania. In: Perkins FO, Cheng TC, eds. Pathology in Marine Science. San Diego, California: Academic Press, Inc; 1990:215-222.
29. Roubal FR, Lester RJG, Foster CK. Studies on cultured and gill - attached *Paramoeba* sp. and the cytopathology of paramoebic gill disease in Atlantic salmon, *Salmo salar* L., from Tasmania. J Fish Dis. 1989;12(5):481-492. doi:10.1111/j.1365-2761.1989.tb00559.x
30. Kent M, Sawyer T, Hedrick R. *Paramoeba pemaquidensis* infestation of the gills of coho salmon *Oncorhynchus kisutch* reared in sea water. Dis Aquat Organ. 1988;5:163-169. doi:10.3354/dao005163
31. Young ND, Crosbie PBB, Adams MB, Nowak BF, Morrison RN. *Neoparamoeba perurans* n. sp., an agent of amoebic gill disease of Atlantic salmon (*Salmo salar*). Int J Parasitol. 2007;37(13):1469-1481. doi:10.1016/j.ijpara.2007.04.018
32. Crosbie PBB, Bridle AR, Cadoret K, Nowak BF. In vitro cultured *Neoparamoeba perurans* causes amoebic gill disease in Atlantic salmon and fulfils Koch's postulates. Int J Parasitol. 2012;42(5):511-515. doi:10.1016/j.ijpara.2012.04.002
33. Mitchell SO, Baxter EJ, Holland C, Rodger HD. Development of a novel histopathological gill scoring protocol for assessment of gill health during a longitudinal study in marine-farmed Atlantic salmon (*Salmo salar*). Aquac Int. 2012;20(5):813-825. doi:10.1007/s10499-012-9504-x
34. Adams MB, Ellard K, Nowak BF. Gross pathology and its relationship with histopathology of amoebic gill disease (AGD) in farmed Atlantic salmon, *Salmo salar* L. J Fish Dis. 2004;27(3):151-161. doi:10.1111/j.1365-2761.2004.00526.x
35. Clark A, Nowak BF. Field investigations of amoebic gill disease in Atlantic salmon, *Salmo salar* L., in Tasmania. J Fish Dis. 1999;22(6):433-443. doi:10.1046/j.1365-2761.1999.00175.x
36. Rogerson A, Laybourn-Parry J. The abundance of marine naked amoebae in the water column of the Clyde estuary. Estuar Coast Shelf Sci. 1992;34(2):187-196. doi:10.1016/S0272-7714(05)80104-0
37. Crosbie PBB, Nowak BF, Carson J. Notes Isolation of *Neoparamoeba pemaquidensis* Page, 1987 from marine and estuarine sediments in Tasmania. Bull Eur Ass Fish Pathol. 2003;23(5):241.
38. Wiik-Nielsen J, Mo TA, Kolstad H, Mohammad SN, Hytterød S, Powell MD. Morphological diversity of *Paramoeba perurans* trophozoites and their interaction with Atlantic salmon, *Salmo salar* L., gills. J Fish Dis. 2016;39(9):1113-1123. doi:10.1111/jfd.12444
39. Valdenegro-Vega VA, Crosbie P, Bridle A, Leef M, Wilson R, Nowak BF. Differentially expressed proteins in gill and skin mucus of Atlantic salmon (*Salmo salar*) affected by amoebic gill disease. Fish Shellfish Immunol. 2014;40(1):69-77. doi:10.1016/j.fsi.2014.06.025

40. Douglas-Helders M, Saksida S, Raverty S, Nowak BF. Temperature as a risk factor for outbreaks of amoebic gill disease in farmed Atlantic salmon (*Salmo salar*). Bull Eur Ass Fish Pathol. 2001;21(3).
41. Stagg HEB, Hall M, Wallace IS, Pert CC, Perez SG, Collins C. Detection of *Paramoeba perurans* in Scottish marine wild fish populations. Bull Eur Ass Fish Pathol. 2015;35(6):217-226.
42. Rozas M, Bohle H, Grothusen H, Bustos P, Program P. Epidemiology of amoebic gill disease (AGD) in Chilean salmon industry between 2007 and 2010. Bull Eur Ass Fish Pathol. 2007;32(5):181.
43. Rodger HD, Mcardle JF. An outbreak of amoebic gill disease in Ireland. Vet Rec. 1996;139(October):348-349. doi:10.1136/vr.139.14.348
44. Steinum T, Kvellestad A, Rønneberg LB, et al. First cases of amoebic gill disease (AGD) in Norwegian seawater farmed Atlantic salmon, *Salmo salar* L., and phylogeny of the causative amoeba using 18S cDNA sequences. J Fish Dis. 2008;31:205-214.
45. Oldham T, Rodger H, Nowak BF. Incidence and distribution of amoebic gill disease (AGD) — An epidemiological review. Aquaculture. 2016;457:35-42. doi:10.1016/j.aquaculture.2016.02.013
46. Munday BL, Zilberg D, Findlay V. Gill disease of marine fish caused by infection with *Neoparamoeba pemaquidensis*. J Fish Dis. 2001;24(9):497-507. doi:10.1046/j.1365-2761.2001.00329.x
47. Feehan CJ, Johnson-Mackinnon J, Scheibling RE, Lauzon-Guay JS, Simpson AGB. Validating the identity of *Paramoeba invadens*, the causative agent of recurrent mass mortality of sea urchins in Nova Scotia, Canada. Dis Aquat Organ. 2013;103(3):209-227. doi:10.3354/dao02577
48. Mullen TE, Russell S, Tucker MT, et al. Paramoebiasis associated with mass mortality of American lobster *Homarus americanus* in Long Island Sound, USA. J Aquat Anim Health. 2004;16(1):29-38. doi:10.1577/H02-045.1
49. Rodger H. Amoebic gill disease (AGD) in farmed salmon (*Salmo salar*) in Europe. Fish Vet J. 2013;14(November):16-26.
50. Robledo D, Matika O, Hamilton A, Houston RD. Genome-wide association and genomic selection for resistance to amoebic gill disease in Atlantic salmon. G3 Genes, Genomes, Genet. 2018;8(4):1195-1203. doi:10.1534/g3.118.200075
51. Shinn AP, Pratoomyot J, Bron JE, Paladini G, Brooker EE, Brooker AJ. Economic costs of protistan and metazoan parasites to global mariculture. Parasitology. 2015;142(1):196-270. doi:10.1017/S0031182014001437
52. Dyková I, Nowak BF, Crosbie PBB, et al. *Neoparamoeba branchiphila* n. sp., and related species of the genus *Neoparamoeba* Page, 1987: Morphological and molecular characterization of selected strains. J Fish Dis. 2005;28(1):49-64. doi:10.1111/j.1365-2761.2004.00600.x

53. Bridle AR, Crosbie PBB, Cadoret K, Nowak BF. Rapid detection and quantification of *Neoparamoeba perurans* in the marine environment. *Aquaculture*. 2010;309(1-4):56-61. doi:10.1016/j.aquaculture.2010.09.018
54. Wright DW, Nowak B, Oppedal F, Bridle A, Dempster T. Free-living *Neoparamoeba perurans* depth distribution is mostly uniform in salmon cages, but reshaped by stratification and potentially extreme fish crowding. *Aquac Environ Interact*. 2017;9:269-279. doi:10.3354/AEI00233
55. Hellebø A, Stene A, Aspehaug V. PCR survey for *Paramoeba perurans* in fauna, environmental samples and fish associated with marine farming sites for Atlantic salmon (*Salmo salar* L.). *J Fish Dis*. 2017;40(5):661-670. doi:10.1111/jfd.12546
56. Choudhuri S. Fundamentals of molecular evolution. In: *Bioinformatics for Beginners*. Elsevier; 2014:27-53. doi:10.1016/b978-0-12-410471-6.00002-5
57. Unda F. Introduction to Phylogenetics. *Sci Creat Q*. May 2005.
58. Pace NR, Sapp J, Goldenfeld N. Phylogeny and beyond: Scientific, historical, and conceptual significance of the first tree of life. *Proc Natl Acad Sci*. 2012;109(4):1011-1018. doi:10.1073/pnas.1109716109
59. Wang X, King Jordan I, Mayer LW. A phylogenetic perspective on molecular epidemiology. In: *Molecular Medical Microbiology: Second Edition*. Vol 1-3. Elsevier Ltd; 2014:517-536. doi:10.1016/B978-0-12-397169-2.00029-9
60. Choudhuri S. Phylogenetic analysis. In: *Bioinformatics for Beginners*. Academic Press; 2014:209-218. doi:10.1016/B978-0-12-410471-6.00009-8
61. Borriess R, Rueckert C, Blom J, Bezuidt O, Reva O, Klenk HP. Whole genome sequence comparisons in taxonomy. In: *Methods in Microbiology*. Vol 38. Academic Press; 2011:409-436. doi:10.1016/B978-0-12-387730-7.00018-8
62. Woese CR, Fox GE. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc Natl Acad Sci*. 1977;74(11):5088-5090. doi:10.1073/pnas.74.11.5088
63. Woese CR. Bacterial evolution. *Microbiol Rev*. 1987;51(2):221-271. doi:10.1128/mmbr.51.2.221-271.1987
64. Dyková I, Nowak B, Pecková H, Fiala I, Crosbie P, Dvůráková H. Phylogeny of *Neoparamoeba* strains isolated from marine fish and invertebrates as inferred from SSU rDNA sequences. *Dis Aquat Organ*. 2007;74:57-65.
65. English CJ, Swords F, Downes JK, et al. Prevalence of six amoeba species colonising the gills of farmed Atlantic salmon with amoebic gill disease (AGD) using qPCR. *Aquac Environ Interact*. 2019;11:405-415. doi:10.3354/aei00325
66. Elliott NG, Wong F, Carson J. Detection and abundance of *Paramoeba* species in the environment. Hobart, Tasmania; 2001.

67. Nassonova E, Smirnov A, Fahrni J, Pawlowski J. Barcoding amoebae: comparison of SSU, ITS and COI genes as tools for molecular identification of naked lobose amoebae. *Protist*. 2010;161(1):102-115. doi:10.1016/j.protis.2009.07.003
68. Hansen H, Botwright NA, Cook MT, et al. Genetic diversity among geographically distant isolates of *Neoparamoeba perurans*. *Dis Aquat Organ*. 2019;137(2):81-87. doi:10.3354/dao03433
69. Volkova E, Kudryavtsev A. Description of *Neoparamoeba longipodia* n. sp. and a new strain of *Neoparamoeba aestuarina* (Page, 1970) from deep-sea habitats. *Eur J Protistol*. 2017;61:107-121. doi:10.1016/j.ejop.2017.09.006
70. English CJ, Tyml T, Botwright NA, et al. A diversity of amoebae colonise the gills of farmed Atlantic salmon (*Salmo salar*) with amoebic gill disease (AGD). *Eur J Protistol*. 2019;67:27-45. doi:10.1016/j.ejop.2018.10.003
71. Hebert PDN, Cywinska A, Ball SL, Dewaard JR. Biological identifications through DNA barcodes. *Proc R Soc Lond*. 2003;270(B):313-321. doi:10.1098/rspb.2002.2218
72. Turon X, Antich A, Palacín C, Præbel K, Wangensteen OS. From metabarcoding to metaphylogeography: separating the wheat from the chaff. *Ecol Appl*. 2020;30(2). doi:10.1002/eap.2036
73. Collins RA, Wangensteen OS, O’Gorman EJ, Mariani S, Sims DW, Genner MJ. Persistence of environmental DNA in marine systems. *Commun Biol*. 2018;1(1):1-12. doi:10.1038/s42003-018-0192-6
74. Peters L, Spatharis S, Dario MA, et al. Environmental DNA: A new low-cost monitoring tool for pathogens in salmonid aquaculture. *Front Microbiol*. 2018;9(DEC):1-10. doi:10.3389/fmicb.2018.03009
75. Ratnasingham S, Hebert PDN. BOLD: The Barcode of Life Data System: barcoding. *Mol Ecol Notes*. 2007;7(3):355-364. doi:10.1111/j.1471-8286.2007.01678.x
76. Pennisi E. DNA barcodes jump-start search for new species. *Science* (80-). 2019;364(6444):920-921. doi:10.1126/science.364.6444.920
77. Drummond AJ, Newcomb RD, Buckley TR, et al. Evaluating a multigene environmental DNA approach for biodiversity assessment. *Gigascience*. 2015;4(1):46. doi:10.1186/s13742-015-0086-1
78. Leray M, Yang JY, Meyer CP, et al. A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. *Front Zool*. 2013;10(1):34. doi:10.1186/1742-9994-10-34
79. Wangensteen OS, Palacín C, Guardiola M, Turon X. DNA metabarcoding of littoral hardbottom communities: High diversity and database gaps revealed by two molecular markers. *PeerJ*. 2018;6:e4705. doi:10.7717/peerj.4705

80. Geller J, Meyer C, Parker M, Hawk H. Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Mol Ecol Resour.* 2013;13(5):851-861. doi:10.1111/1755-0998.12138
81. Guardiola M, Uriz MJ, Taberlet P, Coissac E, Wangensteen OS, Turon X. Deep-sea, deep-sequencing: metabarcoding extracellular DNA from sediments of marine canyons. Duperron S, ed. *PLoS One.* 2015;10(10):e0139633. doi:10.1371/journal.pone.0139633
82. Turon X, Cebrian E, Palacín C. Under the canopy: Community-wide effects of invasive algae in Marine Protected Areas revealed by metabarcoding. *Mar Pollut Bull.* 2018;127(November 2017):54-66. doi:10.1016/j.marpolbul.2017.11.033
83. Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Res.* 2007;35(18):e120. doi:10.1093/nar/gkm541
84. Caporaso JG, Lauber CL, Walters WA, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci.* 2011;108(Supplement 1):4516-4522. doi:10.1073/PNAS.1000080107
85. D'Auriac MBA. COMplementary primer ASymmetric PCR (COMPAS-PCR) applied to the identification of *salmo salar*, *salmo trutta* and their hybrids. *PLoS One.* 2016;11(10):1-15. doi:10.1371/journal.pone.0165468
86. Siegenthaler A, Wangensteen OS, Benvenuto C, Campos J, Mariani S. DNA metabarcoding unveils multiscale trophic variation in a widespread coastal opportunist. *Mol Ecol.* 2019;28(2):232-249. doi:10.1111/mec.14886
87. Coissac E, Riaz T, Puillandre N. Bioinformatic challenges for DNA metabarcoding of plants and animals. *Mol Ecol.* 2012;21(8):1834-1847. doi:10.1111/j.1365-294X.2012.05550.x
88. Pawlowski J, Kelly-Quinn M, Altermatt F, et al. The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. *Sci Total Environ.* 2018;637-638:1295-1310. doi:10.1016/j.scitotenv.2018.05.002
89. Strickler KM, Fremier AK, Goldberg CS. Quantifying effects of UV-B, temperature, and pH on eDNA degradation in aquatic microcosms. *Biol Conserv.* 2015;183:85-92. doi:10.1016/j.biocon.2014.11.038
90. Hansen BK, Bekkevold D, Clausen LW, Nielsen EE. The sceptical optimist: challenges and perspectives for the application of environmental DNA in marine fisheries. *Fish Fish.* 2018;19(5):751-768. doi:10.1111/faf.12286
91. Jin J-J, Yu W-B, Yang J-B, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *bioRxiv.* 2019. doi:10.1101/256479
92. Boratyn GM, Thierry-Mieg J, Thierry-Mieg D, Busby B, Madden TL. Magic-BLAST, an accurate RNA-seq aligner for long and short reads. *BMC Bioinformatics.* 2019;20(1):405. doi:10.1186/s12859-019-2996-x

93. Directorate of Fisheries's aquaculture map. <https://www.yggdrasil.fiskeridir.no>. Published 2020.
94. Regulations on the Marketing of Aquaculture Animals and Products of Aquaculture Animals, Prevention and Control of Communicable Diseases in Aquatic Animals. Norway: <https://lovdata.no/dokument/SF/forskrift/2008-06-17-819>; 2008.
95. Interactive Fish Health Map. <https://www.barrentswatch.no/fiskehelse>. Published 2020.
96. Boyer F, Mercier C, Bonin A, Le Bras Y, Taberlet P, Coissac E. obitools: A unix-inspired software package for DNA metabarcoding. *Mol Ecol Resour.* 2016;16(1):176-182. doi:10.1111/1755-0998.12428
97. Ammon U von, Wood SA, Laroche O, et al. Combining morpho-taxonomy and metabarcoding enhances the detection of non-indigenous marine pests in biofouling communities. *Sci Rep.* 2018;8(1). doi:10.1038/s41598-018-34541-1
98. Deagle BE, Thomas AC, McInnes JC, et al. Counting with DNA in metabarcoding studies: how should we convert sequence reads to dietary data? *Mol Ecol.* 2018;28(2):391-406. doi:10.1111/303461
99. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol.* 2018;35(6):1547-1549. doi:10.1093/molbev/msy096
100. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution (N Y).* 1985;39(4):783-791. doi:10.1111/j.1558-5646.1985.tb00420.x
101. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol.* 1992;9(4):678-687. doi:10.1093/oxfordjournals.molbev.a040752
102. Mitchell JG, Yamazaki H, Seuront L, Wolk F, Li H. Phytoplankton patch patterns: Seascape anatomy in a turbulent ocean. *J Mar Syst.* 2008;69(3-4):247-253. doi:10.1016/j.jmarsys.2006.01.019
103. Breier RE, Lalescu CC, Waas D, Wilczek M, Mazza MG. Emergence of phytoplankton patchiness at small scales in mild turbulence. *Proc Natl Acad Sci.* 2018;115(48):12112-12117. doi:10.1073/pnas.1808711115
104. Jellett JF, Novitsky JA, Cantley JA, Scheibling RE. Non-occurrence of free-living *Paramoeba invadens* in water and sediments of Halifax Harbour, Nova Scotia, Canada. *Mar Ecol Prog Ser.* 1989;56:205-209.
105. Douglas-Helders GM, O'Brien DP, McCorkell BE, et al. Temporal and spatial distribution of paramoebae in the water column - a pilot study. *J Fish Dis.* 2003;26(4):231-240. doi:10.1046/j.1365-2761.2003.00452.x
106. Stigebrandt A. Hydrodynamics and circulation of fjords. In: *Encyclopedia of Earth Sciences Series.* Springer Netherlands; 2012:327-344. doi:10.1007/978-1-4020-4410-6_247

107. Uglem I, Dempster T, Bjørn P, Sanchez-Jerez P, Økland F. High connectivity of salmon farms revealed by aggregation, residence and repeated movements of wild fish among farms. *Mar Ecol Prog Ser.* 2009;384:251-260. doi:10.3354/meps08001
108. Enger O, Husevåg B, Goksøyr J. Presence of the fish pathogen *Vibrio salmonicida* in fish farm sediments. *Appl Environ Microbiol.* 1989;55(11):2815-2818. doi:10.1128/aem.55.11.2815-2818.1989
109. Tan CKF, Nowak BF, Hodson SL. Biofouling as a reservoir of *Neoparamoeba pemaquidensis* (Page, 1970), the causative agent of amoebic gill disease in Atlantic salmon. *Aquaculture.* 2002;210(1-4):49-58. doi:10.1016/S0044-8486(01)00858-4
110. Collins C, Hall M, Fordyce MJ, White P. Survival and growth in vitro of *Paramoeba perurans* populations cultured under different salinities and temperatures. *Protist.* 2019;170(2):153-167. doi:10.1016/j.protis.2018.11.003
111. Buchwald RT, Feehan CJ, Scheibling RE, Simpson AGB. Low temperature tolerance of a sea urchin pathogen: Implications for benthic community dynamics in a warming ocean. *J Exp Mar Bio Ecol.* 2015;469:1-9. doi:10.1016/j.jembe.2015.04.006

Appendix A – Reference Sequences

Table 3 Paramoeba reference sequences used for genome assembly, MOTU ID, and phylogenetic analysis. Full mitogenome(Mito), Shotgun assembly(SA), *EcoTag* database(ED), phylogenetic analysis(PA),

Species	Gene	Strain	Uses	Original Source	Ascension #
<i>Cunea profundata</i>	COI	DIVA3 602/13/16 h1	ED, PA	Estuarine sediment	KP862853
<i>Cunea russae</i>	COI	RC CCMAM 0458	PA	Estuarine sediment	MN317567
<i>Cunea thuwala</i>	COI	RS14G2.1	ED, PA	Estuarine sediment	KP862852
<i>Korotnevela heteracantha</i>	COI	Valamo 4 clone 218.10	PA	Freshwater sediment	KU659852
<i>Korotnevela stella</i>	COI	Valamo 1 1 clone 32.12	PA	Freshwater sediment	KU659838
<i>Korotnevela venosa</i>	COI	SPb clone 28.16	PA	Freshwater sediment	KU659860
<i>Paramoeba aestuarina</i>	18S	--No Recorded Name--	SA	Marine sediment	AY686574
		ATCC 50806	SA	Marine sediment	AY121852
<i>Paramoeba aparasomata</i>	Mito	RC CCMAM0454	SA,	Marine sediment	MK518072
	COI	RC CCMAM0454	SA, ED, PA	Marine sediment	MK518072
<i>Paramoeba atlantica</i>	18S	CCAP 1560/9	SA	Marine sediment	JN202436
<i>Paramoeba branchiphila</i>	COI	RP	SA, ED, PA	<i>Callinectes sapidus</i> (gut)	MK990594
	18S	RP	SA	<i>Callinectes sapidus</i> (gut)	EF675603
		O5 clone h1	SA	Intertidal sediment	KY465836
		5G5 clone h1	SA	Intertidal sediment	KY465845
		KPF3 clone h1	SA	Intertidal sediment	KY465831
<i>Paramoeba eilhardi</i>	COI	CCAP1560/2 clone 6.8	SA	Marine sediment	MK168799
		CCAP1560/2 clone 6.3	SA	Marine sediment	MK168798
		CCAP1560/2 clone 6.1	SA	Marine sediment	MK168797
	18S	106KRT	SA	Marine sediment	MH535952
		107-1HRT	SA	Marine sediment	MH535953
<i>Paramoeba invadens</i>	18S	A-11	SA	<i>Strongylocentrotus droebachiensis</i> (radial nerve)	KC790384
		S-5	SA	<i>Strongylocentrotus droebachiensis</i> (radial nerve)	KC790385
		SMB-60	SA	<i>Strongylocentrotus droebachiensis</i> (radial nerve)	MH934206
<i>Paramoeba karteshi</i>	COI	clone 4.9	SA	<i>Halisarca dujardini</i>	MK168802
		clone 4.3	SA	<i>Halisarca dujardini</i>	MK168801
		clone 4.1	SA	<i>Halisarca dujardini</i>	MK168800
	18S	clone 4.3	SA	<i>Halisarca dujardini</i>	MK168787

		clone 4.6	SA	<i>Halisarca dujardini</i>	MK168788
		clone 4.11	SA	<i>Halisarca dujardini</i>	MK168789
		clone 5.3	SA	<i>Halisarca dujardini</i>	MK168794
		clone 5.9	SA	<i>Halisarca dujardini</i>	MK168796
		clone 5.10	SA	<i>Halisarca dujardini</i>	MK168795
<i>Paramoeba longipodia</i>	COI	DIVA3 574/3	PA	Marine sediment	MF140256
<i>Paramoeba Pemaquidensis</i>	Mito	CCAP 1560/4	SA	Marine environment	KX611830
	COI	CCAP 1560/4	SA, ED, PA	Marine environment	MN025475
		ATCC 50172	SA, ED, PA	<i>Oncorhynchus kisutch</i> (water)	MK990593
	18S	GILL-NOR2	SA	<i>Salmo salar</i> (gills)	AY714354
		GILL-RICH3/I	SA	<i>Salmo salar</i> (gills)	EF675606
		NET-H2T3	SA	Biofouling	AY714350
		WT2708/I	SA	<i>Salmo salar</i> (gills)	EF675605
		TUN1/I	SA	<i>Thunnus maccoyii</i> (gills)	EF675607
		ST8V	SA	Estuarine sediment	AY714355
		SED-ST!	SA	Marine sediments	EU884479
		SED-CT1	SA	Estuarine sediments	EU884477
		SED-5A	SA	Marine sediments	AY714360.
		PA027	SA	<i>Salmo salar</i> (gills)	AF371967
	NP251002	SA	<i>Salmo salar</i> (gills)	AY714351	
<i>Paramoeba Perurans</i>	COI	haplotype 1	SA, ED, PA	<i>Salmo salar</i> (gills)	MK990592
		haplotype 2	SA, ED, PA	<i>Salmo salar</i> (gills)	MK990589
		haplotype 3	SA, ED, PA	<i>Salmo salar</i> (gills)	MK990591
		haplotype 4	SA	<i>Salmo salar</i> (gills)	MK990590
		haplotype 5	SA, ED, PA	<i>Salmo salar</i> (gills)	MN025478
		haplotype 6	SA, ED, PA	<i>Labrus bergylta</i> (gills)	MN025488
		haplotype 7	SA	<i>Salmo salar</i> (gills)	MK990584
		haplotype 8	SA, ED, PA	<i>Salmo salar</i> (gills)	MH535934
	18S	GD-D1/1/1	SA	<i>Salmo salar</i> (gills)	EF216902
		GD-D1/2	SA	<i>Salmo salar</i> (gills)	EF216899
		GD-D1/3	SA	<i>Salmo salar</i> (gills)	EF216900
		GD-D1/4	SA	<i>Salmo salar</i> (gills)	EF216901

Appendix B – Sampling Protocols

Skogshamn eDNA project sampling guidelines

Water sample collection:

Before navigating to a sampling point, clean (10% klor mixture) and rinse (saltwater) the Niskin bottle and attached line in the black storage bucket and make them ready for sampling. Then clean and rinse the surfaces of the boat where sampling will occur.

Once at the intended sampling point:

1. Spring load the bottle lids and rinse the Niskin bottle with saltwater before use.
2. Lower the Niskin bottle to the surface depth (~2m), send the weight to close the lids, then retrieve the bottle.
3. If the lids are not fully closed, reload and try again.
4. Label a sterile plastic bag with the sample location(A-Z) and collection depth(2, 15, or 200), remove the plastic seal, then open it and fill it by carefully pouring collected seawater directly from the top of the Niskin bottle.
5. Close the bag by squeezing out the remaining air, rolling the top 3-5 times, then twist the wire ties.
6. Bring the sealed bag to the plastic bin and place it upright to prevent leakage.
7. Reload the lid spring and slide weight back up the line in preparation for the next collection.
8. Navigate back to the same GPS point for the second depth collection if any drift has occurred or on to the next sampling site.
9. Repeat water sample collection protocol from step 1 for each remaining sample, then immediately return to the float to begin filtration process.

Static pump setup, water sample filtration, and storage protocols

Filtration station setup and pump assembly:

Bring container of water samples from boat to pumping area.

Set up folding table.

Spray with bottles and wipe with lint free paper towels to clean(klor mixture) and rinse(dH2O):

Table surface

Outside of tubing

Inside and outside of all 500mL and 1L plastic containers

Any other materials or surfaces used during the pumping process

Unpack pump and place on raised surface above table. Carefully mount pump head with Allen wrench and 4 screws provided. Plug in. Test for any noises or visible issues.

Unpack 3 cassettes and attach to pump with tubing running through them. Cassette tube size markers should be set to 17 on both sides and pressure applicator loose enough for cassettes to click into place with gentle downward pressure.

Fill a 1L container with klor mixture and another with dH₂O and place both on input side of pump.

Place 3rd 1L container for waste on output side. This will be used for Klor, dH₂O, and priming saltwater

Pumping procedures:

1. Clean surgical tubing of residual DNA by pumping klor mixture through all 3 tubes into the waste container. Run pump until tubes are completely full of klor mix, then shut off and let sit for 2min.
2. Remove input tubes from solution and pump remaining liquid from all 3 tubes.
3. Rinse chlorinated tubing by pumping dH₂O through all 3 tubes and into the waste container. Use ~200mL dH₂O, then repeat step 2
4. Select a sterile water bag and note location and depth labels. Place it in an empty filters box(for better stability) on the input side, carefully open it, and place all 3 pieces of tubing in it.
5. Prime all 3 tubes with saltwater by pumping ~200mL through them into the waste container, then shut off pump.
6. Change nitrile gloves.
7. One at a time, remove filters from sterile packaging and attach filters to output end of each tube. The exposed end of each filter should rest on the rim of an empty 500mL container, pointing downwards so all filtered water is collected in the container without immersing and contaminating the filter.
8. Begin pumping. Keep at 60rpm to prevent excess pressure from popping the filters off of the tubes.
9. When filtered water levels near the 500mL mark, remove the input end of each tube from the water sample bag. Continue pumping air until the filter becomes visibly dry.
10. Once all filters appear dry, turn off pump, remove filters, and place each into a separate pre-labeled falcon tube. Each should be labeled with location letter and depth number from the water sample bag. Also include a replicate number(r1, r2, or r3) for each of the 3 filters.
11. Place all 3 falcon tubes into a small ziplock bag labeled with the sampling letter, depth number, and date of collection.
12. Place that bag into a large ziplock bag with all water sample filters from that same sampling date. Close the larger bag between samples to reduce contamination from other DNA sources.
13. Repeat from step 1 for each water sample.

After completing filtration of all samples, double bag the samples and place in the eDNA only freezer onsite as soon as possible.

Freezer guidelines:

The freezer should only be used for eDNA water filters. Avoid opening it unless adding samples for storage or removing bags for immediate transportation to UiT.

Benthic sediment collection:

Benthic sediment collection should always take place after water collection. The grab and other sampling equipment should remain on the float until water sampling is completed and sterile water bags have been removed from the vessel.

Before navigating to a sampling point, clean and rinse the sediment grabber, attached line, and sediment trays in the black storage bucket and make them ready for sampling. Clean and rinse the surfaces of the boat where sampling will occur.

Once at the intended sampling point

1. Rinse the sediment grabber and trays thoroughly with saltwater before each use.
2. Mount the sediment grabber on the winch.
3. Quickly lower the sediment grabber to the ocean bottom through the winch pullie, but maintain contact with the line to prevent tangles.
4. When the grabber hits bottom and tension on the line is relieved, pull upwards to activate closing mechanism. Note: consider redropping the grab 2 or 3 times from ~5m off the bottom to ensure closing mechanism is activated.
5. Once closed, raise grabber with winch at agreed upon maximum speed. Coil line from the winch directly into the black bucket to prevent tangles on next deployment. Note: watch for the 15m indicator(black tape) on the line as you coil
6. Carefully bring the full sediment grabber onto the vessel by hand and place in a clean tray.
7. Open the doors on top of the grabber to check for adequate sediment, then pour off any excess water.
8. Take 3 replicate samples in pre-labeled falcon tubes. Note: by drilling a hole in the end of the falcon tubes, air can be released, and sediment easily enters the tube as a core sample.
9. Clean and dry surfaces of the falcon tubes, close air hole with plastic wrap, then place falcon tubes into a small ziplock bag pre-labeled with the sampling location and date.
10. Place bag into a secure storage container.
11. Repeat from step 1 at the next sample station until all sediment samples are collected. Once sampling is complete, place all sample bags into a large ziplock bag labeled with the sampling date and place that bag into the specimen freezer on arrival at the shoreside facility.

Before leaving the float:

Clean(klor mixture) and rinse(freshwater) all sampling equipment in black buckets and pack away neatly in the designated area.

Spray/wipe clean(klor mixture) and rinse(dH₂O) all pumping and filtration equipment before placing it in eDNA storage boxes(aluminum) and packing them away neatly in designated area.

Ensure doors between workshop and living areas are closed, interior lights are off, and rolling door is completely shut and bolted.

Appendix C – Sediment Extraction Protocol

19/02/2019
Paulina Urban

DNeasy PowerSoil Kit

Safety Information:

1. **Solution 5** contains ethanol – it is flammable
2. **PowerBead tubes** and **Solution 4** contain guanidine salts, which should not be combined with **bleach**. If liquid containing these buffers is spilt, clean with a suitable laboratory detergent and water. If the spilt liquid contains potentially infectious agents: clean area with laboratory detergent and water, and then with 1% (v/v) sodium hypochlorite.

Extensions: Bead beating options

This kit does not require homogenization using a high velocity bead beater. If homogenization other than provided by a vortex is needed use PowerLyzer®24 Homogenizer (110/220V) (cat.no 13155) in combination.

Equipment needed:

- Microcentrifuge (10,000 x g)
- Pipettors (50 µl-500µl)
- Vortex & Vortex Adapter for 24 (1.5-2.0ml) tubes

Important Notes:

- Make sure the 2ml **PowerBead tubes** rotate freely in your centrifuge without rubbing
- Shake to mix **Solution C4** before use
- Perform all centrifugation steps at room temperature (15-25°C)
- If **Solution C1** has precipitated, heat at 60°C until precipitate dissolves
- 2 ml collection tubes are provided

Procedure:

1. Add 0.25 g of soil sample to the PowerBead tube provided. Gently vortex to mix

Note: After your sample has been loaded into the PowerBead tube, the next step is a homogenization and lysis procedure. The PowerBead Tube contains a buffer that will (a) help disperse the soil particles, (b) begin to dissolve humic acids and (c) protect nucleic acids from degradation Gentle vortexing mixes the components in the PowerBead tube and begins to disperse the sample in the buffer.

2. If **Solution C1** has precipitated, heat at 60 °C until precipitate dissolves. Add 60 µl of **Solution C1** to sample and invert several times or vortex briefly

Note: **Solution C1** may be added to the PowerBead tube before adding soil sample. **Solution 1** contains SDS and other disruption agents required for complete cell lysis. In addition to aiding cell lysis, SDS is an anionic detergent that breaks down fatty acids and lipids associated with the cell membrane of several organisms. If it gets cold, it will form a white precipitate in the bottle. Heating to 60 °C will dissolve the SDS but will not harm it or the other disruption agents. **Solution C1** can be used while it is still warm.

3. Secure PowerBead tubes horizontally using a Vortex Adapter for 24 (1.5-2 ml) tubes
4. Vortex at max. Speed for 2h preferably at 60°C

Note: If using the 24-place Vortex Adapter for more than 12 preps, increase the vortex time by 5-10 min. Vortexing is critical for complete homogenization and cell lysis. Cells are lysed by a combination of chemical agents from steps 1-4 and mechanical shaking introduced at this step. By randomly shaking the beads in the presence of disruption agents, collision of the beads with microbial cells will cause the cells to break open. Use the Vortex Adapter will max. homogenization efficiency, which can lead to higher DNA yields. Avoid using tape, which can become loose and result in reduced homogenization efficiency, inconsistent results and reduced yields.

5. Centrifuge tubes at 10,000 x *g* for 1min
6. Transfer the supernatant to a clean 2 ml collection tube

Note: Expect between 400-500 µl of supernatant. Supernatant may still contain some soil particles

7. Add 250 µl of Solution C2 and vortex for 5 s. Incubate at 2-8 °C for 5 min

Note: You can skip the 5 min incubation. However, if you have already validated the DNeasy PowerSoil extractions with this incubation we recommend you retain the step. Solution C2 is patented Inhibitor Removal Technology (IRT). It contains a reagent that can precipitate non-DNA organic and inorganic material including humic substances, cell debris and proteins important to remove to increase DNA purity.

8. Centrifuge the tubes at 10,000 x *g* for 1 min
9. Avoiding the pellet, transfer up to 600 µl of supernatant to clean 2 ml collection tube

Note: The pellet at this point contains non-DNA organic and inorganic material including humic acid, cell debris and proteins. For best DNA yields and quality, avoid transferring any of the pellet.

10. Add 200 µl of Solution C3 and vortex briefly. Incubate at 2-8 °C for 5 min

Note: You can skip the 5 min incubation. However, if you have already validated the DNeasy PowerSoil extractions with this incubation we recommend you retain the step. Solution C3 is patented Inhibitor Removal Technology (IRT), see above.

11. Centrifuge the tubes at 10,000 x *g* for 1 min
12. Avoiding the pellet, transfer up to 700 µl of supernatant to a clean 2 ml collection tube (bigger tubes from the store).

Note: The pellet at this point contains non-DNA organic and inorganic material including humic acid, cell debris and proteins. For best DNA yields and quality, avoid transferring any of the pellet.

13. **Shake to mix** Solution C4 and add 1200 µl to the supernatant . Vortex for 5 s

Note: Solution C4 is a high-concentration salt solution. Since DNA binds tightly to silica at high salt concentrations, this will adjust the DNA solution salt concentrations to allow binding of DNA, but not non-DNA organic and inorganic material that may still be present at low levels, to the MB Spin Columns

14. Load 600 µl onto an MB Spin Column and centrifuge at 10,000 x *g* for 1 min. **Discard flow through**

Note: DNA is selectively bound to the silica membrane in the MB Spin Column.

15. Repeat step 14 twice, until all of the sample has been processed
16. Add 500 µl of Solution C5. Centrifuge again at 10,000 x *g* for 30 min

Note: Solution C5 is an ethanol based wash solution used to further clean the DNA that is bound to the silica filter membrane in the MB Spin Column. This wash solution removes residual salt, humic acid, and other contaminants while allowing the DNA to stay bound to silica membrane

17. **Discard the flow-through.** Centrifuge again at 10,000 x *g* for 1 min

Note: This flow-through is non-DNA organic and inorganic waste removed from the silica MB Spin Column membrane by the ethanol wash solution. The second spin removes residual Solution C5 that may interfere with many downstream DNA applications such as PCR or gel electrophoresis.

18. Carefully place the MB Spin Column into a clean 2ml collection tube. Avoid splashing any Solution C5 onto the column

19. Add 100 µl of Solution C6 to the center of the white filter membrane. Alternatively, you can use sterile DNA-free PCR-grade water for this step.

Note: Placing the Solution C6 (sterile elution buffer) in the center of the small white membrane will make sure the entire membrane is wet. This will result in a more efficient and complete release of the DNA from the silica MB Spin Column membrane. Solution C6 lacks salts, DNA gets released.

20. Centrifuge at room temperature at 10,000 x *g* for 30 min. **Discard the MB Spin Column.** The DNA is now ready for downstream applications.

Note: We recommend storing DNA frozen (-20 °C to -80 °C) as Solution C6 does not contain EDTA. To concentrate DNA see below (Troubleshooting Guide)

Troubleshooting Guide:

Soil processing:

A) Soil sample is high in water content: remove contents from PowerBead Tube (beads and solution) and transfer into another sterile microcentrifuge tube. Add soil sample to PowerBead Tube and centrifuge at room temperature for 30 s at 10,000 x *g*. Remove as much liquid as possible with a pipet tip. Add beads and bead solution back to PowerBead Tube, gently vortex to mix and resume protocol from Step 2.

DNA:

A) DNA does not amplify: Try diluting DNA template (although it should not be necessary with this isolation kit)

B) Concentrating eluted DNA: Add 4 µl of 5M NaCl and inverting 3-5 times to mix. Next, add 200 µl of 100% cold ethanol and invert 3-5 times to mix. Centrifuge at 10,000 x *g* for 5 min at room temperature. Decant all liquid. Remove residual ethanol in a speed vac., a desiccator or air dry. Resuspend precipitated DNA in sterile water or sterile 10mM Tris.

Alternative lysis:

A) Cells are difficult to lyse: After adding Solution C1, incubate at 70°C for 10 min. Resume protocol from step 3

B) Reducing of shearing of DNA: After adding Solution C1, vortex 3-4 seconds, then heat to 70°C for 5 min. Vortex 3-4 s. Heat another 5 min. Vortex 3-4 s. This alternative procedure will reduce shearing but may also reduce yield.

Appendix D – Water Extraction Protocol



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

EXTRACTION PROTOCOL FOR STERIVEX FILTERS

EDNA EXTRACTION BASED ON QIAGEN DNEASY BLOOD & TISSUE KIT

IMPORTANT NOTES

- Make sure that the incubator is set to 56°C before starting the work. The equipment you are going to use for the extraction protocol should always be cleaned.
- Always shake Eppendorf tubes out of the bag, don't put your hand inside of it. Discard any excess tubes.
- Only open the bags containing Eppendorf tubes, or other tubes inside the flowhood.
- Only use pipette tips with barriers/filters and only open the boxes inside the flowhood.
- Always follow the workflow or any precautions given for the eDNA clean lab working routines.
- Always discard tips/tubes/gloves if you have the slightest suspicion about contamination (e.g. if the tip touches the table before entering a tube or buffer bottle).
- Always work with at least one extraction blanks per extraction round (i.e. 24 samples). However, if you are working with 22 samples to extract, then to complete the number to 24, you work with two blanks.
- Always start with the lowest concentration i.e. air blanks and water blanks (if any) except extraction blanks, which should be treated as any regular sample.
- However, the extraction blanks
- If extracting samples from several species/locations, sterilize everything between samples.
- Do not touch the ends of the Sterivex filters or the inside of the tube caps with hands or tweezers.
- Always be careful when you open the Eppendorf tubes not to touch the inside of the cap. Hold them in your hand and flick them open with the tip of your thumb.
- MAKE SURE YOU HAVE ENOUGH TIPS! You will mainly use 1000µl tips but also stock up on 20µl and 200µl ones. You also need Eppendorf tubes (both 1.5ml and 2.0ml), 50ml falcon tubes. Always have enough of these things before you start working.

DAY 0:

1. Find filters in -80°C freezer and place them in the fridge in the lock at 4°C for gentle thawing. It takes approx. 1-2 hours but since the freezer is located in a “contaminated” area the preference is to take the samples out the day before. The day after you start your extractions by showing up in clean clothes and freshly showered.



DAY 1:

2. Follow the descriptions in the 'Clean Lab Routines' of how to enter the labs.
3. Clean the outside of the 50ml falcon tubes containing the filters with bleach. Alternatively, if your filters are in ziplock bags, clean the outside of the bag. Do not use ethanol if there is any labelling on the tube/bag.
4. To remove excess water inside the filters, place the inlet of the filter (narrow end) in a 1.5 ml Eppendorf tube and gently slide filter and tube into the 50 ml falcon tube that contained the filter (or in a new 50 ml tube if samples were stored in ziplock bags). If more than one filter is in the tube, label a new tube for the second filter. When done with filters from one species/station, clean everything again (forceps, gloves, working surface) with bleach, MilliQ water and ethanol, before proceeding to the next species/station.
5. Centrifuge the tubes at 1500 x g for 3 minutes to remove the remaining seawater from the filters.
6. Make extraction buffer solution for adding 2.5X the recommended volume = 500µl per filter.
 - Recommended volume is 20µl Proteinase K + 180µl Buffer ATL per sample:
 - 2.5 * 20µl ProK = 50µl
 - 2.5 * 180µl ATL = 450 ul
 - Total amount of extraction buffer per sample = 500µl
 - E.g. for 20 samples: 1000µl ProK, 9000µl ATL. First, pipette the 9ml with a sterile glass pipette into a clean 50ml or 15ml tube (if you have 24 or less samples the smaller tube is enough). Then pipette 1 ml of ProK into the same tube. Close with lid and invert solution, avoiding foaming.
7. Add 500µl of the extraction solution to each filter, starting with blanks, by pushing the 1000µl tip tight **into the outlet** end of the filter and gently aspirating the solution into the filter. Take care that all the solution goes into the filter. If the filter is clogged, then aspirate from the inlet end of the filter.
8. Cap the filters with sterile caps. Make sure that its **completely sealed**.
9. **MAKE SURE YOU LABEL ALL THE FILTERS CORRESPONDING TO THE TUBES**, by writing the label and the replicate letter (A, B, C etc.) on the filter and **cover with tape**.
10. Place the filters in rotator and fasten them with the elastic band.
11. When done with all filters, move the rotator to the incubator oven (56°C). Make sure that the rotator is moving at 6 rpm and not hitting the oven. Check the filters after a couple of hours and leave them overnight for the 2nd day of extractions. Minimum 8-12 hours incubation.
12. Note: Always use similar incubation time for all filters within a project. Note the time for when incubation in the incubator oven started.

DAY 2

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



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

19. "Measure" the approximate volume of 2-3 samples using a pipette with NEW tips for each sample. Round the mean volume to nearest 50µl.
20. Add an equal volume of the Buffer AL as the one determined above (7.) and ensure to mix it with the pipette immediately using new tips for each sample.
21. Add an equal volume of 100% EtOH as the one determined above (7.) and ensure to mix it with the pipette immediately using new tips for each sample.
22. Vortex and spin down the samples to make sure it is mixed and liquid from the cap is removed.
23. Place the spin columns in front of the samples in the rack.
24. Transfer 630µl of the sample into corresponding spin column. Be careful not to make any bubbles but at the same time try not to leave liquid in the tip because it is precious DNA.
25. Centrifuge the columns at 15.000 x g for 2 mins.
26. Discard the collection tube with the flow-through and transfer the spin column to a new collection tube. Make sure the flow-through has not spilled back to the column when you removed it from the centrifuge.
27. Transfer the rest of the sample to the corresponding spin column. If more than 630µl, three rounds of spinning are required.
28. Centrifuge the columns at 15.000 x g for 2 mins.
29. Discard the collection tube with the flow-through and transfer the spin column to a new collection tube. Make sure the flow-through has not spilled back to the column when you removed it from the centrifuge.
30. Add 500µl Buffer AW1 (check EtOH has been added to buffer) using new tips for each tube.
31. Centrifuge at 15000 x g for 2 mins.
32. Discard the collection tube with the flow-through and transfer the spin column to a new collection tube. Make sure the flow-through has not spilled back to the column when you removed it from the centrifuge.
33. Add 500µl Buffer AW2 and centrifuge for 4 mins at 20.000 x g.
34. While centrifuging, clean flowhood, pipettes, and pens with bleach, MilliQ and ethanol.
35. TAKE GREAT CARE that no flow-through is present on the sides of the spin columns. If so, spin the columns again in a new collection tube at 20.000 x g for 2 mins. Note what samples that have been centrifuged twice.
36. Transfer the spin-columns to the corresponding Eppendorf tubes. Make sure that the lid/tap of the spin column does not touch the cap of the Eppendorf tube to avoid contamination.
37. Add 75µl of Buffer AE to each spin columns. Make sure to add the buffer at the center of the membrane without touching the membrane. Incubate for 1 min, then spin the samples at 20.000 x g for 2 mins.
38. Discard the spin columns and transfer a 12µl aliquot of the extracted DNA from each sample to a PCR plate or PCR strips. It is very important the plate/strip is labeled properly with all necessary information (if using strips, use empty pipette tip boxes as racks). Wrap aliquots in two bags before temporary storage. Place the aliquot in the fridge at 4°C if you are certain it will be processed within the next 2-3 weeks or in the aliquot freezer if longer.
39. Store the rest of the DNA as stock in the freezer located in the extraction lab. Store the 1.5ml tubes in a cryobox that you have purchased at the store and brought with you. Make sure to label the box properly. Put the cryobox in two bags before storing it and **ONLY** thaw the stock if absolutely necessary.
40. Clean flowhood and all equipment according to the guidelines.



Appendix E – COI Leray-XT metabarcoding Protocol

Protocol for COI metabarcoding using Leray-XT primers and Metafast library preparation (PCR-free ligation procedure)

Owen S. Wangensteen. January 2018.

METABARCODING PRIMERS

We use the Leray-XT primer set (Wangensteen et al., 2018). This is a highly-degenerated primer pair able to amplify a 313 bp fragment of cytochrome *c* oxidase subunit I (COI) from a wide array of eukaryotic groups, including virtually all metazoans. The sequences (where “I” stands for deoxy-inosine) are:

Forward, **miCOIint-XT**: 5'-GGWACWRGWTGRACWITITAYCCYCC-3'
Reverse, **igHCO2198**: 5'-TAIACYTCIGGRTGICCRAARAAYCA-3'

DNA AMPLIFICATION

We use a simple 1-step PCR protocol to amplify the Leray fragment. The metabarcoding primers have an 8-base sample-tag attached (each tag with at least 3 differences out of 8 bases). Also, we add a variable number (2-4) of leading Ns, in order to increase sequence variability to improve Illumina sequencing. Each forward and reverse primer has the same sample-tag attached in both ends. E.g.:

Primer F1: NNaacaagccGGWACWRGWTGRACWITITAYCCYCC
Primer R1: NNNNaacaagccTAIACYTCIGGRTGICCRAARAAYCA

Primer F2: NNNggaatgagGGWACWRGWTGRACWITITAYCCYCC
Primer R2: NNNggaatgagTAIACYTCIGGRTGICCRAARAAYCA

Primer F3: NNNNaattgccgGGWACWRGWTGRACWITITAYCCYCC
Primer R3: NNaattgccgTAIACYTCIGGRTGICCRAARAAYCA

We have 96 such different pairs, so we can multiplex up to 96 samples in one library.

The PCR protocol uses AmpliTaq Gold 360 master mix (ThermoFisher)
<https://www.thermofisher.com/order/catalog/product/4398886>
and bovine serum albumin (BSA)
<https://www.thermofisher.com/order/catalog/product/B14?ICID=search-B14>

The PCR mix is as follows:

AmpliTaq Gold Master Mix	10.00	μl
BSA 20 μg/μl	0.16	μl
H2O	5.84	μl
Forward primer 5 μM	1	μl
Reverse primer 5 μM	1	μl
DNA Template	2	μl

Note that the primers cannot be added to the PCR master mix for aliquoting (as is common practice for preparing normal PCRs). They have to be added to every individual sample, since every sample will be amplified with a different version of the primer set.

The PCR programme is:

95°C	10 min	(needed for denaturing the blocking antibody of Taq polymerase)
94°C	1 min	
45°C	1 min	x 35 cycles
72°C	1 min	
72°C	5 min	(extension time)

LIBRARY POOLING AND CONCENTRATION

Once all samples are amplified, the success of amplifications may be checked by gel electrophoresis in 1% agarose. Note that the samples must be prepared in a clean room to avoid contaminations. They should never be opened in a common electrophoresis laboratory. We routinely use 2 µl of the PCR products for the electrophoresis. The rest (18 µl per sample, including the blank samples) will be pooled together in a single Eppendorf tube and this pool is then thoroughly homogenized by vortexing.

The pool is then purified using MinElute columns for removing DNA fragments below 70 bp. This step will also concentrate the amplified DNA around 10 times. <https://www.qiagen.com/qdm/aw/cup/pcr-purification/>

These MinElute columns have a maximum sample volume capacity of 130 µl per sample. So you will probably need to use 10 or 12 of such columns, depending on the total volume of your pool. Follow the protocol in the kit. In the final step, you can elute every column in 12-15 µl of elution buffer. Then pool all the eluates together and homogenize thoroughly by vortexing.

You can measure the DNA concentration in the final pool using a Qubit fluorimeter with the Broad-Range DNA quantification kit. You need a minimum concentration of 75 ng/µl in the final pool for a best performance of the next ligation step.

LIBRARY PREPARATION

For library preparation, we use a PCR-free ligation protocol, the NEXTflex PCR-Free DNA Sequencing Kit from BIO Scientific: <http://www.bioscientific.com/Next-Gen-Sequencing/Illumina-Library-Prep-Kits/NEXTflex-PCR-Free-DNA-Sequencing-Kit>

We use 3 µg of DNA (up to 40 µl of the previous pool) as starting material. The instructions for preparing a COI library are exactly the ones described in the kit manual: <http://www.bioscientific.com/Portals/0/Manuals/NGS/5142-01-NEXTflex-PCR-Free-DNA-Seq-Kit.pdf>

Note this protocol is valid for selecting fragment sizes of 300-400 bp, exactly the right size for the Leray fragment. If you want to use a different metabarcoding marker with a shorter fragment, then you need to change Step B of the protocol (size selection).

With this kit, you will get to ligate your amplicons to the Illumina adapters and a 6-base library tag. The basic kit includes just one such library-tag, which is enough for multiplexing 96 samples with our set of 96 sample-tags. If you wish to multiplex over 96 samples, you could use two or more library tags. For this, you would need to buy an extra box of BIO barcodes, which come in 6, 12, 24, 48 or 96 versions: <http://www.bioscientific.com/Next-Gen-Sequencing/Illumina-Adapters/DNA-Seq/NEXTflex-DNA-Barcodes>

You will need to use magnetic beads for some steps of this protocol. The original Agencourt AMPure XP beads are quite expensive, but they are most convenient. <http://uk.beckman.com/nucleic-acid-sample-prep/purification-clean-up/pcr-purification?geolocation=gb>

LIBRARY CHECKING

We usually analyse the final library using either an Agilent TapeStation or Bioanalyzer, in order to check that the ligation has gone well. If you don't have any of these analyzers available, then you could use just a gel electrophoresis to check the right migration of the fragment. Note that the library fragments are the result of a special Y-shaped adapter ligation and they will not be linear DNA. So they will migrate anomalously in all this analytical methods. The library peak will not appear at the expected size of ~ 510 bp, but it will produce a broad peak of ~ 800 bp. This strange migration behaviour is normal and won't interfere with the MiSeq sequencing.

LIBRARY QUANTIFICATION

In order to load the right concentration of the library in the MiSeq, it is essential to check the exact concentration of the library using a specific qPCR method. This method will use a specific probe for the Illumina adapter sequence, so it allows to quantify exactly which molarity of adapter you will be loading into the MiSeq, which is crucial for not overclustering the Illumina flow-cell.

For this purpose, we use the NEBNext Library Quant Kit from New England Biolabs: <https://www.neb.com/products/e7630-nebnext-library-quant-kit-for-illumina>

We usually analyse library dilutions of 1:5000, 1:10,000 and/or 1:50,000.

You will need to use a qPCR machine. In Salford, we use the Rotor-Gene Q from QIAGEN but, of course, any qPCR machine will work: <https://www.qiagen.com/us/search/rotor-gene-q/>

LIBRARY DILUTION AND MiSeq LOADING

The final target concentration for the MiSeq loading will depend if you want to use a v2 or v3 MiSeq sequencing kit. With a v2 kit, you can get up to 15 M reads, and you will use a sample with up to 10 pM DNA concentration. With a v3 kit you will get up to 25 M reads, and you will use a sample with up to 20 pM DNA concentration. We usually target at 9 pM for a v2 or 18 pM for a v3, so to prevent overclustering of the flow-cell.

We will prepare our sample including a 1% of PhiX library, which will be used as an internal sequencing control for calculating error rates per cycle. <https://www.illumina.com/products/by-type/sequencing-kits/cluster-gen-sequencing-reagents/phix-control-v3.html>

The protocol for <https://www.illumina.com/products/by-type/sequencing-kits/cluster-gen-sequencing-reagents/phix-control-v3.html>

- Prepare a mix of up to 10 µl of your library and PhiX-library mix (in the right molar proportions) and put it in the bottom of a 2-ml Eppendorf tube.
- Denature with the same volume of 0.2N NaOH during 5 min. During this time, you may vortex once and spin in a centrifuge for recovering the sample.
- Add HT1 hybridization buffer (included with your the MiSeq reagent kit) to a total volume of 2 ml and vortex thoroughly.
- Load 600 µl of this denatured sample into the the MiSeq for sequencing.

References:

- Wangenstein OS, Palacín C, Guardiola M, Turon X (2018) DNA metabarcoding of littoral hard-bottom communities: high diversity and database gaps revealed by two molecular markers. PeerJ 6, e4705. <https://peerj.com/articles/4705/>

Appendix F – Programming script for RStudio

```
# Set Local Working Directory
setwd("C:/Users/Nathan/Desktop/UiT Resources/Thesis/R Code")

# Load Libraries
library(dplyr)

library(ggplot2)

# Read data files
Paramoeba_reads_C01<-read.csv('C:/Users/Nathan/Desktop/UiT Resources/Thesis/Data/SKOC_Curated_LULU_only_Paramoeba.csv')
metadata<-read.csv('C:/Users/Nathan/Desktop/UiT Resources/Thesis/Data/SKOC_sample-metadata_adjusted.csv')
metadata <- data.frame(lapply(metadata, as.character), stringsAsFactors = FALSE)

# Samples
all_samples<-colnames(Paramoeba_reads_C01[17:ncol(Paramoeba_reads_C01)])

# Unify best ID to a single column
tax<-Paramoeba_reads_C01[1:12,8:13]

# Condense database by species_name
Para_reads_col<-rowsum.data.frame(Paramoeba_reads_C01[1:12,17:398],group = tax$species_name)

# Convert raw reads to relative abundance
relab<- t(as.data.frame(mapply('/',Para_reads_col,Paramoeba_reads_C01[15,17:398])*10000))
rownames(relab)<-rownames(Para_reads_col)

# Merge rel-abundance + metadata tables
rownames(metadata)<-metadata$aliquot_name
metadata<-t(metadata)

relab_meta<-rbind(relab,metadata)

relab_meta<-t(relab_meta)
relab_meta<-as.data.frame(relab_meta)

# Define blanks, feed, sediment, and water samples
sediment_samples <- relab_meta[16:87,]
sediment_samples$`Paramoeba sp.` <- as.numeric(as.character(sediment_samples$`Paramoeba sp.`))

water_samples <- relab_meta[88:382,]
water_samples$`Paramoeba sp.` <- as.numeric(as.character(water_samples$`Paramoeba sp.`))

water_samples_mOZ <- water_samples%>%
  filter( Distance != 0 & Distance != 1000) #all water samples except 0s &Zs

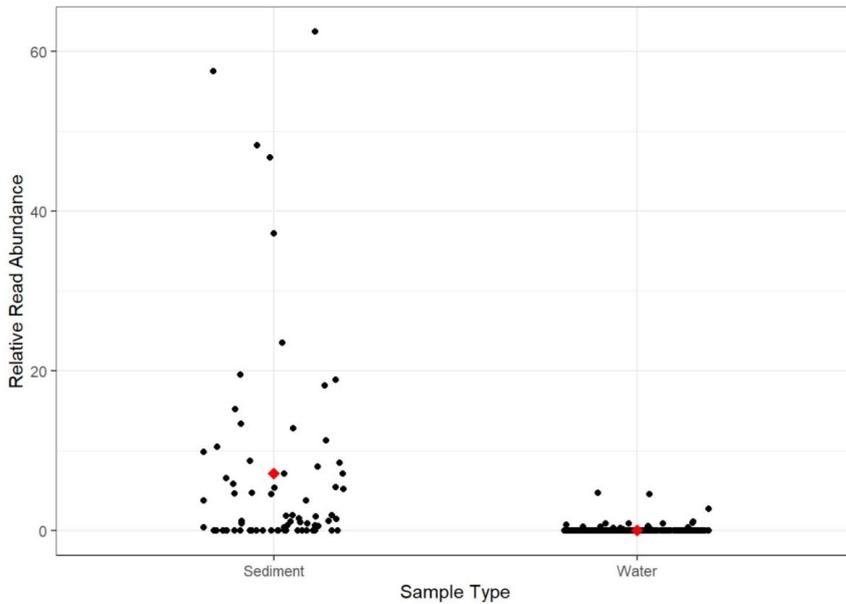
feed_samples<-relab_meta[12:15,]

##### SED vs WATER #####

relab_meta$`Paramoeba sp.` <- as.numeric(as.character(relab_meta$`Paramoeba sp.`))

relab_sed_water <- ggplot(relab_meta[16:382,], aes(x=sample_type, y= relab_meta[16:382,1] )) +
  geom_jitter(shape=16, position=position_jitter(0.2))+
  stat_summary(fun=mean, geom="point", color="red", shape=18, size=3 )+
  ylab('Relative Read Abundance')+ xlab('Sample Type')+ theme_bw()+
  theme(legend.position="none")

relab_sed_water
```



```
res.aov.sed_water <- aov(relab_meta[16:382,1]~relab_meta[16:382,3], data = relab_meta[16:382,])
summary(res.aov.sed_water)
```

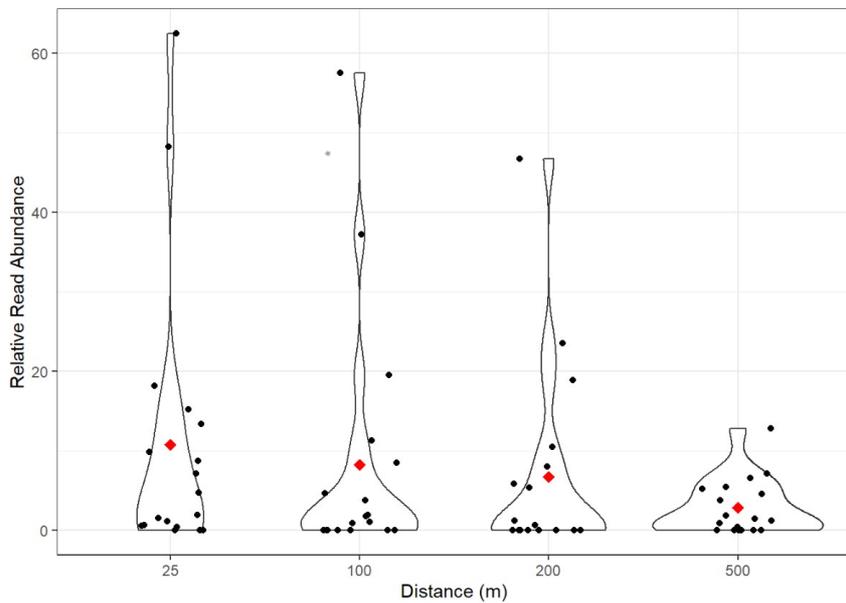
```
##           Df Sum Sq Mean Sq F value Pr(>F)
## relab_meta[16:382, 3] 1 2904 2903.7 83.89 <2e-16 ***
## Residuals          365 12633 34.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

SEDIMENTS

```
# Mean read abundance in sediment by distance
sediment_samples$Distance <- factor(sediment_samples$Distance, levels = c("25", "100", "200", "500"))

relab_sed_dist <- ggplot(sediment_samples, aes(x=Distance, y= sediment_samples[,1] )) +
  geom_violin(trim=TRUE)+
  geom_jitter(shape=16, position=position_jitter(0.2))+
  stat_summary(fun=mean, geom="point", color="red", shape=18, size=3 )+
  ylab('Relative Read Abundance')+ xlab('Distance (m)')+ theme_bw()+
  theme(legend.position="none")

relab_sed_dist
```



```
res.aov.sed_dist <- aov(sediment_samples[,1]~sediment_samples$Distance, data = sediment_samples)
summary(res.aov.sed_dist)
```

```
##
## sediment_samples$Distance 3 595 198.3 1.126 0.345
## Residuals 68 11982 176.2
```

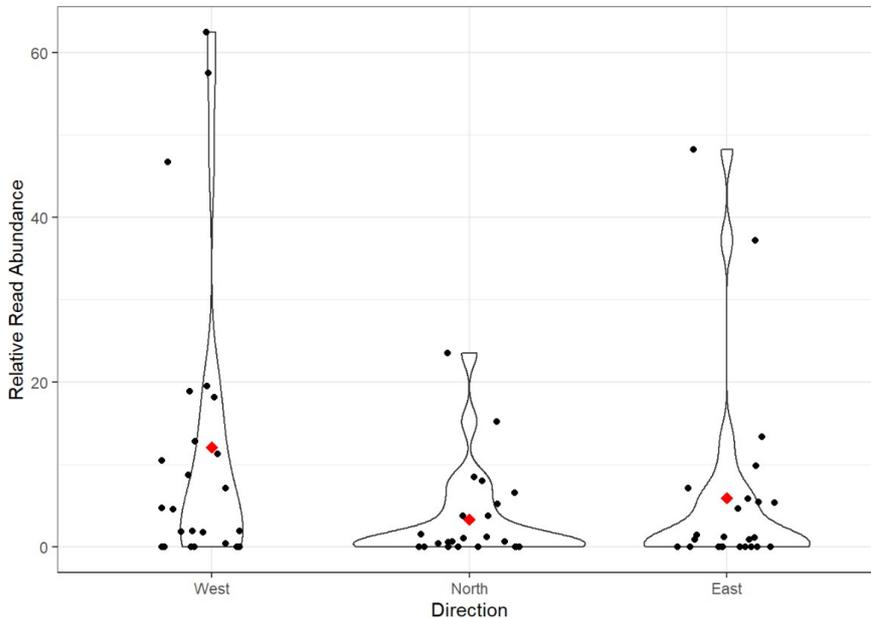
```
TukeyHSD(res.aov.sed_dist)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = sediment_samples[, 1] ~ sediment_samples$Distance, data = sediment_samples)
##
## $`sediment_samples$Distance`
## diff lwr upr p adj
## 100-25 -2.561151 -14.21462 9.092315 0.9381060
## 200-25 -4.089870 -15.74334 7.563596 0.7919411
## 500-25 -7.934359 -19.58782 3.719107 0.2854939
## 200-100 -1.528719 -13.18218 10.124747 0.9857047
## 500-100 -5.373208 -17.02667 6.280258 0.6199658
## 500-200 -3.844489 -15.49795 7.808977 0.8208614
```

```
# Mean read abundance in sediment by direction
sediment_samples$Direction <- factor(sediment_samples$Direction,levels = c("West","North","East"))

relab_sed_dir <- ggplot(sediment_samples, aes(x=Direction, y= sediment_samples[,1] )) +
  geom_violin(trim=TRUE)+
  geom_jitter(shape=16, position=position_jitter(0.2))+
  stat_summary(fun=mean, geom="point", color="red",shape=18, size=3 )+
  ylab('Relative Read Abundance')+ xlab('Direction')+ theme_bw()+
  theme(legend.position="none")

relab_sed_dir
```



```
res.aov.sed_dir <- aov(sediment_samples[,1]~sediment_samples$Direction, data = sediment_samples)
summary(res.aov.sed_dir)
```

```
##
## sediment_samples$Direction 2 974 487.1 2.897 0.0619 .
## Residuals 69 11603 168.2
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(res.aov.sed_dir)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = sediment_samples[, 1] ~ sediment_samples$Direction, data = sediment_samples)
##
## $`sediment_samples$Direction`
##          diff      lwr      upr    p adj
## North-West -8.767261 -17.733814  0.1992923 0.0566719
## East-West  -6.183593 -15.150146  2.7829598 0.2311663
## East-North  2.583667  -6.382886 11.5502204 0.7699908
```

```
# Mean read abundance in sediment by depth
sediment_samples$Sed_Dep_Binomial <- factor(sediment_samples$Sed_Dep_Binomial,levels = c("Shallow","Deep"))

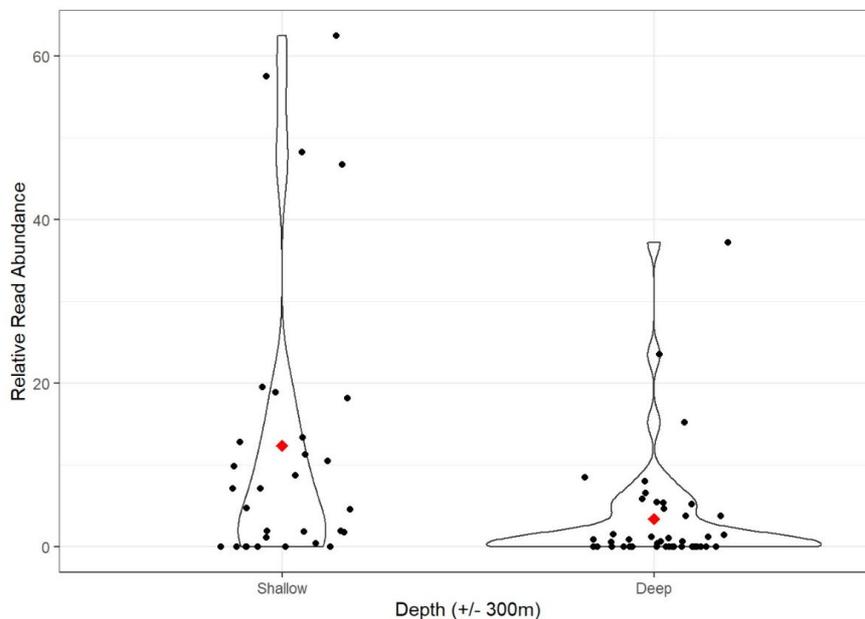
relab_sed_dep <- ggplot(sediment_samples, aes(x=Sed_Dep_Binomial, y= sediment_samples[,1] )) +
  geom_violin(trim=TRUE)+
  geom_jitter(shape=16, position=position_jitter(0.2))+
  stat_summary(fun=mean, geom="point",color="red",shape=18, size=3)+
  ylab('Relative Read Abundance')+ xlab('Depth (+/- 300m)')+ theme_bw()+
  theme(legend.position="none")

relab_sed_dep
```

```
# Mean read abundance in sediment by depth
sediment_samples$Sed_Dep_Binomial <- factor(sediment_samples$Sed_Dep_Binomial,levels = c("Shallow","Deep"))

relab_sed_dep <- ggplot(sediment_samples, aes(x=Sed_Dep_Binomial, y= sediment_samples[,1] )) +
  geom_violin(trim=TRUE)+
  geom_jitter(shape=16, position=position_jitter(0.2))+
  stat_summary(fun=mean, geom="point",color="red",shape=18, size=3)+
  ylab('Relative Read Abundance')+ xlab('Depth (+/- 300m)')+ theme_bw()+
  theme(legend.position="none")

relab_sed_dep
```



```
res.aov.sed_dep <- aov(sediment_samples[,1]~sediment_samples$Sed_Dep_Binomial, data = sediment_samples)
summary(res.aov.sed_dep)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## sediment_samples$Sed_Dep_Binomial  1  1399  1399.3   8.763 0.00419 **
## Residuals                        70  11178   159.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

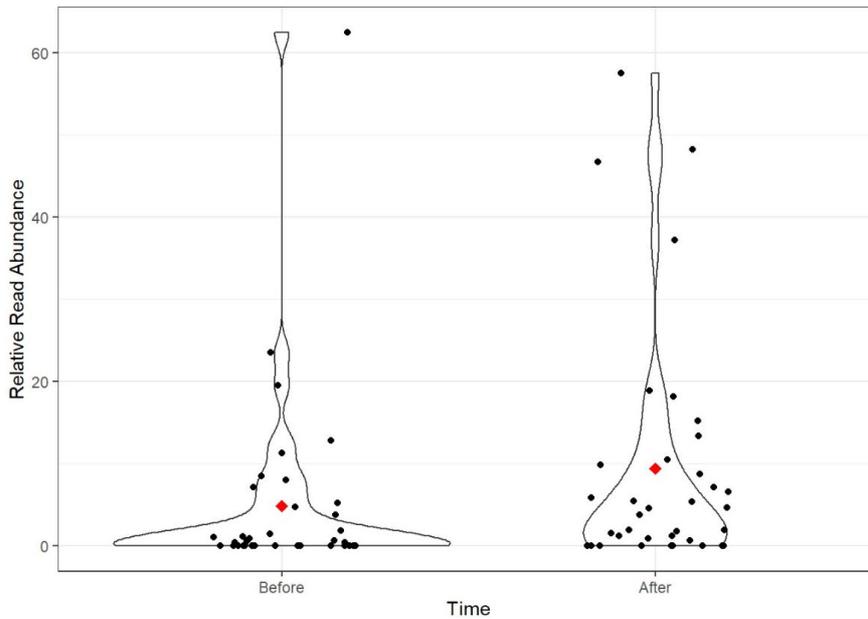
```

# Mean read abundance in sediment by time
sediment_samples$Time <- factor(sediment_samples$Time,levels = c("Before","After"))

relab_sed_time <- ggplot(sediment_samples, aes(x=Time, y= sediment_samples[,1] )) +
  geom_violin(trim=TRUE)+
  geom_jitter(shape=16, position=position_jitter(0.2))+
  stat_summary(fun=mean, geom="point",color="red",shape=18, size=3 )+
  ylab('Relative Read Abundance')+ xlab('Time')+ theme_bw()+
  theme(legend.position="none")

relab_sed_time

```



```

res.aov.sed_time <- aov(sediment_samples[,1]~sediment_samples$Time, data = sediment_samples)
summary(res.aov.sed_time)

```

```

##              Df Sum Sq Mean Sq F value Pr(>F)
## sediment_samples$Time  1    372   371.7    2.132  0.149
## Residuals              70  12205   174.4

```

```

##### WATER #####

```

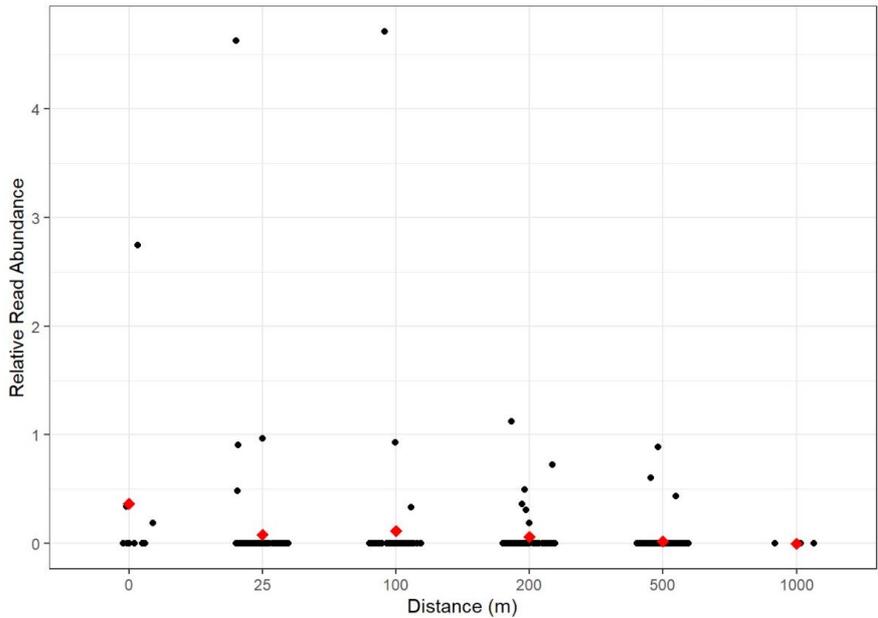
```

## Mean read abundance in water by distance
water_samples$Distance <- factor(water_samples$Distance,levels = c("0", "25", "100", "200", "500", "1000"))

relab_water_dist <- ggplot(water_samples, aes(x=Distance, y= water_samples$'Parameoba sp.' )) +
  geom_jitter(shape=16, position=position_jitter(0.2))+
  stat_summary(fun=mean, geom="point", color="red",shape=18, size=3 )+
  ylab('Relative Read Abundance')+ xlab('Distance (m)')+ theme_bw()+
  theme(legend.position="none")

relab_water_dist

```



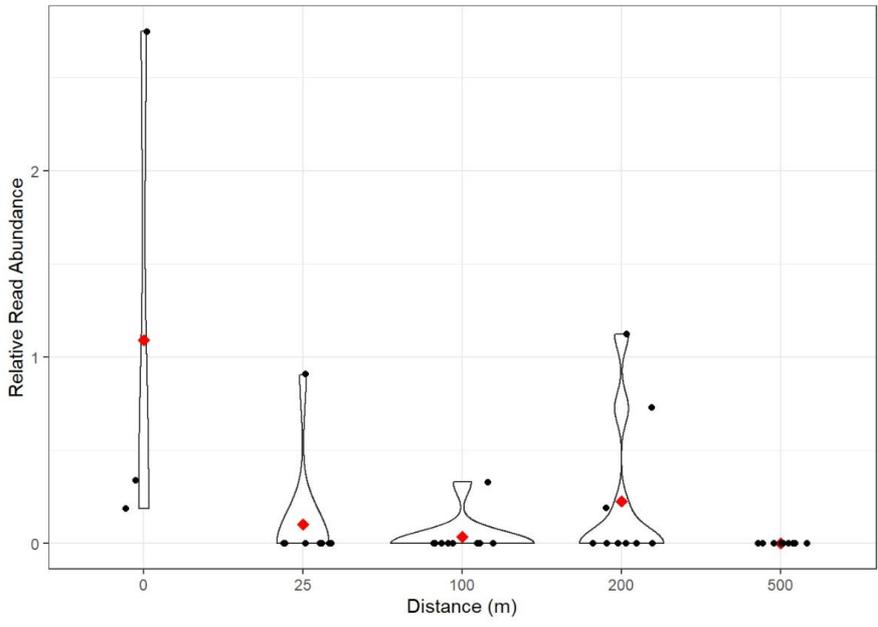
```
res.aov.water_dist <- aov(water_samples$'Paramoeba sp.'~water_samples$Distance, data = water_samples)
summary(res.aov.water_dist)
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## water_samples$Distance  5  1.13  0.2254  1.179  0.319
## Residuals                289 55.25  0.1912
```

```
#summary plot for dates 190912 - only date with distance as significant effect
water_samples_septpeak <- water_samples%>%
  filter( DATE_NUM == 190912)

relab_water_distseptpeak <- ggplot(water_samples_septpeak, aes(x=Distance, y= water_samples_septpeak[,1] )) +
  geom_violin(trim=TRUE)+
  stat_summary(fun=mean, geom="point", color="red",shape=18, size=3) +
  geom_jitter(shape=16, position=position_jitter(0.2))+
  ylab('Relative Read Abundance')+ xlab('Distance (m)')+ theme_bw()+
  theme(legend.position="none")

relab_water_distseptpeak
```



```
res.aov.water_distseptpeak <- aov(water_samples_septpeak[,1]~water_samples_septpeak$Distance, data = water_samples)
summary(res.aov.water_distseptpeak)
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## water_samples_septpeak$Distance  4  3.035  0.7588  4.081 0.00829 **
## Residuals                        34  6.321  0.1859
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(res.aov.water_distseptpeak)
```

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = water_samples_septpeak[, 1] ~ water_samples_septpeak$Distance, data = water_samples)
##
## $`water_samples_septpeak$Distance`
##      diff      lwr      upr      p adj
## 25-0    -0.99004111 -1.8177879 -0.1622944 0.0125057
## 100-0   -1.05408481 -1.8818315 -0.2263381 0.0069590
## 200-0   -0.86393483 -1.6916816 -0.0361881 0.0372382
## 500-0   -1.09066701 -1.9184137 -0.2629203 0.0049395
## 100-25  -0.06404369 -0.6493490  0.5212616 0.9977545
## 200-25  0.12610628 -0.4591991  0.7114116 0.9707981
## 500-25 -0.10062589 -0.6859312  0.4846794 0.9872982
## 200-100 0.19014997 -0.3951554  0.7754553 0.8810347
## 500-100 -0.03658220 -0.6218875  0.5487231 0.9997525
## 500-200 -0.22673217 -0.8120375  0.3585732 0.7972921
```

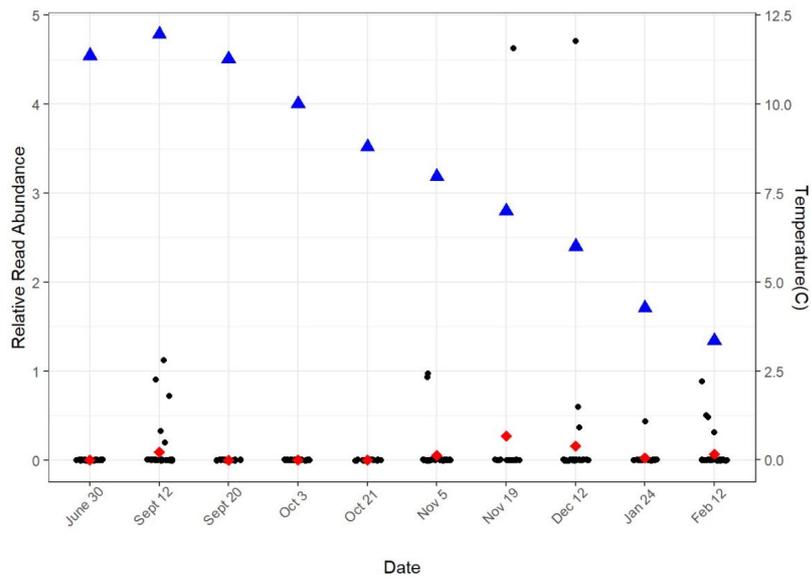
```
# Temperature by date
water_samples_moz$Temp<-as.numeric(as.character(water_samples_moz$Temp))
tempbydate <- aggregate.data.frame(list(water_samples_moz$Temp),
                                     by = list(water_samples_moz$DATE_NUM),
                                     FUN = function(x) c(mean(as.numeric(as.character(x))))))
colnames(tempbydate)<- c("DATE_NUM", "Temp")
dates_character<-c('June 30', 'Sept 12', 'Sept 20', 'Oct 3', 'Oct 21',
                  'Nov 5', 'Nov 19', 'Dec 12', 'Jan 24', 'Feb 12')

# Mean read abundance in water by date only using points A-N
relab_water_date <- ggplot(water_samples_moz, aes(x=DATE_NUM, y= water_samples_moz$'Paramoeba sp.' )) +
  geom_jitter(shape=16, position=position_jitter(0.2))+
  stat_summary(fun=mean, geom="point", color="red",shape=18, size=3)+
  geom_point(data = tempbydate,aes(x=DATE_NUM, y= Temp/2.5 ), shape=17, size=3, color = "blue")+
  ylab('Relative Read Abundance')+
  scale_x_discrete(name='Date',labels=dates_character)+
  theme_bw()+
  theme(axis.text.x = element_text(angle=45))+

  theme(legend.position="none")

relab_water_datetemp <- relab_water_date +
  scale_y_continuous(sec.axis = sec_axis(trans = ~ . * 2.5,
                                       name = "Temperature(C)"))

relab_water_datetemp
```



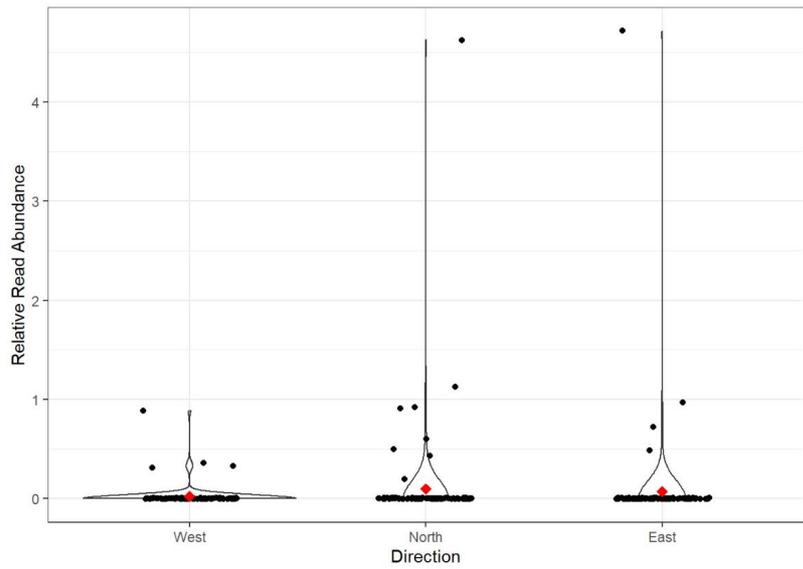
```
res.aov.water_date <- aov(water_samples_mOZ$'Paramoeba sp.'~water_samples_mOZ$DATE_NUM, data = water_samples_mOZ)
summary(res.aov.water_date)
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## water_samples_mOZ$DATE_NUM  9  1.55  0.1726  0.988  0.45
## Residuals                272 47.51  0.1747
```

```
# Mean read abundance in water by direction only using points A-N
water_samples_mOZ$Direction <- factor(water_samples_mOZ$Direction,levels = c("West","North","East"))

relab_water_dir <- ggplot(water_samples_mOZ, aes(x=Direction, y= water_samples_mOZ$'Paramoeba sp.' )) +
  geom_violin(trim=TRUE)+
  geom_jitter(shape=16, position=position_jitter(0.2))+
  stat_summary(fun=mean, geom="point", color="red",shape=18, size=3) +
  ylab('Relative Read Abundance')+ xlab('Direction')+ theme_bw()+
  theme(legend.position="none")

relab_water_dir
```



```
res.aov.water_dir <- aov(water_samples_mOZ$'Paramoeba sp.'~water_samples_mOZ$Direction, data = water_samples_mOZ)
summary(res.aov.water_dir)
```

```
##                Df Sum Sq Mean Sq F value Pr(>F)
## water_samples_mOZ$Direction  2  0.28  0.1407  0.805  0.448
## Residuals                279 48.78  0.1749
```


(This page was intentionally left blank)

(This page was intentionally left blank)

