UiT The Arctic University of Norway

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

# Whole genome sequencing reveals development of structured salmon lice (*Lepeophtheirus salmonis,* Krøyer, 1838) populations among aquaculture net pens through production

Nora Brække
Master's thesis in Aquamedicine, BIO-3955, May 2023

# Preface

The last year of the study, the Master´s thesis, has come to an end. During the spring of 2022 my supervisors and I agreed that I would be able to contribute to their LICE-MAP project. I felt an excitement to be able to write about my favorite parasite, salmon louse. After 5 years of studying Aquamedicine I have become very interested in how diseases and parasites affect the fish health of farmed fish. On the other hand, I wanted to challenge myself by choosing a slightly unfamiliar topic. I am satisfied with my choice as I have understood the importance of genetic studies, and how this is necessary information to be able to answer today´s problems that occur due to salmon lice and the treatments that follows.

I want to thank my supervisor, Kim, for your incredible dedication to genetic research and how your novel ideas become a reality, not to mention that I had the honor of taking part in this research group. I also want to thank my co-supervisor Roy for always being friendly and supportive of my work. A BIG thanks to Shripathi who has spent a lot of time helping me with the bioinformatics; without you, the results would not have made any sense. Simon was my lab supervisor, helping me with any problems that occurred during DNA extraction and other laboratory work. Thank you for all the fun sampling trips to Skogshamn! Thank you SalMar for letting us/me use your facilities to accomplish the research.

A big hug to my cool classmates Hanna, Solveig, Vilde, Mina, and Anne-Marja who made the "modul-life" a great life. Thank you for five amazing years living under the same roof Johanne. After a long-awaited summer holiday, a new chapter will start in Vesterålen with Endre, where salmon lice will continue to be a part of my daily life.

Thank you UiT and Norwegian College of Fishery Science, for five great years.

Tromsø, May 2023

Nora Brække

# Abstract

Salmon louse, (*Lepeophtheirus salmonis,* Krøyer, 1838), is an ectoparasite that causes multiple health and economic problems in the farming of Atlantic salmon (*Salmo salar* L.), but also threatens wild salmonid species. The industry is struggling to identify proper measures for limiting lice infections, and to identify effective delousing treatments. One of the major production and management challenges, but interesting biological observations, is that each treatment, as seen for medicinal or chemical treatment, becomes less effective due to the development of resistance in salmon lice. This causes reduced fish health, large financial losses for the industry, and, so far, unsolvable challenges for the management of aquaculture. The current management regime assumes that salmon lice along the Norwegian coast is a panmictic population, despite it is known that delousing treatments act as strong selective agents that create phenotypes who display adaptive traits towards resistance. However, there is no prior research addressing the population genomics underlying this observation on smaller geographical scales through time. The objective of this study was to investigate within farm and among net pens genomic population structure of salmon lice within an aquaculture facility located in northern Norway through time. The results of the present study, based on whole genome sequencing, and analyses of 66,769 SNPs, revealed that a production of Atlantic salmon can be viewed as a continuum of genomic differentiation, where salmon lice colonize the hosts at the onset of production to developing weak, but significant, population structure among net pens and through time. The application of the random forest machine learning identified 107 SNPs and 260 SNPs discriminatory for genomic differences among net pens and timepoints, respectively. These two sets of SNPs increased significant differentiation among net pens, and significantly supported the development of genetic population differentiation over time of production. Overall, these findings show that salmon lice within aquaculture net pens evolve into genetically differentiated populations during production time.

# Table of Contents

# Introduction

Aquaculture has globally expanded within the last decades, and Norway is currently the world's largest producer of Atlantic salmon (Iversen et al., 2020; Straume et al., 2020). In 2021, Norway sold over 1.6 million tons of farmed fish, an export industry that supplies fish products to a global market, where farmed Atlantic salmon represent the largest percentage (Fiskeridirektoratet, 2022). Norway is in a unique position to operate traditional salmon farming, due to its long coastline with deep fjords and natural advantages such as oxygen-rich water and cold temperatures (NOU, 2019). Even with good environmental conditions, open net pens have their disadvantages because of continuous seawater flow through the system, which enables free drift of pathogens (Conte, 2004). Open net pens are the most used form of farming, enabling lower costs and simpler construction. However, keeping the fish for the growth-out phase (12-24 months) in an open aquaculture system also mediates an environment where the fish is more susceptible to external factors like diseases, parasites, and other natural microorganisms that may threaten the fish welfare (Hastein et al., 2005). Additionally, an open aquaculture system also allows nutrients, parasites, and diseases from the high-intensity farming to spread into the natural ecosystem, with yet unknown consequences (Olaussen, 2018).

The salmon louse, *Lepeophtheirus salmonis*, is an ectoparasitic copepod that infects both farmed and wild fish (Johnsen et al., 2020), hereafter referred to as salmon lice. Salmon lice live naturally in the marine environment and feed on the skin, mucus, and blood of Atlantic salmon, and other salmonid species such as seatrout *(Salmo trutta* L.) (Johnsen et al., 2021; Sommerset et al., 2022). It represents one of the largest challenges in the farming industry in Norway today as it impairs the fish welfare, increases the production costs, and represents a major parameter to control towards the expansion of the industry (Tully & Whelan, 1993; Torrissen et al., 2013; Overton et al., 2019). The pathology occurs primarily due to the eating habits of the parasite. By destroying the skin of the fish, referred to as the first barrier, the fish will become more susceptible to other infectious such as bacteria, viruses, and fungi. Salmon lice can therefore cause multiple health problems to their host, and even mortality can be the outcome for heavily infested fish (Bjørn & Finstad, 1998; Dawson, 1998; Harðardóttir et al., 2021).

The life cycle of salmon lice consists of eight stages, where the development of the louse is temperature and salinity-dependent (Boxaspen, 2006; Hamre et al., 2019). Temperature and salinity have also been identified as major drivers of sea lice population dynamics (Heuch et al., 2002), and various life stages of salmon lice have different salinity tolerances, depending on the water temperature (Johnson & Albright, 1991; Bricknell et al., 2006). The nauplius and copepodite stages of salmon lice are nonfeeding planktonic larvae that drift with the water currents (Torrissen et al., 2013; Thorstad et al., 2015). In areas with strong currents, nauplii larvae and infective copepodites can be widely dispersed from their starting point, with up to 100 km or more (Asplin et al., 2011; Asplin et al., 2014). Once the copepodite attaches to the host, it molts into chalimus I and II, preadult I and II, and the adult stage. Adult females will carry offspring in a pair of long strings, where one louse has the capacity to produce 11 pairs of egg strings with up to 700 eggs in each egg string, throughout its lifetime (Heuch et al., 2000). The optimum temperature range for salmon lice is not fully understood, but the species requires temperatures of 4 °C or higher to complete its lifecycle successfully (Boxaspen & Næss, 2000).

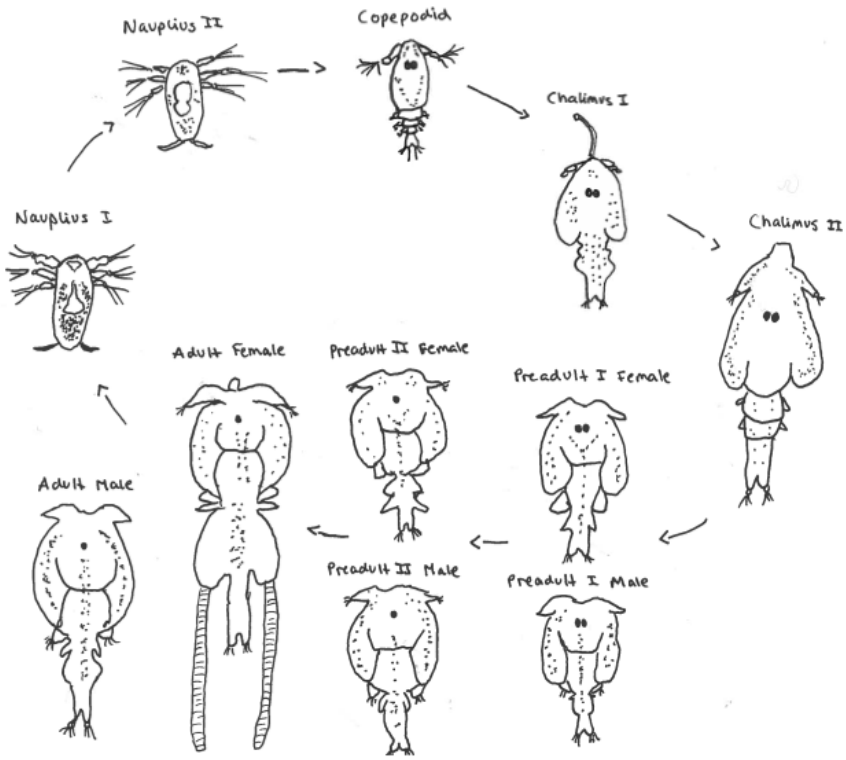



*Figure 1. Life cycle of salmon lice involves 8 phases including nauplius I and II, copepods, chalimus I and II, preadults I and II, and adults.* Illustration by *N. Brække*.

Water temperature is a key regulator of the development of all life stages of salmon louse, including the planktonic and parasitic stages (Johnson & Albright, 1991; Heuch et al., 2000). A temperature rule purposes that ectoparasites develop faster in warmer sea temperatures (Stien et al., 2005). The Institute of Marine Research (2019) observed salmon lice with a 100% hatching success between 15-20 ℃, and poor hatching success below 5 ℃. Even if the female louse hatch at low temperatures, the nauplius larvae will have problems developing into copepodite (Boxaspen & Næss, 2000). With a sea temperature of 5 ℃, salmon lice proved to have a generation time of 111 days (555 day degrees), with subsequent ovum production after 21 days. With a sea temperature of 10 ℃, salmon lice proved to have a generation time of 47 days (470 day degrees), with subsequent ovum production after 9 days (Dalvin & Oppedal, 2019). However, to what extend salmon lice can evolve adaptations towards different environmental conditions like cold temperatures and low salinities remains to be investigated. A study reported that egg string development appeared to be adapted to low temperatures by reducing the time for egg development to be able to survive and develop to the infectious stage (Boxaspen & Næss, 2000).

## Salmon lice regulation

The number of available hosts for salmon lice has increased due to the expanded farming production of Atlantic salmon, and assorted studies have claimed that the salmon farming industry is causing higher lice infections and mortality among wild fish (Heuch & Mo, 2001; Penston et al., 2008; Penston & Davies, 2009). As an attempt to reduce the transmission of lice between farmed and wild salmon, salmon lice-induced mortality on wild fish is used as a parameter directly to regulate biomass production, implemented by the traffic light system in the Norwegian aquaculture sector in 2017 (Olaussen, 2018; Myksvoll et al., 2020). The impact of aquaculture on the environment is monitored according to the currently applicable environmental indicators. In case of, unacceptable environmental impacts are reached in a given production area, the authorities will require a reduction of the production in that production area (Norwegian Ministry of Trade, 2015).

All aquaculture facilities with production of Atlantic salmon are required to keep lice levels below an average of 0.5 adult female lice per fish, except during the migration period for wild salmon where this average is set to 0.2 adult female lice per fish (Forskrift om lakselusbekjempelse, 2012). The farming industry is pursuing multiple strategies to control

salmon lice infection rates down to acceptable levels (Torrissen et al., 2013). The most used methods to mitigate salmon lice in aquaculture have traditionally been the use of medication through feed or bath administration (Aaen et al., 2015). However, repeated medical treatments, insufficient dispersal, and inaccurate dosage have led to the development of delouse agent resistance among salmon lice (Rae, 2002; Aaen et al., 2015; Overton et al., 2018). Therefore, mechanical treatment, also referred to as non-medicinal substitutes, is today the most frequently used delousing treatment in Norway (Sommerset et al., 2022; Walde et al., 2022).

## Population genetics

Population genetics can be defined as the study of genetic makeup of biological populations and the changes of genetic makeup because of action of various evolutionary processes such as mutation, random genetic drift, gene flow, and natural selection, using genetic markers (Wilding et al., 2001; Luikart et al., 2003). In population genetics, evolution is defined as a change in the allele frequencies in a population over time (Hartl et al., 1997). Allele frequencies are an estimation of how frequently an allele is present in a population of the same species. When allele frequencies become different between populations, it is called genetic differentiation (Luikart et al., 2003). $F_{ST}$ is a valuable tool in population genetics as it provides a quantitative measure of genetic differentiation among populations, explained as the proportion of the total genetic variance ($F$) contained in a subpopulation ($_S$), relative to the total genetic variance ($_T$) (Holsinger & Weir, 2009). High $F_{ST}$ values may indicate elevated levels of gene flow and strong genetic drift, while low $F_{ST}$ values may indicate high levels of gene flow and low genetic differentiation due to migration (Larson et al., 1984). This is based on the theory of the Hardy-Weinberg equilibrium (HWE), which states that allele frequencies remain constant in future generations in large random mating populations when the conditions of no mutation, no migration, and no selection hold (Chen, 2010).

Natural selection works on heritable variation in fitness, which is a crucial factor in allele frequency change and environmental adaptation. Selection pressure results in the evolution of phenotypes that are best adapted to the environment, especially in a population with a high genetic variation in alleles (Wade & Kalisz, 1990 ). Individuals with a high fitness rate have a higher chance of surviving and reproducing. Other mechanisms like genetic drift and gene flow, however, can also have an impact on the allele frequencies, but, in contrast to natural selection, these forces are more random components of variation in reproductive success. Although

genetic drift occurs in a population of all sizes, its effect tends to be stronger in smaller populations by reducing the genetic variation and increasing the homozygosity (Colonna et al., 2013). Genetic drift and a bottleneck represent a random subset of alleles and their associated frequencies, compared to the original population (Nei et al., 1975). Surviving individuals of a genetic bottleneck can colonize a new isolated environment, but with a random subset of alleles and frequencies of the ancestral population, which is known as the founder effect (Matute, 2013).

Evolution depends on genetic variation among individuals. All genetic variations have their origin in mutations and are determined as a change in the nucleotide sequence of the genome from parent to offspring (Wright, 2001). The mutation rate varies among individuals and populations, from changes in single nucleotides to rearrangements of sequences of DNA. Small mutational changes in the germ cells can be transferred from one generation to another and may give the next generation advantages against natural selection in the future (Hartl et al., 1997). However, population-wide accumulation of mutations with positive effects on the population fitness landscape is strongly dependent on the effective population size. It will often take many generations before the mutation will occur as the organism gradually changes its genetic composition over time (Nadler, 1995).

## Population genetics of salmon lice

Despite the importance of knowing the genetic structure and population dynamics of salmon lice, few studies have been conducted to assess the population genetics of salmon lice in Norway. The last decades advancement in DNA and RNA sequencing technologies have resulted in a new area of population genomic insight (Etter et al., 2011). For salmon lice, however, many earlier studies have struggled with technical challenges, low samples size and small numbers of polymorphic markers, or a combination of these challenges, or from not using up-to-date sequencing and analytical techniques (Nolan et al., 2000; Glover et al., 2011).

Previous population genetic studies on salmon lice found no significant genetic differences for geographically separated populations of salmon lice within the Atlantic Ocean and for that reason thought to be part of a larger panmictic North-East Atlantic population (Todd et al., 2004; Nolan & Powell, 2009; Glover et al., 2011). These studies have used classical genetic markers such as microsatellite loci, which can be useful on a large geographical scale due to

their high degree of polymorphism (Tautz, 1989; Chistiakov et al., 2006). On the other hand, microsatellites are allegedly neutral genetic markers and rarely located in, or near, genes or coding regions of the genome. Hence, resolving genetic differences on smaller geographical scales, in populations with large effective population sizes, and for populations that experience high rates of gene flow using microsatellites is difficult as they will not reveal that adaptive differences often responsible for driving population structure under such scenarios (Goldstein et al., 1995; Muneer, 2014). Recently, Restriction-site Associated DNA-sequencing (RADseq) was developed, which obtains a mid- to high-density representation of genetic markers across the genome of any organism (Etter et al., 2011). Studies using this method have resulted in genetic differentiations of salmon lice populations, where the population differentiation was shown to be driven mainly by genetic markers associated with resistance against chemical treatments (Carmichael et al., 2013; Jacobs et al., 2018)

Whole genome sequencing (WGS) (Bentley, 2006) is a step towards improving the resolution of genetic population studies and allows us to obtain sequences of both the nuclear genome and the mitochondrial genome simultaneously from an individual (Yue & Wang, 2017). More recently, single nucleotide polymorphisms (SNPs) analysis has become a common approach in molecular studies such as e.g., studying genetic markers in WGS studies (Chen et al., 2018). SNPs are defined as loci with alleles that differ at a single base with a frequency of at least 1% in a random set of individuals in a population (Keats & Sherman, 2013). The genetic markers are distributed throughout the entire genome, and thus reflect the complete genetic signature of an organism, population, or species (Nadler, 1995). WGS has identified peaks of differentiation in Atlantic cod in the Gulf of Maine driven by genes associated with spawning and temperature (Clucas et al., 2019b) that were not detected with RAD-seq (Clucas et al., 2019a). Despite being, still, a high-cost sequencing method, whole genome high-density coverage of SNPs can be necessary to understand mechanisms of adaptation. Recent improvements in SNP genotyping make it more attractive for population studies, especially when the improved technology has made the sequencing more affordable (Lou et al., 2021).

WGS presents challenges for statistical and bioinformatical data analyses as the high dimensionality of genomic features makes the work no longer easily achievable (Chen & Ishwaran, 2012). Random forest is an ensemble of decision trees that perform classification and or regression tasks, where each decision tree is constructed using different bootstrap values from training dataset (Chen et al., 2004). Such a machine learning approach is useful in

population genetic studies to identify SNPs important for the study question and to reduce the complexity and size of the dataset (Sylvester et al., 2018). Important SNPs, also called outlier loci, are a subset of genetic markers that deviate from the neutral expectations of the genome-wide distribution of genetic variation (Luikart et al., 2003). The increased use of random forest has provided insight into the genomic regions that are often involved in local adaptations, environmental stress, or selective pressures (Andrews et al., 2023).

This master project has been conducted as an evolve and resequencing experiment (E&R) to investigate the genomic responses of adaptation during experimental evolution over a relatively short time-scales. An E&R experiment involves several steps: (1) establishing a starting population, (2) experimental treatment in terms of selective pressure, (3) monitoring population growth by regular sampling intervals, (4) resequencing by whole genome sequencing, and (5) analyses of the evolved population (Long et al., 2015). A similar study using an E&R experiment setup proved evolved resistance against emamectin benzoate (SLICE®) among salmon lice in the Pacific Ocean (Godwin et al., 2022). Upon exposure to salmon lice-specific selective pressures, for instance delousing treatment, such approach would generate insights into how the genetic variation of salmon lice is evolving over time.

# Objectives

The high fecundity of female salmon lice contributes to a rapid population growth of salmon lice. This, in addition to having a high dispersal rate, leads to a need for repeated lice treatments in aquaculture facilities. Regulation of salmon lice in Norwegian aquaculture is based on the existence of a panmictic population of salmon lice and all fish from an aquaculture site are treated the same way. Nevertheless, the high load of salmon lice is still a problem for production, as it reduces the fish welfare of Atlantic salmon.

The aim of this master´s project was to study the temporal population genomics of salmon lice within an Atlantic salmon farming facility continuously over the period of one year. The combination of obtaining genome-wide coverage of SNP markers, and performing an E&R experiment, represents a powerful approach to study salmon lice populations within an aquaculture facility over time. As all adaptations are genetically based, I therefore hypothesize:

$H_0$: Salmon lice are genetically similar the entire production throughout the entire production cycle.
$H_1$: Salmon lice are genetically different and population structure occurs within the net pens over time.

## Expectations of the Hypothesis

The null hypothesis anticipates that the first salmon lice entering the net pens represent a panmictic population, as shown in previous studies (Todd et al., 2004; Nolan & Powell, 2009; Glover et al., 2011). As the population must be considered infinite, the genotype frequencies will be in Hardy Weinberg equilibrium throughout the production implying that natural selection are not considered to influence the population genomics of Norwegian salmon lice (Mayo, 2008).

It has been suggested that the swimming ability of salmon lice and possible transfer between hosts is limited (Bruno & Stone, 1990). A study has shown that salmon lice, even the adult stage, are host specific but can jump off the chosen host and attach to another host. Most adult male lice changed host within 24 hours, and adult female lice remained on their initial host for

a longer period (Ritchie, 1997). Hence, combined with free-flowing eggs and nauplii larvae, an alleged panmictic population of salmon lice, and free gene flow among net pens and other farms, no change in genetic variation is expected for the samples analyzed through time in this study.

The alternative hypothesis proposes that genetic drift, natural selection, and reproduction will affect the population structure of salmon lice within net pens through production. The fish in this study was treated with medical feed from the period 23.08.2021 to 02.09.2021, to avoid a large population of copepodites attaching and establishing in the net pens at an early stage of production. A delousing treatment can cause a bottleneck effect in which the population number is rapidly limited in population size. It must be expected that individuals that survive a delousing treatment harbor alleles that induce resistance and that these alleles will be represented at a higher frequency after the treatment. Mutation is not expected to be a component in resistance mechanisms as they cannot accumulate in the entire population within a few generations. Despite this, microevolution can occur where changes in gene frequency or chromosome number occur in a population within a brief time from one generation to the other (Nadler, 1995; Helmstetter et al., 2022).

The first salmon lice were sampled in August 2021, after the given slice treatment, when the sea temperature was at its highest (Appendix I). There is a high probability that larvae have hatched and developed during this period. The coldest temperatures were measured in March, which leads to limited development and hatching success. Considering the changes in temperature, it is expected that approximately four generations of salmon lice were included in this study.

# Material and methods

## Location and sampling

The project was based on the aquaculture facility Skogshamn which is in Solbergsfjorden at Dyrøya, Northern Norway (Figure 2). The facility operates under an experimental research permit which is owned by the Tromsø Aquaculture Station (Havbruksstasjonen i Tromsø), UiT the Arctic University of Norway, and NOFIMA, and daily operations were performed by SalMar Farming AS.

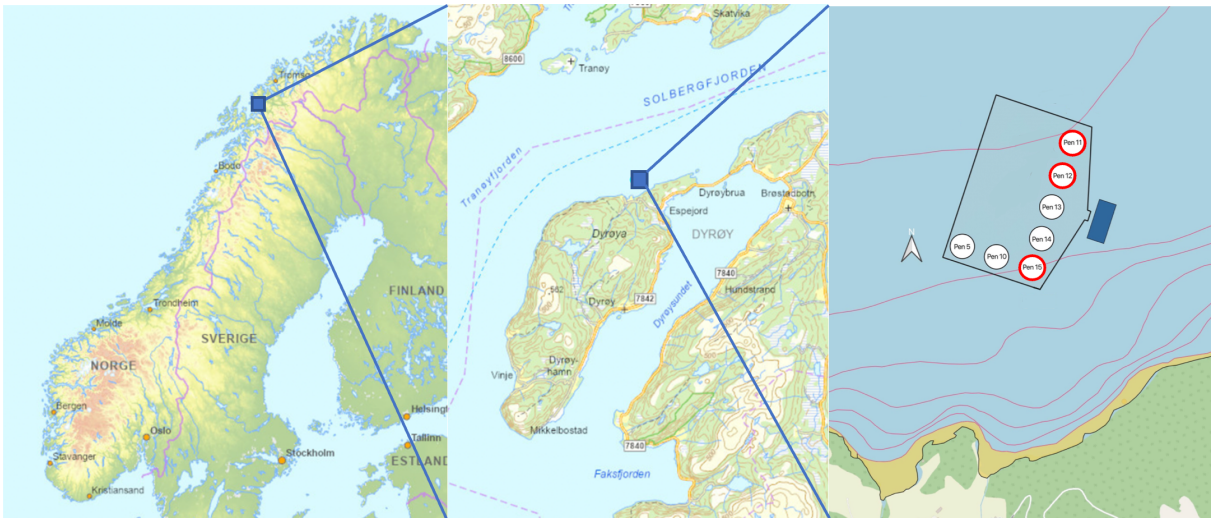



*Figure 2. Location of Skogshamn aquaculture site in Solbergsfjorden at Dyrøya. The illustration on the right side shows the aquaculture facility with associated net pens. The net pens used for this project are highlighted with red circles (net pen 11, net pen 12 and net pen 15). The map sections are obtained from Fiskeridirektoratet (Fiskeridirektoratet, 2023) and Illustrated by N. Brække.*

In the project (LICE-MAP2) that this Master´s thesis a part of, salmon lice were collected every week from August 2021 to January 2023, and thus, include specimens from the entire production cycle at the Skogshamn facility. A total of 152 samples were included in this thesis project (149 individual salmon lice and three blanks (20µL AL buffer)). The parasites were hand-picked directly from the skin of Atlantic salmon and preserved in a 2ml cryotube with 96% EtOH in a -20 ºC freezer. Four different sampling points were selected with an interval of 13 or 14 weeks apart (Figure 2). The Atlantic salmon was placed in the net pens during week 35 (year 2021) with subsequent occurrence of salmon lice. For that reason, the first sampling

point is referred to as the 'base population' and involves the different life stages chalimus, preadult I and II, and adult. The first sampling point includes several weeks, to get an understanding of what kind of individuals are present from the start, and how they establish in the aquaculture facility. Recording to the other sampling points there has been a random selection of 12 salmon lice from each net pen with approximately the same number of adult female and adult male lice. Salmon lice have been collected from net pens 11, 12, and 15 from four different timepoints (Figure 3).

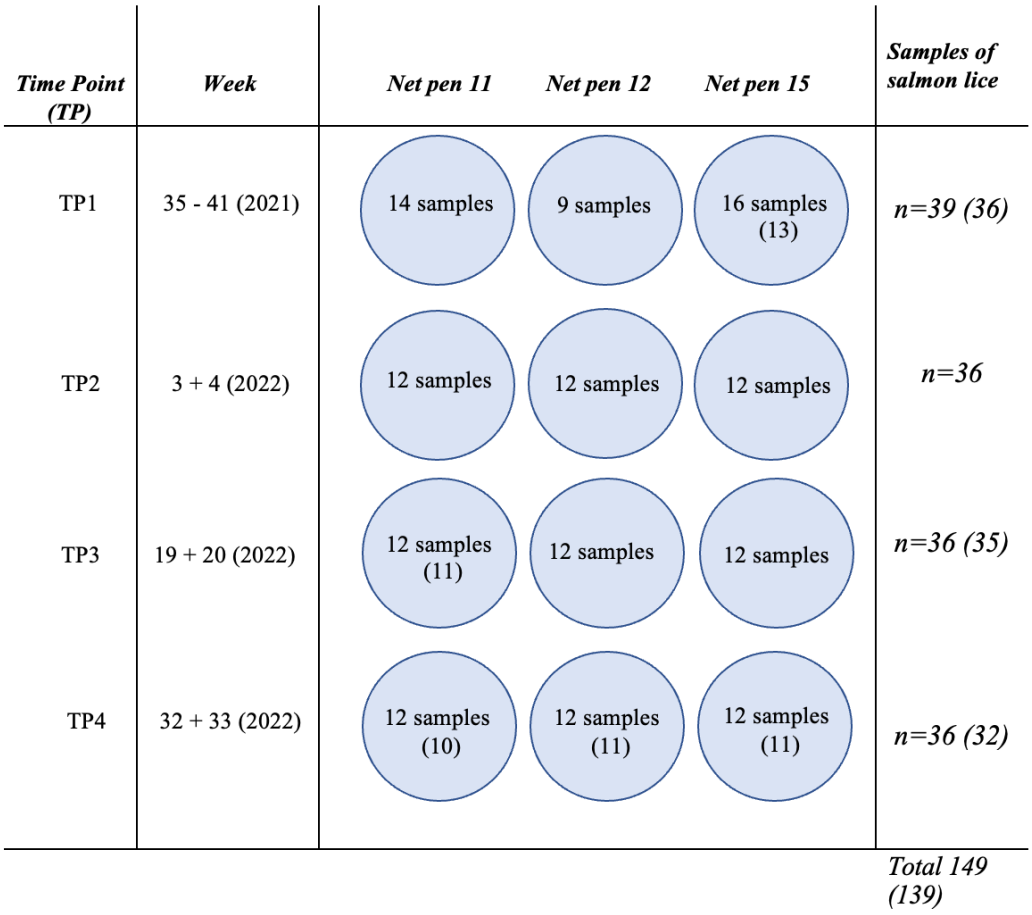



*Figure 3. Overview of samples collected for this project. Samples were collected at four different timepoints with an interval of 13 or 14 weeks apart, during one year of production from August 2021 to August 2022. TP1 represents the base population with a total of 39 salmon lice. The other timepoints include 36 salmon lice each, distributed equally among net pen 11, net pen 12 and net pen 15. A total of 139 salmon lice was used for this Master´s thesis after filtering of dataset. The number inside the brackets shows number of individuals remaining after following the genotype filtering steps. For details of filtering steps see below (Table 1).*

## Dissection and DNA extraction of salmon lice

The different life stages of the salmon lice were determined using a stereomicroscope (Olympus SZX12). Female lice are usually larger than males, and the genital segment is distinguishable by two folds, and a narrowing of the abdomen as it meets the genital complex. There is a microscopically differentiation between female and male chalimus II, where the female louse has developed a larger frontal body part (cephalothorax) (Samsing et al., 2016). For the dissection of adult lice, it was sufficient to use parts of the louse as DNA material for extraction due to its large size. Both sides of the cephalothorax were cut off with a scalpel and used for extraction (Appendix A) (Figure 4). Between each dissection, scalpel was washed with 10% bleach, milli-Q water and 96% ethanol. DNA extraction was performed in spin columns using the DNeasy® Blood & Tissue Kit (QIAGEN). Minor modifications to the manufactures protocol included (Appendix A); a) at Step 2, vortex halfway through the 1.5 hours lysis in the thermomixer, vortex + spin, adding RNase A and followed by lysis in the thermomixer at 56 ºC for 10 min, and b) at Step 7, heated buffer AE in the thermomixer, added 60 mL buffer AE to the DNeasy column membrane, centrifuged for 1 min at 17.000x g to elute, and repeated step 7 by eluting with 100 mL buffer AE in a second 1.5 mL microcentrifuge tubes.

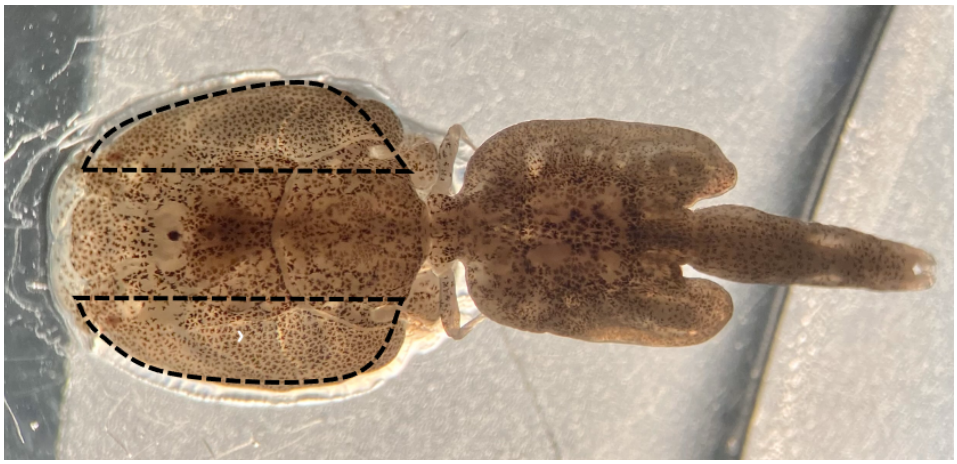


*Figure 4. Illustration of an adult female louse with marked areas of cephalothorax, used for DNA extraction with DNeasy® Blood & Tissue Kit (QIAGEN). Illustration by N. Brække.*

Chalimus and preadult stages are often smaller in size (Hamre et al., 2013). To get acceptable material for downstream analyses, the entire louse was used for extraction, following Isolation of Genomic DNA from Tissues (QIAGEN). Modifications of the protocol (Appendix B)

involved; a) at Step 4, added 6 mL RNase A, lysis for 10 min and incubated for 2 min, and b) at Step 12, heated buffer AE in thermomixer and added 40 mL buffer AE to the membrane.

To ensure samples with high molecular DNA, the extractions were analyzed by gel electrophoresis on a 1% agarose gel (see Appendix C for gel electrophoresis results). The concentrations of double-stranded DNA in each sample were measured using both Qubit4™ dsDNA BR Assay Kit (Thermo Fisher Scientific) and Quant-iT™ PicoGreen™ dsDNA (Thermo Fisher Scientific). The purity of the DNA extracts was assessed using NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific), by measuring the absorbance of the samples at wavelengths 260/280 nm and 260/230 nm. The 260/280 nm ratio is a measure of the purity of nucleic acid samples. A ratio of 1.8 is generally accepted as pure DNA and deviating values may indicate the presence of RNA, protein, phenol, or other contaminants (ThermoFisherScientific, 2010). The 260/230 nm is another measure of nucleic acid sample purity. A ratio of around 2.0 is generally considered indicative of a pure nucleic acid sample, as it suggests minimal contamination used during the DNA extraction. A ratio significantly lower than 2.0 indicates the presence of contaminants. Each sample was measured three times, with subsequent calculations of the average of A260/A280 and A260/A230 to increase the accuracy of the measurement. The instrument was calibrated by running a blank (Milli-Q water) between each sample.

## Library preparation and Whole Genome Sequencing

DNA from 36 samples did not meet the requirements of a commercial sequencing service provider (Novogene Co, Ltd) for library preparation. Therefore, we made the libraries in the labs by the research group of genetics (RGG labs). Aliquots from 36 samples were pipetted into Polymerase Chain Reaction (PCR) well plates and sequencing libraries were constructed for each sample according to NEBNext® Ultra™ II FS DNA Library Prep Kit for Illumina® (New England BioLabs), using the > 100ng input version of the protocol. Modifications of the protocol (Appendix D) involved; a) in Step 1.1, the fragmentation process was carried out for 13 min at 37 ºC, 30 min at 65 ºC, and held for 12 min at 4 ºC, with a fragmentation size of 150 bp, b) at Step 1.2, used 33 ng as a total input based on the Qubit BR and Picogreen results, followed by a 10-fold (1:10) adaptor dilution and 1.5 mM working adaptor concentration, and c) in Step 1.4.3, used six cycles for the annealing/extension step. To facilitate multiplexing of samples for sequencing, samples were barcoded with dual indexes. The individual libraries

were quantified using qPCR and further assembled into a single pool in equimolar concentration. This step was carried out by Kristel Berg, RGG, UiT. Library preparations for Illumina paired-end sequencing on the 116 remaining samples were performed by Novogene Co, Ltd. A total of 152 samples, including three blanks were sequenced paired-end, with 150 bp read length on an Illumina NovoSeq 6000 sequencing platform (Novogene Co, Ltd) to achieve estimated depth of 15x of the 640 Mb genome (LSalAtl2s and GCA_905330665.1) per salmon louse sample (9.6 Gb per sample).

## Bioinformatics analysis

The bioinformatic analysis of the data were led by Dr. Shripathi Bhat, Norwegian College of Fishery Science (NFH). Initially, demultiplexed raw sequence data in *FASTQ* format was quality checked in FASTQC v0.11.9 (Andrews, 2010), to investigate the length of reads, presence of adaptors, uncalled bases, and base quality of reads. *Cutadapt* (Martin, 2011) was applied to trim the adapters and bases from paired-end reads, and included sequences with a base quality score below 20 or less, discard reads when ≥ 5% of the bases were not called (e.g., Ns), and reads with less than 50 bases. The trimmed *FASTQ* files were re-analyzed in *FASTQ* to confirm the adapter removal and quality trimming.

The quality trimmed reads were mapped against two different salmon lice genomes (Atlantic strain), (1) Sea Lice Research Centre, Bergen, Norway: LSalAtl2s, Jan 2013, and (2) Institute of Aquaculture, Stirling, Scotland: GCA_905330665.1, (GenBank genomes) using *BWA-MEM* (Li, 2013) read mapper. *SAMtools* (Li et al., 2009) with the commands *SAMtools view* and *SAMtools sort* were used to filter the SNPs to obtain a minimum mapping quality of 20 and to convert it into a bam file format with mapping coordinates sorted by location. To call the genotypes, the final dataset was selected based on mapping to the Scottish version of the salmon lice genome as this genome represents a chromosome-level assembly. Although the Norwegian version was assembled at a scaffold level, no noticeable difference in mapping quality was observed between the Scottish and the Norwegian genomes.

*ANGSD* (Korneliussen et al., 2014) was used to call genotypes (in our case) from bam files generated from mapping processes. Strict genotyping parameters were followed to get the genotypes as opposed to genotype likelihoods. Default settings with minor modifications included: reads mapped as proper pair, minimum genotype depth (geno_minDepth) 9x,

minimum number of individuals where genotypes were called for a site (min_ind) 135 (i.e., 90%), minor allele frequency (-minMaf) was set to 5%, base quality, mapping quality was set to 20, and genotypes were stored in *BCF* format. *BCFtools* (Li, 2011) was used to convert genotypes from *BCF* format to *VCF* format. *VCFtools* (Danecek et al., 2011) and *plink2* (Chang et al., 2015) were utilized for filtering the dataset to ensure a *VCF* file with high-quality filtering. Mitochondrial sequences and non-chromosomal were removed and only sequences associated with chromosomes were retained. The sequencing depth per locus (SNP site) and per individual were calculated, to be used in the next filtering step. Further filtering contained a minimum mean sequencing depth of 15x, and a max sequencing depth of 90x (twice the mean depth of ~45x). Only biallelic SNPs were retained in the filtered *VCF* file. The *VCF* file was further filtered to linkage disequilibrium between the loci in *plink2* with following options: *-indep-pairwise 50 5 0.2* to remove linked SNPs. The last filtering step included removal of individuals with missing data of 20% or more.

## Population genetic analyses

Two datasets were prepared to test the study hypotheses. The first dataset consisted of all the SNPs filtered following the above explained filtering steps and are referred to as the "full dataset". From this dataset, reduced datasets were generated containing SNPs important for the hypotheses were generated using random forest (Liaw & Wiener, 2002), which will be referred to as the "important dataset (net pens)" and "important dataset (timepoints)" in later chapters, following the below explained filtering steps.

Identification of important SNPs were conducted by random forest analysis and performed by Dr Shripathi Bhat. The full dataset was divided randomly into a training dataset consisting of 2/3 of the data which was used to determine the association between classes within classification variables (net pens or timepoints) and SNP markers. The remaining SNP data, 1/3 out of bag (OOB), were used to evaluate the predictive power of SNPs to correctly classify the individuals into the right origin (net pens or timepoints). The *VCF* file containing all the SNPs was first converted to numerical codes in *plink2*, and the genotypes was coded as 0 (homozygous for reference allele), 1 (heterozygotes), and 2 (homozygous for alternative allele). The data was prepared for random forest using a combination of inhouse script and a modified version of the openly available script implemented by Brieuc (2018). Missing genotypes were

imputed by using the function *rflmpute* in the random forest package. Each random forest run was performed 3 times with 10.000 trees. Mean Decrease in Accuracy values (MDA) was used as filtering metrics (importance scores) to remove the less important SNPs. Based on the importance of scores of the loci in each initial run (on the full dataset), a small subset of data was generated based on the different quantile probabilities (1%, 3%, 5% and 10% of the lowest OOB error rate) and random forest run was run 3 times. OBB error rate was recalculated. SNPs subsets (the ones based on quantile, here 1% subset gave least OBB in comparison to other subsets) having the lowest OOB error observed in this study were further used for backward purging approach and were removed from the dataset (Holliday et al., 2012). Doing this initial quantile-based sub setting helps to reduce the dataset to run the final backward purging based on random forest analysis. The backward purging continued until two SNPs were left and chose groups of SNPs where the OOB error rate was at the lowest (elbow shape curve with the number of SNP Vs OOB error rate).

## Population structure analyses

The following analyses were performed separately for the full dataset, and the two important datasets based on net pens and timepoints: principal component analysis (PCA) (Appendix F), Discriminant Analysis of Principal Components (DAPC), and pairwise $F_{ST}$ (see Appendix E for R-scripts).

### $F_{ST}$ calculation

The R-package called *StAMPP* was used to calculate the pairwise $F_{ST}$ between the net pens and the timepoints, and the associated *p*-values. Associated p-values were obtained with bootstrap value of 1000 and the *p*-values were adjusted for multiple comparisons with Benjamini-Hochberg (Benjamini & Hochberg, 1995) method using *p.adjust* function in R-studio. A heatmap was used to present pairwise $F_{ST}$-values and values were plotted using the R-package *heatplus*.

### PCA and DAPC analyses

The initial population structure exploration was performed with an inbuilt PCA option in *plink2*. Eigenvalue and eigenvector files were imported and plotted in R-studio using R-package *tidyverse* and *ggplot2*. Discriminant analysis of Principal Components (DAPC) plots are extensively used in population genetics to explore and visualize the genetic population structure

of a dataset and it was performed with the R-package *adegenet* (Jombart & Ahmed, 2011). The data was first transformed and submitted to a PCA, and the second principal components of PCA were submitted to a Linear discriminant analysis (LDA). The command *find.cluster* was used to find the putative number of clusters present in the dataset *de novo*. In prior-based DAPC, the DAPC analysis was fed with priors, as in this case were either net pens (prior of 12 net pens or prior of 3 net pens each timepoint) or timepoints (prior of 4 timepoints). The function *xval* was used to optimize the choice of the number of PCs required for DAPC analysis, which was recommended by the authors of *adegenet* (Jombart & Ahmed, 2011). *Xval* was run 5-10 times, depending on the size of the data. When many xval runs have the same number of PCs selected, the lowest number of PCs was selected as the final PC to use in DAPC analysis. If too many PCs are included in a DAPC analysis, this can result in overfitting the model.

All the tasks described above are illustrated in a flow diagram (Figure 5). The flow diagram includes laboratory work and bioinformatic analysis. All the steps except for sampling and sequencing were performed at NFH, UiT.

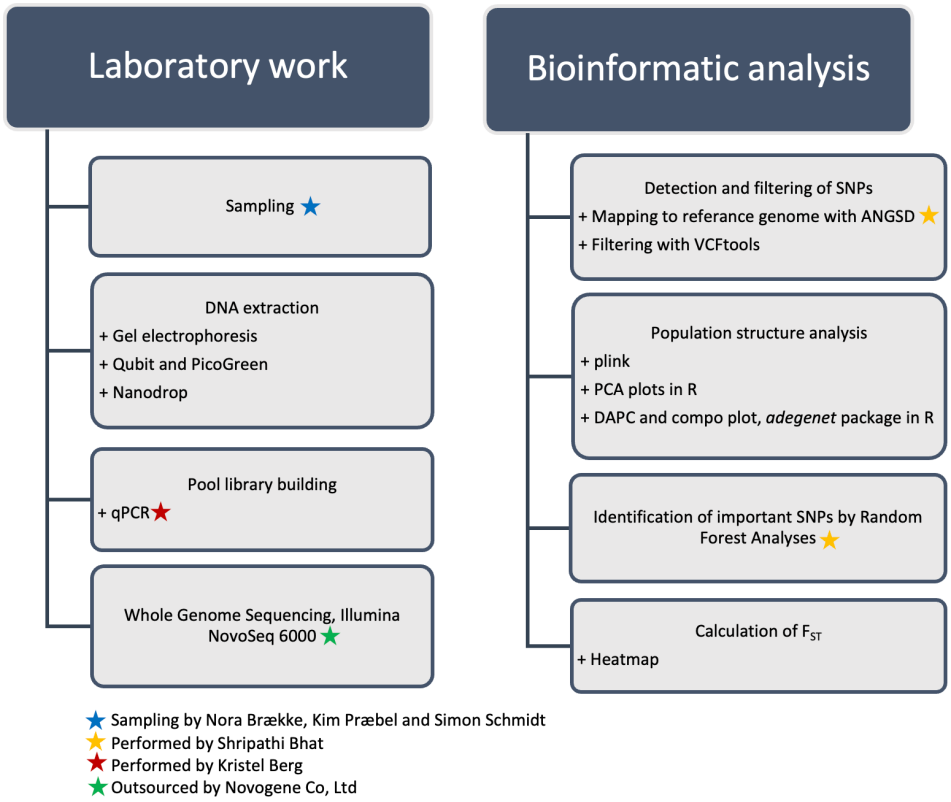



*Figure 5. Flow diagram that shows a summary of the entire workflow of this Master's thesis. It is divided into laboratory work and bioinformatic analyses. The stars represent contributors to this project. No star means that Nora Brække performed the tasks.*

# Results

## Sequencing and genotyping

All libraries passed qPCR with high concentrations (from 5-200nM adjusted for fragment size). A total of 152 samples including three blanks were sequenced on an Illumina NovoSeq 6000 sequencing platform which obtained 10,572,097,494 raw reads, corresponding to a mean of 66,912,009 (SD ± 7,366,985) raw reads per individual.

The *ANGSD* output consisted of 377,579 SNPs and was further filtered based on different filtering criteria (Table 1). 10 individuals having 20% or more missing genotypes was removed from the dataset as these individuals did not meet the filtering criteria, which in this case seemed to be randomly assorted among the timepoints and the net pens (Figure 3). After the filtering steps, the full dataset contained 66,769 SNPs per individual. Further random forest analysis using timepoints as classification variables resulted in 261 SNPs (obtained after the OOB error rate was reduced to ~9% from the initial ~67% in the full dataset). The random forest analysis with net pens as classification variables resulted in 107 SNPs (obtained after OOB error rate was reduced to ~47% from initial ~86% in the full dataset).

*Table 1. Summary table of SNP dataset filtering showing filtering steps and following information regarding the filtering steps. The highlighted numbers represent the full dataset with a total of 66,769 SNPs, and the important datasets of 261 SNPs (timepoints) and 107 SNPs (net pens).*

| Filtering step | Filtering | Retained salmon lice | Retained SNPs |
|---|---|---|---|
| 0 | Raw output from *ANGSD* | 149 | 377,579 |
| 1 | Only SNP in chromosomes | 149 | 136,787 |
| 2 | Biallelic<br>Local MAF ≥ 0.05<br>Coverage from 15 to 90x | 149 | 130,555 |
| 3 | LD test ($r^2 \leq 0.2$) | 139 | **66,769** |
| 4 | Random forest based on timepoints | 139 | **261** |
| 5 | Random forest based on net pens | 139 | **107** |

## Genetic population structure analyses

The first approach was to investigate whether population structure happened based on net pens. The net pens were represented as the combination of timepoint (TP) and net pen number (example: 1_12, where 1 represents TP1 and 12 represents net pen 12). The second approach was to investigate population genetic structure based on timepoints (TP).

### Analyses based on net pens

Figure 6 illustrating the DAPC analysis based on the full dataset using net pens as prior (12 net pens) revealed the genetic variation among all net pens for all timepoints. The DAPC analysis suggest the presence of weak genetic structures (Figure 6). Net pens belonging to the same timepoints show a higher frequency of overlap, proposing individuals in the net pens are genetically related. The composition plot revealed similar patterns as the DAPC plot (Figure 7).

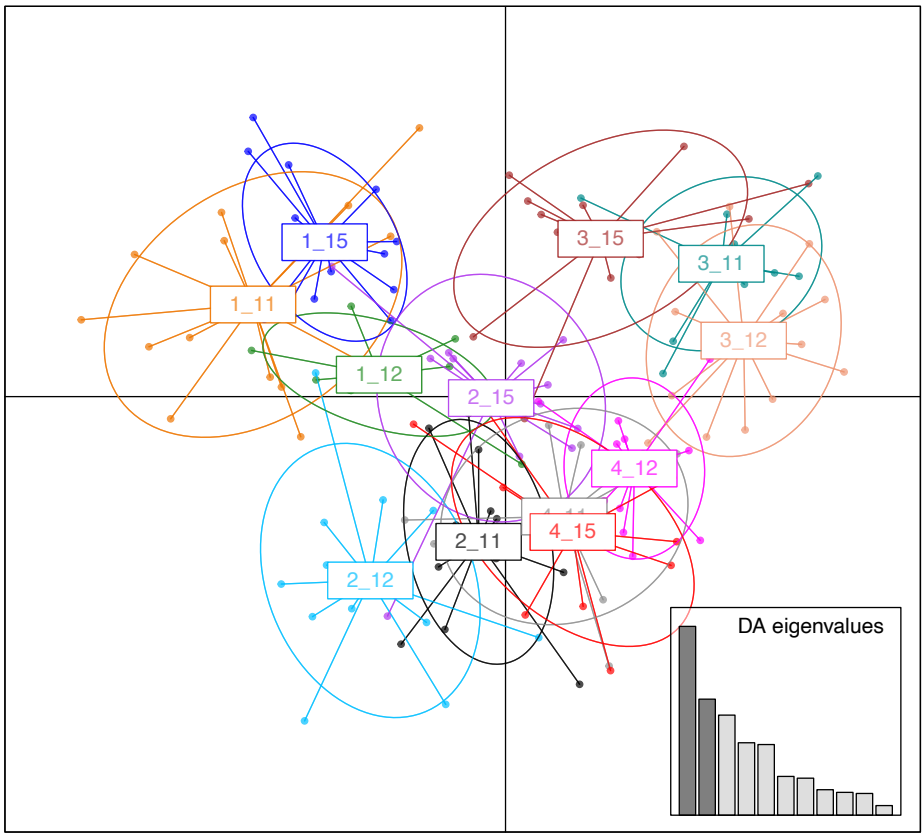



*Figure 6. DAPC scatter plots based on the full dataset of 66,769 SNPs performed with R, with a given prior of 12. The obtained graph represents the 139 salmon lice individuals as dots and the clustering of individuals are represented by 95% inertia eclipses. The net pens are represented as the combination of timepoint and net pen number (example: 1_12, where 1 represents TP1 and 12 represents net pen 12). Every net pen at every timepoint is represented by a color. Eigenvalues of the analysis are displayed in the inset and correspond to the ratio of the variance between groups over the variance within groups for each discriminant function. In total 11 linear discriminants explained (the two eigenvalues with the highest percentage): 26.8% and 16.4%.*

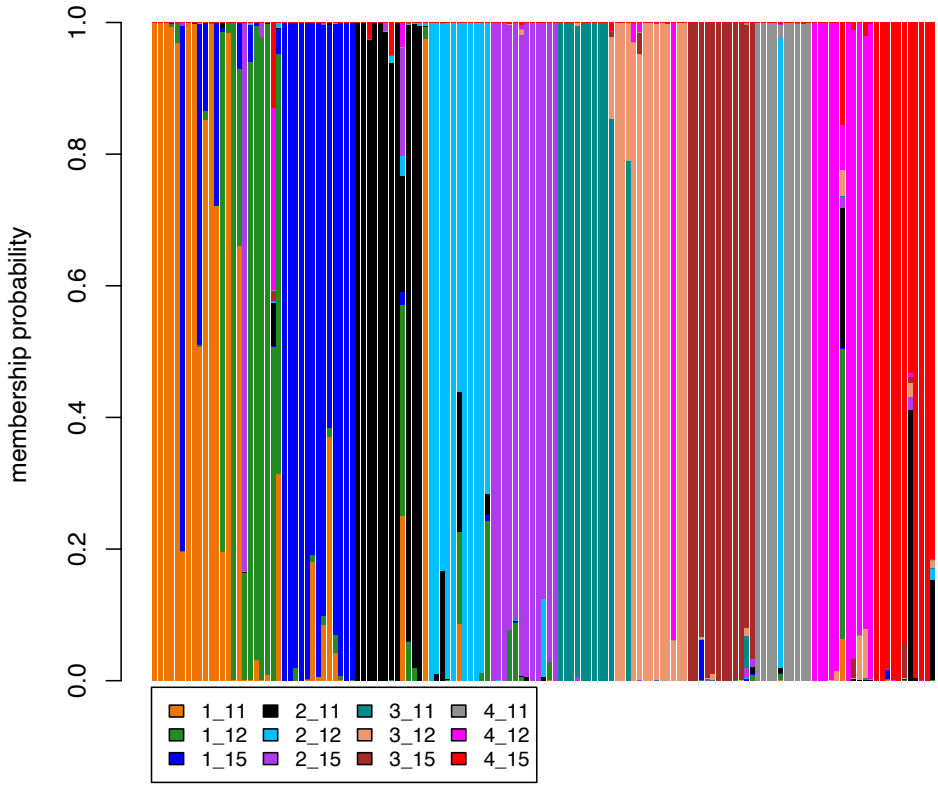



*Figure 7. Composition plot based on the full dataset of 66,769 SNPs performed with R, with a given prior of 12. The net pens are represented as the combination of timepoint and net pen number (example: 1_12, where 1 represents TP1 and 12 represents net pen 12), and every combination is represented by a color. The x axis represents the salmon lice individuals as bars, and the y axis represents the membership probabilities. Membership probability showing the population assignment probability for each salmon lice.*

The DAPC analysis was performed using the important dataset (net pens), using net pens as prior (12 net pens). A stronger pattern of genetic differentiation is seen between the net pens, compared to the distribution seen in Figure 6 (Figure 8). Net pens 1_11-15 and 2_11-15 form a dense cluster. Net pens 3_11-15 and 4_11-15 formed two, less dense, separate clusters. In the representation of 3_11-15, we observe that net pen 3_12 is distant from the other net pens in the cluster without overlapping individuals. These observations are emphasized in the composition plot (Figure 9).

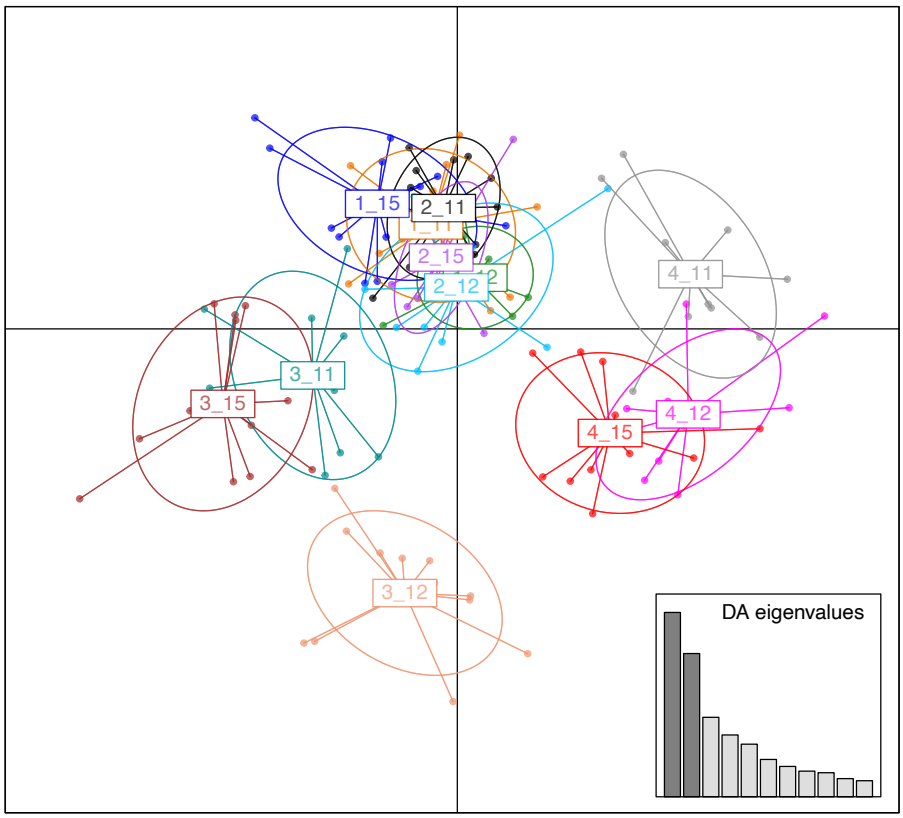



*Figure 8. DAPC scatter plots based on the important dataset (net pens) of 107 SNPs performed with R, with a given prior of 12. The obtained graph represents the 139 salmon lice individuals as dots and the clustering of individuals are represented by 95% inertia eclipses. The net pens are represented as the combination of timepoint and net pen number (example: 1_12, where 1 represents TP1 and 12 represents net pen 12). Every net pen at every timepoint is represented by a color. Eigenvalues of the analysis are displayed in the inset and correspond to the ratio of the variance between groups over the variance within groups for each discriminant function. In total 11 linear discriminants explained (the two highest percentage): 26.4% and 19.8%.*

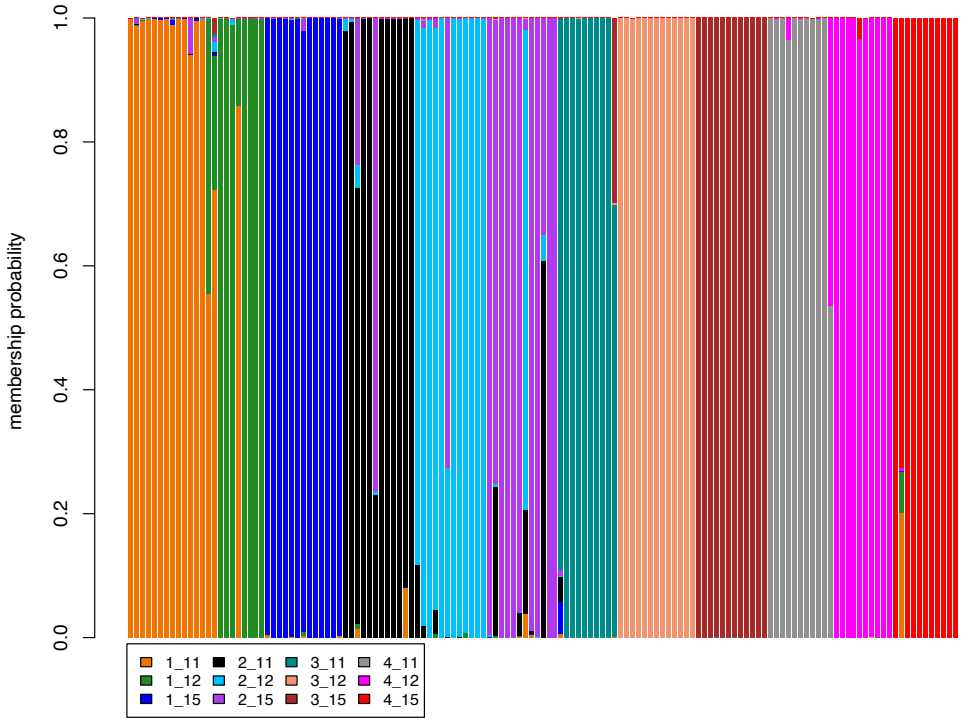



*Figure 9. Composition plot based on the important dataset (net pens) of 107 SNPs performed with R, with a given prior of 12. The net pens are represented as the combination of timepoint and net pen number (example: 1_12, where 1 represents TP1 and 12 represents net pen 12), and every combination is represented by a color. The x axis represents the salmon lice individuals as bars, and the y axis represents the membership probabilities. Membership probability showing the population assignment probability for each salmon lice.*

To further investigate the genetic differentiation among the net pens, DAPC analyses were performed for the three net pens separately at each timepoint using the full dataset, with net pens as prior (3 net pens). Figure 10 show that salmon lice belonging to TP1, TP2 and TP3 reveal a weak population structure within and among the net pens, with no observed clusters. TP4 shows a more cluster-based population distribution among the net pens, hence, we can still observe a genetic mixture of individuals from the different net pens. Equal patterns are seen in the composition plot (Figure 11).

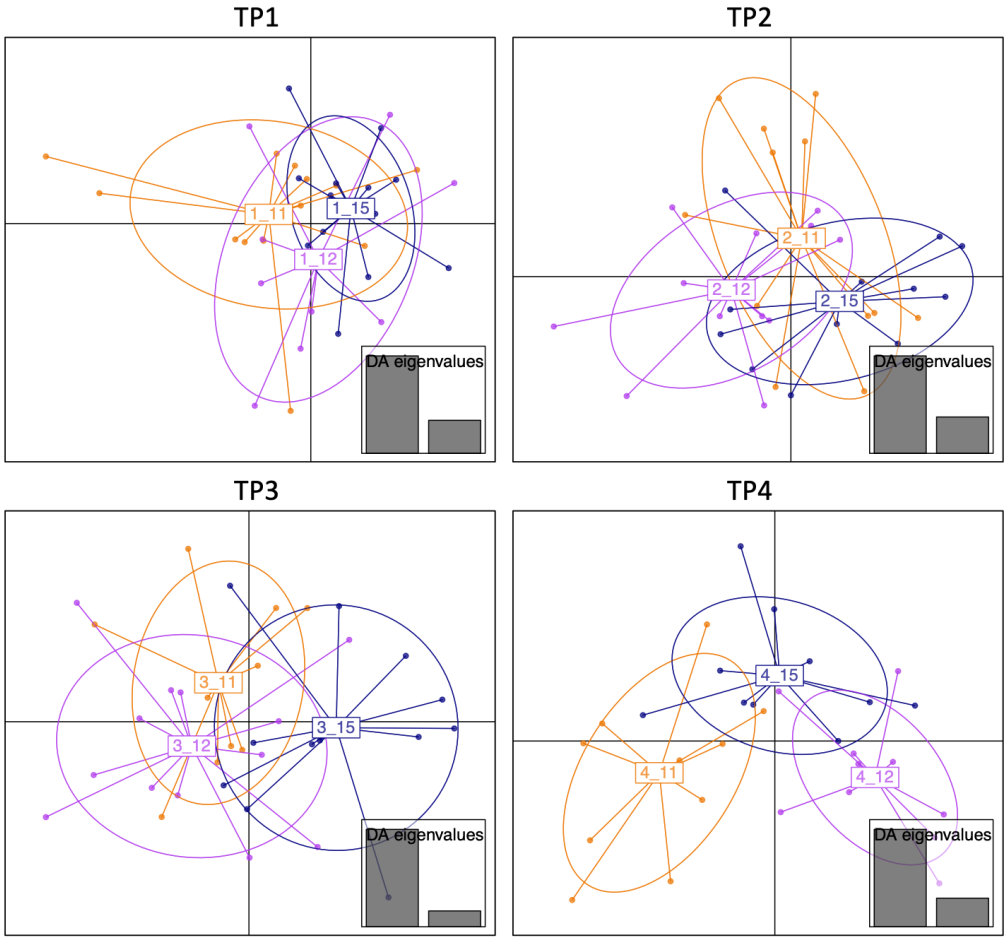



*Figure 10. DAPC scatter plots based on the full dataset of 66,769 SNPs performed with R, with a given prior of 3. The obtained graph represents the salmon lice individuals as dots and the clustering of individuals are represented by 95% inertia eclipses. The net pens are represented as the combination of timepoint and net pen number (example: 1_12, where 1 represents TP1 and 12 represents net pen 12). Every net pen is represented by a color: net pen 11 (orange), net pen 12 (pink) and net pen 15 (blue). Eigenvalues of the analysis are displayed in the inset and correspond to the ratio of the variance between groups over the variance within groups for each discriminant function. In total 2 linear discriminants explained TP1 (74.7% and 25.3%), TP2 (72.8% and 27.2%), TP3 (86% and 14%) and TP4 (77.3% and 22.7%).*

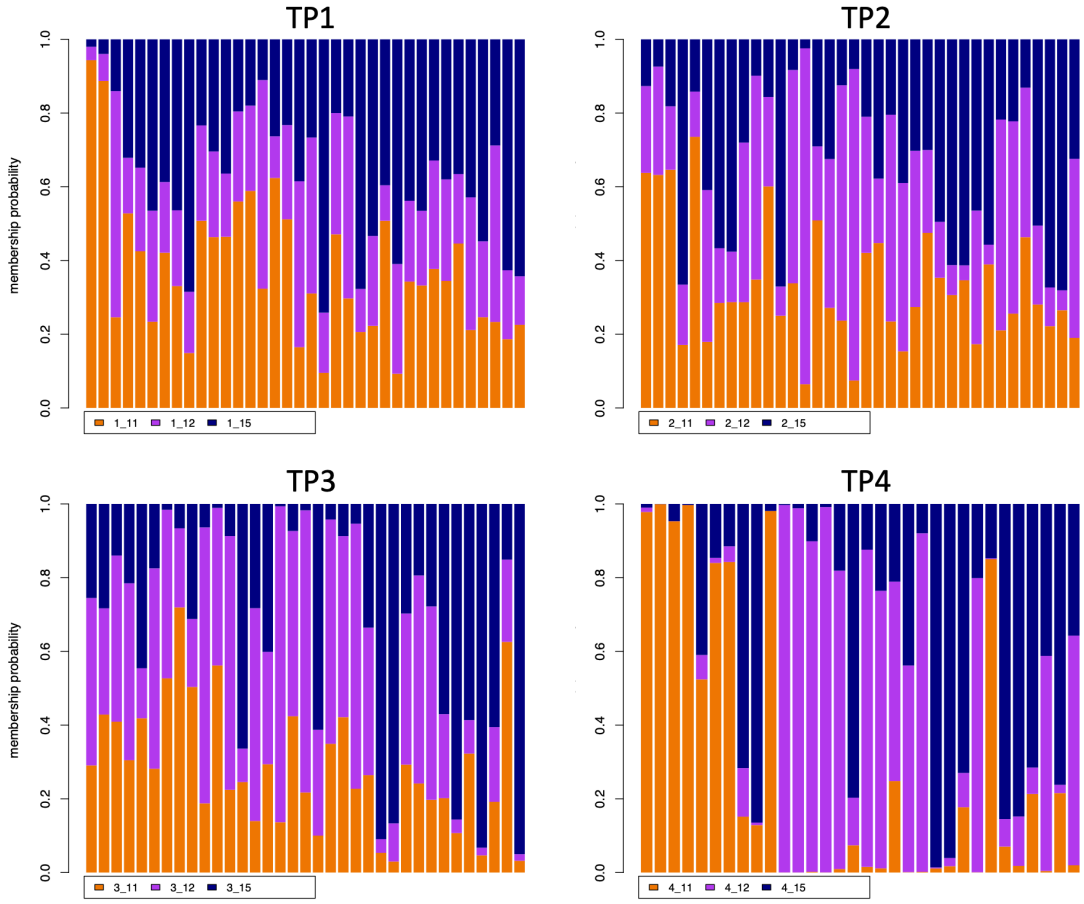



*Figure 11. Composition plots based on the full dataset of 66,769 SNPs performed with R, with a given prior of 3. Each timepoint has their own composition plot and every net pen is represented by a color: net pen 11 (orange), net pen 12 (pink) and net pen 15 (blue). The net pens are represented as the combination of timepoint and net pen number (example: 1_12, where 1 represent TP1 and 12 represent net pen 12). The x axis represents the salmon lice individuals as bars, and the y axis represent the membership probabilities. Membership probability showing the population assignment probability for each salmon lice.*

The pairwise comparisons of $F_{ST}$ among net pens were visualized in a heatmap and performed using the important dataset (net pens). The $F_{ST}$-values ranged from 0.020 to 0.080. Apart from the net pens 1_12 and 1_15, which obtained a $p$-value > 0.05 after correcting for multiple testing, all comparisons were significant ($p < 0.05$) (Figure 12). $F_{ST}$-values were the highest between the net pens belonging to timepoint 1 and 3 (0.061 – 0.080), and lowest between net pens belonging to timepoint 1 and 2 (0.020 – 0.040). Significant pairwise $F_{ST}$-values between net pens decreased using the full dataset (Appendix H).

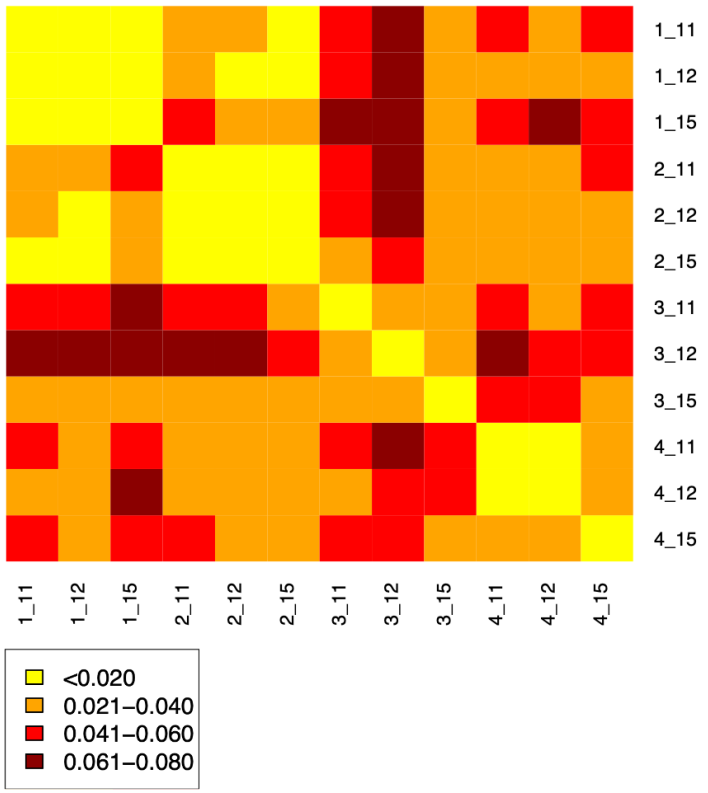



*Figure 12. Heatmap showing pairwise $F_{ST}$-values between net pens at the different timepoints, based on the important dataset (net pens) of 107 SNPs, performed in R. The net pens are represented as the combination of timepoint and net pen number (example: 1_12, where 1 represent TP1 and 12 represent net pen 12). The colors represent the range of $F_{ST}$-values: < 0.020 (yellow), 0.021 – 0.040 (orange), 0.041 – 0.050 (red) and 0.051 – 0.080 (dark red).*

## Analyses based on timepoints

The analyses based on net pens revealed a genetic population structure among the salmon lice over time. To further investigate the differentiation between the timepoints, the DAPC plot was performed using the full dataset, with given timepoints as prior (4 timepoints). The plot is illustrated in Figure 13, suggesting four clusters among the timepoints. Between the clusters we observe some overlapping individuals from each timepoint, proving a certain degree of genetic exchange. Both the DAPC plot and the composition plot reveal similar patterns of the population structure among the timepoints (Figure 13; Figure 14).

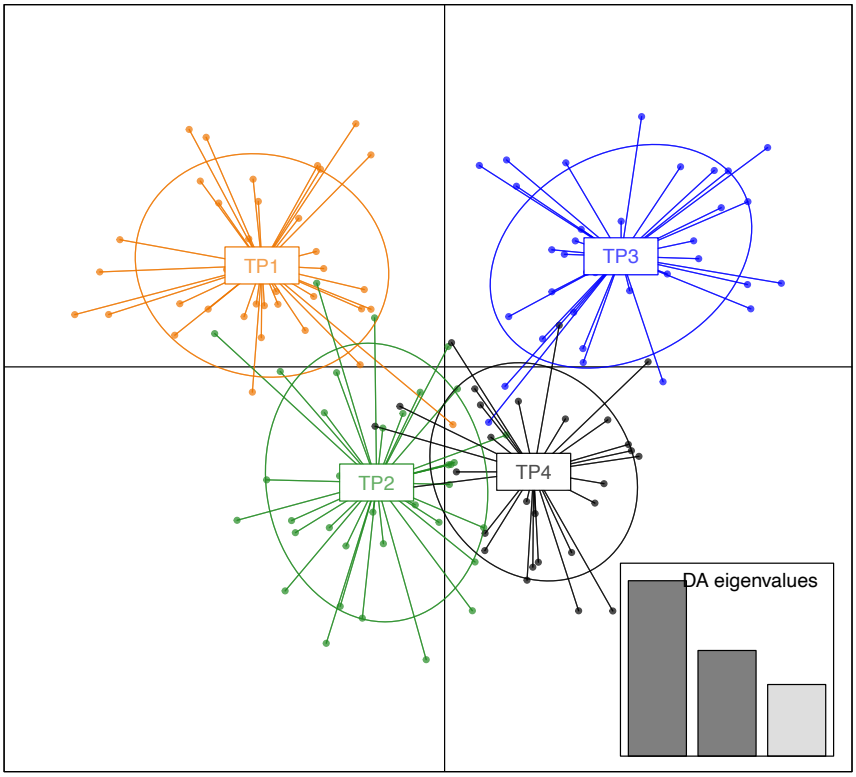



*Figure 13. DAPC scatter plot based on the full dataset of 66,769 SNPs performed with R, with a given prior of 4. The obtained graph represents the salmon lice individuals as dots and the clustering of individuals are represented by 95% inertia eclipses. Every timepoint is represented by the color TP1 (orange), TP2 (green), TP3 (blue) and TP4 (black). Eigenvalues of the analysis are displayed in the inset and correspond to the ratio of the variance between groups over the variance within groups for each discriminant function. In total 3 linear discriminants explained 49.7%, 30% and 20.3%.*

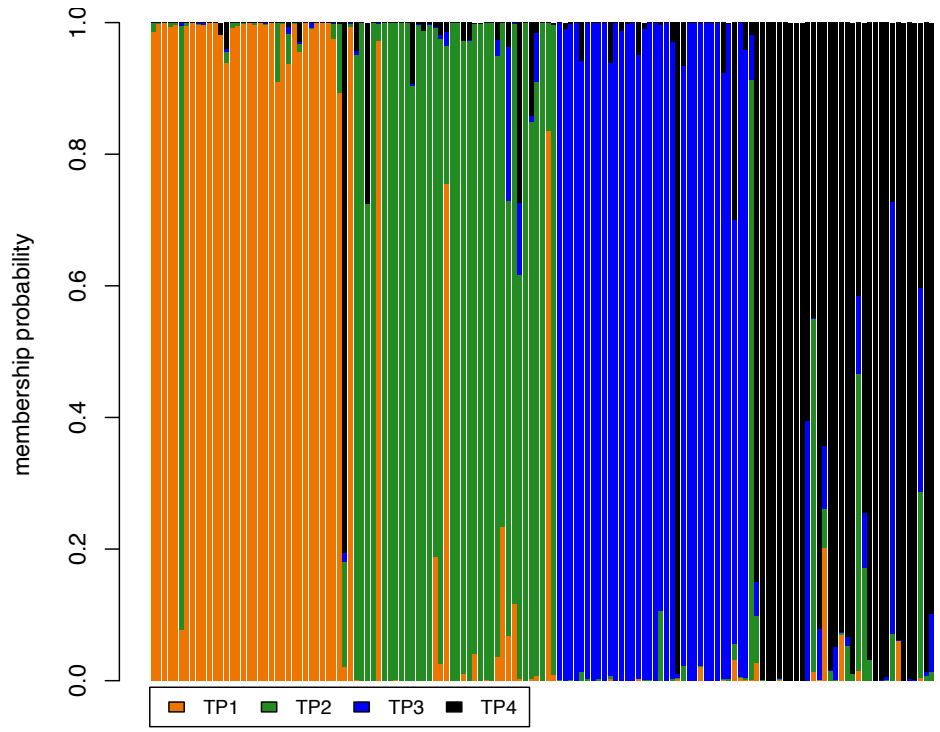



*Figure 14. Composition plot based on the full dataset of 66,769 SNPs performed with R, with a given prior of 4. Each timepoint is represented by the color TP1 (orange), TP2 (green), TP3 (blue) and TP4 (black), which indicate the population of origin. The x axis represents the salmon lice individuals as bars, and the y axis represents the membership probabilities. Membership probability showing the population assignment probability for each salmon lice.*

The DAPC plot performed on the important dataset (timepoints) increased the genetic population structure among the salmon lice within the timepoints (Figure 15). The DAPC plot suggested 4 clusters among the timepoints, including overlapping of a few individuals. The overlapping individuals are less frequent than seen in the preceding analyses (Figure 13), suggesting higher individuality within each cluster, and thereby greater genetical differences between the clusters. The analyses of the important SNPs resulted in separation of clusters, seen in both illustrations (Figure 15; Figure 16).

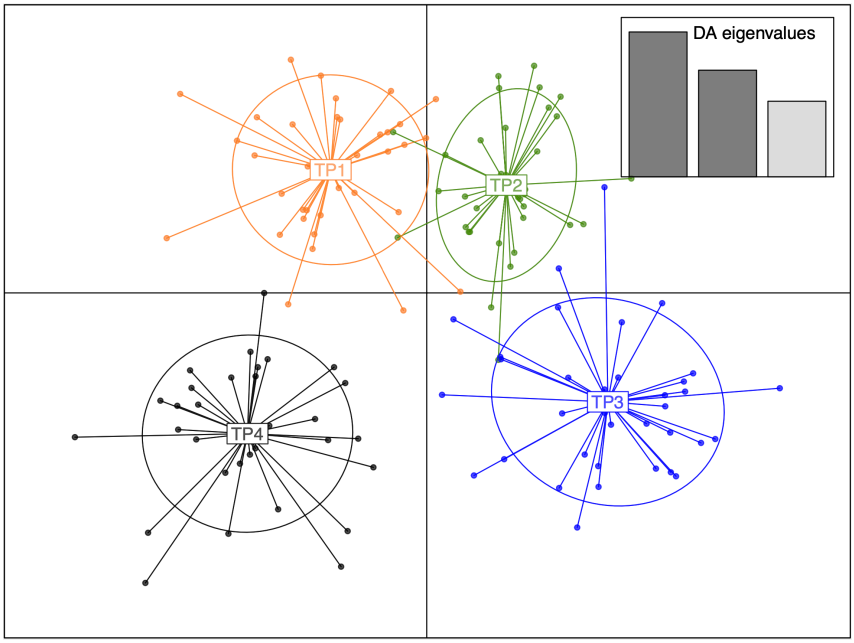



*Figure 15. DAPC scatter plot based on the important dataset (timepoints) of 261 SNPs performed with R, with a given prior of 4. The obtained graph represents the salmon lice individuals as dots and the clustering of individuals are represented by 95% inertia eclipses. Every timepoint is represented by the color TP1 (orange), TP2 (green), TP3 (blue) and TP4 (black). Eigenvalues of the analysis are displayed in the inset and correspond to the ratio of the variance between groups over the variance within groups for each discriminant function. In total 3 linear discriminants explained 40.6%, 33% and 26.4%.*

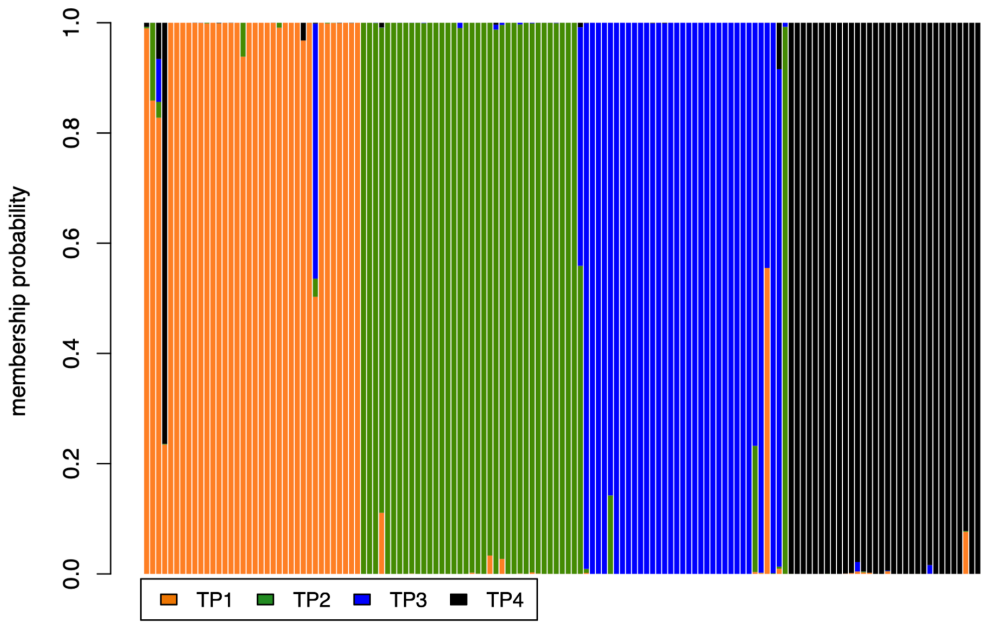



*Figure 16. Composition plot based on the important dataset (timepoints) of 261 SNPs performed with R, with a given prior of 4. Each timepoint is represented by the color TP1 (orange), TP2 (green), TP3 (blue) and TP4 (black), which indicate the population of origin. The x axis represents the salmon lice individuals as bars, and the y axis represents the membership probabilities. Membership probability showing the population assignment probability for each salmon lice.*

The pairwise comparisons of $F_{ST}$ among timepoints were visualized in a heatmap and performed using the important dataset (timepoints). The $F_{ST}$-values ranged from 0.025 to 0.050. All the comparisons were significant ($p < 0.05$) following the Benjamini-Hochberg correction (Figure 17). The $F_{ST}$-values were highest between TP1 and TP3 (0.050). Pairwise $F_{ST}$-values based on the full dataset were significant, but weaker (Appendix H).

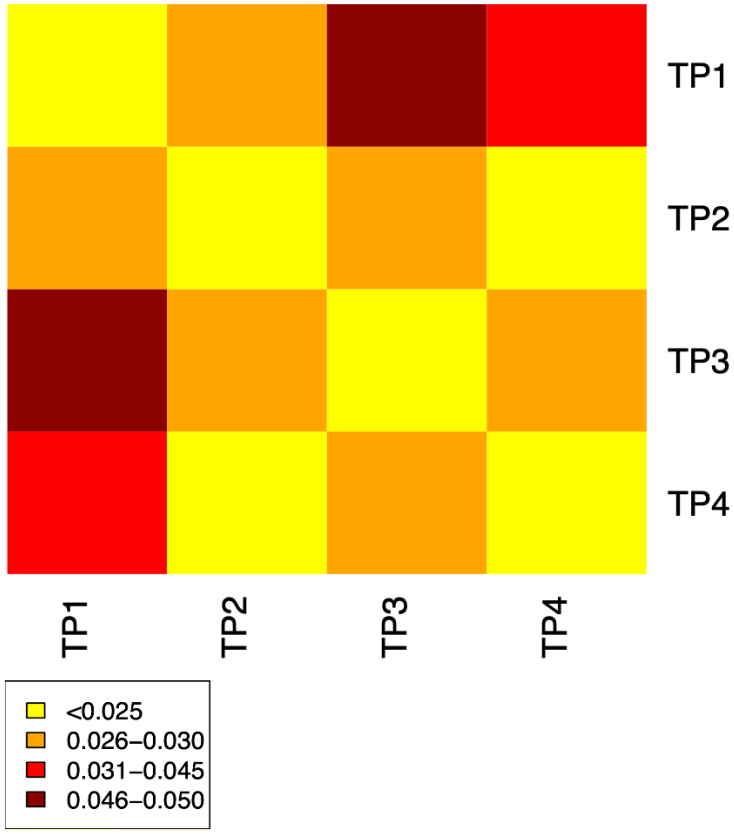



*Figure 17. Heatmap showing $F_{ST}$-values between the different timepoints, based on the important dataset (timepoints) of 261 SNPs, performed in R. The colors represent the range of $F_{ST}$ values: <0.025 (yellow), 0.026 - 0.030 (orange), 0.031 – 0.045 (red) and 0.046 – 0.050 (dark red).*

# Discussion

Advances in genomic technologies, such as whole genome re-sequencing, and the continued accumulation of reference genomes to map whole genome re-sequencing data, have enabled researchers to better understand the evolution of non-model species and the genetic mechanisms that induce diversity and adaption in populations (Lou et al., 2021). In this study, whole genome re-sequencing was applied to obtain a genome-wide genomic structuring of salmon lice populations within an aquaculture facility over time. Analyses based on the full dataset of 66,769 SNPs revealed weak, yet significant, genetic population structure of salmon lice among net pens through time. However, when combining salmon lice from net pens at each timepoint, a stronger genetic clustering was observed developing through the duration of the production. The random forest machine learning approach obtained 107 SNPs based on net pens and 261 SNPs based on timepoints, and when using these two datasets, increased genetic population clustering was observed among net pens and through time. Hence, random forest analysis in conjunction with analysis from the full SNP dataset, resulted in a comprehensive and representative understanding of the data. The results revealed that salmon lice indeed from population genetic structure among net pens within salmon farms and that the genetic differences become larger through time of production.

In the following sections, these results will be discussed in relation to spatial and temporal distribution of salmon lice, salmon lice biology and fish health implications.

## Population genetic differences of salmon lice among net pens

Structuring analysis based on net pens with the full dataset did not reveal genetically significant population structure within the 139 individuals of salmon lice. However, genome-wide differences cannot be expected to accumulate throughout all salmon lice within each net pen during such a short time-frame. The random forest machine learning approach where therefore used to explore whether genetic differences had developed among salmon lice in the three net pens. The increased resolution provided by random forest machine learning revealed significant population clustering of salmon lice among the net pens. Population studies, conducted on species such as rainbow trout and wrasse, have observed similar population patterns when using random forest (Jacobs et al., 2018; Jansson et al., 2020; Andrews et al., 2023). Observations of the DAPC analyses, using the important dataset (net pens), indicated salmon lice within net

pens belonging to the same timepoints were closely related, whereas a stronger significant divergent, clustering, was observed among the salmon lice within net pens later in production compared to the start of the production. Pairwise $F_{ST}$-values based on the important dataset (net pens) supported these observations by the presence of significant genetic differences between similar net pens belonging to different timepoints (e.g., 1_12 vs 3_12). Combined, these results show that significant population genetic structure developed among the salmon lice in the net pens, where possible drivers may be strong selection, random genetic drift, or both.

Most population studies, including sample locations in Norway, have been conducted over large geographical scales. Such studies have shown that salmon lice populations can exhibit various levels of population structure, from none or weak population structures (Nolan & Powell, 2009; Glover et al., 2011), to distinct population structures (Jacobs et al., 2018; Guragain et al., 2022). Glover et al. (2011) suggested that salmon lice sampled from Norway, Faroe Island, Shetland, Ireland, and Canada displayed a weak, but significant, genetic differentiation among these locations. However, no overall genetic variation was observed among the 19 samples collected in Norway, with a global $F_{ST}$-value below 0.0001 (Glover et al., 2011). Microsatellite loci were used as genetic markers in the study, which have shown reduced population-level resolution for non-model organisms compared to genetic markers such as SNPs with genome wide coverage (Zimmerman et al., 2020). A recent study by Jacobs et al. (2018) used SNPs obtained by RAD sequencing and found significant population genomic structuring of salmon lice individuals from Ireland, Scotland, and Norway. The observed pairwise $F_{ST}$-values, using a SNP dataset obtained using random forest, ranged from 0.081 to 0.096, suggesting a stronger genetic differentiation compared to the results obtained in this master project (Jacobs et al., 2018). However, the observed significant population structure was detected at a wide geographical scale, compared to short geographical distances used in this project. Genetic differences are expected to increase with geographical distance between the populations (Eckert et al., 2008), and may explain the higher $F_{ST}$-values.

## Development of population genetic differences of salmon lice through a production cycle

In this study, time was observed to be a crucial factor regarding the genetic differences observed among the net pens. The structure analysis based on the full dataset revealed a genetically

distinct population pattern of salmon lice within the timepoints. However, a significantly stronger genetic clustering was observed between salmon lice based on the important dataset of 261 SNPs. The pairwise $F_{ST}$ values based on the important dataset ranged between 0.025 – 0.050, and suggested significant genetic differentiation, especially between salmon lice belonging to TP1 and TP3. Salmon lice belonging to TP3 was sampled nine months after the base population was sampled, which indicated a population structure establishment over a short period of time. A small-scale population genetic study based on microsatellite markers, conducted in Ireland observed a weak, but significant, population differentiation within an aquaculture facility, by looking at salmon lice as one population over a 10-month period (Nolan & Powell, 2009). The pairwise $F_{ST}$ comparisons between the samples conducted from the aquaculture site in Ireland of 0.014 – 0.102 suggested a higher significant temporal genetic differentiation among the salmon lice, compared to the $F_{ST}$ values observed in this study. However, the method included only four microsatellite loci and a low sample size, which may have influenced the results.

Fruit flies (*Drosophila*) are a widely used model organism in E&R experiments due to its simple genetics with only four pairs of chromosomes, many offspring per generation, and a rapid life cycle (Magwire et al., 2012). At the optimum temperature of 25°C, a new generation of fruit flies is produced every 10 days, which (Spieth, 1974). Despite the fact that salmon lice have a longer generation time and thus a longer adaptation period, the present results show that salmon lice also evolve genomic differences on short time scales and on small geographic distances. Factors that might have led to temporal genetic differentiation of salmon lice within an aquaculture facility; (1) bottleneck effect at the establishment of salmon lice, (2) fouling of nets at a site leading to lice retention within the net pens, and thereby influencing the genetic structure, (3) environmental conditions such as temperature and salinity which affected the survival rate and the distribution of salmon lice, and (4) reproductive success among adult salmon lice (David et al., 1997; Larson & Julian, 1999; Nolan & Powell, 2009). A combination of these factors might have led to localized changes in the gene pool of salmon lice at a particular aquaculture facility through production. It was not a part of this thesis to identify the factors important for the observed genetic structuring, but several factors such as genetic drift, selection pressures, and reproductive success must play a central role (Larson & Julian, 1999; Pinceel et al., 2013). Temporal trends in salmon lice larval densities have been observed (Penston et al., 2008), which may have affected the influx of salmon lice larvae with different genetic backgrounds into the aquaculture facility at Skogshamn. Salmon lice with "new genetic

signatures" within each net pen would therefore have been expected to occur over the time of the study. However, the results revealed herein did not support influx of larva with different genetic background, suggesting the presence of a retention mechanisms within the net pens. Further studies are needed to investigate and verify this observation in detail.

## Biology of salmon lice

Salmon lice have a complex life cycle that involves parasitizing multiple hosts, including wild and farmed salmon, and depending on the life stage be transported by water currents or host movements (Asplin et al., 2014). Previous studies suggests that marine organisms display weak population genetic structure due to high dispersal potential, rapid reproduction, and large population sizes (Ward et al., 1994; Waples, 1998). Even though salmon lice are known to be quite site-specific parasites, adult lice are mobile and can move between different attachment sites on the host, in addition to migrating from one host to another, both within and between populations of hosts (Ritchie, 1997). What kind of populations enter the net pens in the aquaculture facility is unknown, as the net pens work as a flow-through system where salmon lice either swim or get transferred by water currents in and out of the aquaculture facility (Morton et al., 2011). In view of an open aquaculture system, the spread of salmon lice among net pens are expected. However, this may lead to genetic exchange among the individuals, hence does not support the results of distinct populations within the net pens that occurred through time.

The four sampling points were distributed over a year, where there was measured a small difference in seawater temperature between the different sampling points (Appendix I). Considering the changes in temperature, the dataset contains approximately four generations of salmon lice from the first to the last timepoint. This means that the population genetic changes observed have evolved fast, from an evolutionary point of view (Nadler, 1995). Temperature and salinity, for instance, are proven to have profound effect on salmon lice survival, development, and reproduction (Johnson & Albright, 1991). In general, higher water temperatures have led to increased reproduction, faster development, and higher survival rates of salmon lice, which can affect the population structure of salmon lice (Stien et al., 2005). Local environmental conditions can therefore lead to selective pressures driving allele frequency differences among populations (Samsing et al., 2016).

A significant genetic difference was observed between TP1 and TP3 based on the important dataset of 261 SNPs. As genetic differences between populations often develop over time through evolutionary processes (Versace et al., 2014), it would in this case be expected a stronger significant differentiation among salmon lice belonging to TP1 and TP4, hence given a longer period between the timepoints. Salmon lice belonging to TP1 were sampled in August (2021) when the sea temperature was 10.4℃, and salmon lice belonging to TP1 were sampled in May when the sea temperature was 4.2℃. As variation in temperature has been shown to have an effect on ectothermic organisms and their expression of genes (Abram et al., 2017; Ugelvik et al., 2022), the sampling time could be a crucial factor contributing to the genetic differences between TP1 and TP3. However, further research is needed to determine whether and how temperature contributes as a driver of local population structure of salmon lice.

## Production implications for salmon lice

Despite a slice treatment that was given at the start of the production to prevent development of chalimus and preadult, no more delousing treatments were carried out during the sampling for this Master´s thesis. Salmon lice have been shown to be highly adaptive organisms, which have led to an evolution of resistance to various drug treatments (Rae, 2002; Aaen et al., 2015). The survival rate after treatment depends on the level of resistance, which has been shown to differentiate in salmon lice populations in Norway (Espedal et al., 2013; Jacobs et al., 2018). A reduction in genetic diversity among the salmon lice could be one potential effect of the feed treatment that was given early in the production, like a bottleneck effect (Nei et al., 1975). A study observed a small genomic region in salmon lice sampled from different countries that were strongly linked to emamectin benzoate resistance, which indicated a rapidly spread of salmon lice across the Atlantic between the period 1999 when the chemical was first introduced, to 2010 when the samples for the project were obtained (Besnier et al., 2014). Within 11 years the mutated genetic material in the salmon louse genome had spread to salmon lice throughout the North Atlantic. Several models of sea louse larvae transportation have been developed and have demonstrated that louse larvae can be transported several kilometers (Asplin et al., 2004; Brooks, 2005). There has been a development of resistance which has resulted in better salmon lice survival (Aaen et al., 2015; Helgesen et al., 2022). Considering that the spread of genes associated with resistance is high, optimization of delousing treatment regime is even more important. Different populations of lice in net pens within the same aquaculture facility may respond differently to delousing treatments due to their genetic background. However, different

responses to delousing treatments would depend on whether the resistant genes are driving the genetic differences between the populations in the net pens, or other genetic drivers.

Regulation of salmon lice in Norwegian aquaculture is based on the existence of a panmictic population of salmon lice and all fish in an aquaculture are treated the same way (Glover et al., 2011). Today, the traffic light system is used as an indicator for the impact of salmon lice on wild salmon in the Norwegian Institute of Marine Research´s (HI) risk assessment (Karlsen et al., 2023). Background data for the traffic light system is based on model simulations that take water currents, dispersion of salmon lice, and the development and behavior of salmon lice into account (Sandvik et al., 2020). Parameters involving modeling of salmon lice genetics and population structures are not in use today, despite previous research showing genetic differentiation among salmon lice within wide-scale geographical studies (Nolan & Powell, 2009; Jacobs et al., 2018). Wide-scale studies showing development of distinct salmon lice populations is also further strengthened as this study shows significant lice populations in net pens located 50-100 meters apart. Overall, research on the genetic composition of salmon lice populations suggests that these parasites are evolving over time, and that adaption to drug treatment or local environmental conditions could be potential drivers for a genetic change (Yue & Wang, 2017). Hopefully, this small-scale genome-wide study can form the basis of the structure of salmon lice population within an aquaculture facility and the knowledge may contribute to an increased understanding that favors population-specific delousing treatment.

# Conclusion

The genome-wide coverage of SNPs used in this project revealed patterns of salmon lice population structuring among net pens within a single aquaculture facility through time. Thereby, the alternative hypothesis of the present study is confirmed, and the null hypothesis discarded. Using the full genome-wide dataset of 66,769 SNPs, DAPC analyses, and $F_{ST}$-values revealed development of weak, but significant, genetic differences in salmon lice within net pens during the start of the production. However, random forest classification approach revealed a clear genetic differentiation between net pens observed later in the production. The application of random forest appears to be a valuable tool to better comprehend the extent of a population structure. Nevertheless, there is a need for future E&R experiments to include supplementary timepoints, extend the sample sizes, and potentially increase the frequency of delousing treatments during the production period. By doing this, we may obtain an improved understanding of the factors and genetic drivers of the population structure among salmon lice in aquaculture.

# Future perspective

Although my goal of the present Master´s thesis was to investigate differentiation among salmon lice populations within an aquaculture facility, and not gender differentiation, there has been discovered a gender pattern in some of the plots used in this master project (Appendix G). Sex-specific population patterns have also been observed in previous studies (Li & Merilä, 2010; Igboeli et al., 2014; Poley et al., 2016), and have suggested that sex determination in salmon lice is likely a combination of genetic and environmental factors, with multiple genes potentially involved. Tolerance against temperatures and delousing chemicals has been shown to be different in females and males (Whyte et al., 2013; Igboeli et al., 2014; Sutherland et al., 2015). However, the specific genes or mechanisms responsible for sex determination in salmon lice are not yet fully understood and further research is needed to examine whether gender-specific salmon lice treatment can be beneficial.

Salmon lice have shown the ability to develop resistance to certain delousing treatments, such as the slice treatment. Another approach for removing salmon lice that have been seen increasing over the recent years is by bathing the salmon in freshwater for a few hours (Thompson et al., 2023). The rise of freshwater treatment has led to concerns that this strategy

may impose a selection pressure strong enough for lice to evolve resistance to freshwater. As freshwater acts as a natural method delousing amongst wild salmon, this type of treatment might therefore have deficient consequences for the wild salmon populations in fjords and rivers. However, the reported correlations connecting the number of freshwater treatments and salmon lice´s tolerance to freshwater are small. Conducting an E&R experiment similar to the present study may exhibit valuable information regarding salmon lice and potential resistance mechanism against freshwater. This could produce worthwhile results of interest to the aquaculture sector, in particular concerning the wellness of wild tribes of salmon and future salmon farming.

# References


Abram, P. K., Boivin, G., Moiroux, J., & Brodeur, J. (2017). Behavioural effects of temperature on ectothermic animals: unifying thermal physiology and behavioural plasticity. *Biological Reviews*, *92*(4), 1859-1876.

Andrews, K. R., Seaborn, T., Egan, J. P., Fagnan, M. W., New, D. D., Chen, Z., Hohenlohe, P. A., Waits, L. P., Caudill, C. C., & Narum, S. R. (2023). Whole genome resequencing identifies local adaptation associated with environmental variation for redband trout. *Molecular Ecology*, *32*(4), 800-818. https://doi.org/10.1111/mec.16810

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence *Babraham Bioinformatics*. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Asplin, L., Boxaspen, K., & Sandvik, A. D. (2004). Modelled distribution of salmon lice in a Norwegian fjord. *Institute of Marine Research, Bergen, Norway*.

Asplin, L., Boxaspen, K. K., & Sandvik, A. D. (2011). Modeling the distribution and abundance of planktonic larval stages of *Lepeophtheirus salmonis* in Norway. *Institute of Marine Research, Bergen, Norway.*, 31-50.

Asplin, L., Johnsen, I. A., Sandvik, A. D., Albretsen, J., Sundfjord, V., Aure, J., & Boxaspen, K. K. (2014). Dispersion of salmon lice in the Hardangerfjord. *Marine Biology Research*, *10*(3), 216-225. https://doi.org/10.1080/17451000.2013.810755

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bentley, D. R. (2006). Whole-genome re-sequencing. *Current opinion in genetics & development*, *16*(6), 545-552. https://doi.org/10.1016/j.gde.2006.10.009

Besnier, F., Kent, M., Skern-Mauritzen, R., Lien, S., Malde, K., Edvardsen, R. B., Taylor, S., Ljungfeldt, L. E., Nilsen, F., & Glover, K. A. (2014). Human-induced evolution caught in action: SNP-array reveals rapid amphi-atlantic spread of pesticide resistance in the salmon ectoparasite *Lepeophtheirus salmonis*. *BMC genomics*, *15*(1), 1-18. https://doi.org/10.1186/1471-2164-15-937

Bjørn, P. A., & Finstad, B. (1998). The development of salmon lice *(Lepeophtheirus salmonis)* on artificially infected post smolts of sea trout (*Salmo trutta*). *Canadian Journal of Zoology*, *76*(5), 970-977. https://doi.org/10.1139/z98-003

Boxaspen, K. (2006). A review of the biology and genetics of sea lice. *ICES Journal of marine Science*, *63*(7), 1304-1316. https://doi.org/10.1016/j.icesjms.2006.04.017

Boxaspen, K., & Næss, T. (2000). Development of eggs and the planktonic stages of salmon lice (*Lepeophtheirus salmonis*) at low temperatures. *Contributions to Zoology*, *69*(1-2), 51-55. https://doi.org/10.1163/18759866-0690102005

Bricknell, I. R., Dalesman, S. J., O'Shea, B., Pert, C. C., & Luntz, A. J. M. (2006). Effect of environmental salinity on sea lice *Lepeophtheirus salmonis* settlement success. *Diseases of aquatic organisms*, *71*(3), 201-212. https://doi.org/10.3354/dao071201

Brieuc, M. S. O., Waters, C. D., Drinan, D. P., & Naish, K. A. (2018). A practical introduction to random forest for genetic association studies in ecology and evolution.

*Molecular ecology resources*, *18*(4), 755-766. https://doi.org/10.1111/1755-0998.12773

Brooks, K. M. (2005). The effects of water temperature, salinity, and currents on the survival and distribution of the infective copepodid stage of sea lice (*Lepeophtheirus salmonis*) originating on Atlantic salmon farms in the Broughton Archipelago of British Columbia, Canada. *Reviews in Fisheries Science*, *13*(3), 177-204. https://doi.org/10.1080/10641260500207109

Bruno, D. W., & Stone, J. (1990). The role of saithe, *Pollachius virens L.*, as a host for the sea lice, *Lepeoptheirus salmonis* Krøyer and *Caligus elongatus* Nordmann. *Aquaculture*, *89*(3-4), 201-207. https://doi.org/10.1016/0044-8486(90)90125-7

Carmichael, S. N., Bekaert, M., Taggart, J. B., Christie, H. R., Bassett, D. I., Bron, J. E., Skuce, P. J., Gharbi, K., Skern-Mauritzen, R., & Sturm, A. (2013). Identification of a sex-linked SNP marker in the salmon louse (*Lepeophtheirus salmonis*) using RAD sequencing. *PLoS One*, *8*(10), e77832.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, *4*, 7. https://doi.org/10.1186/s13742-015-0047-8

Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. *University of California, Berkeley*, *110*(1-12), 24. https://statistics.berkeley.edu/sites/default/files/tech-reports/666.pdf

Chen, J. J. (2010). The Hardy-Weinberg principle and its applications in modern population genetics. *Frontiers in Biology*, *5*(4), 348-353.

Chen, X., & Ishwaran, H. (2012). Random forests for genomic data analysis. *Genomics*, *99*(6), 323-329. https://doi.org/10.1016/j.ygeno.2012.04.003

Chen, Z., Farrell, A. P., Matala, A., Hoffman, N., & Narum, S. R. (2018). Physiological and genomic signatures of evolutionary thermal adaptation in redband trout from extreme climates. *Evolutionary applications*, *11*(9), 1686-1699. https://doi.org/10.1111/eva.12672

Chistiakov, D. A., Hellemans, B., & Volckaert, F. A. (2006). Microsatellites and their genomic distribution, evolution, function and applications: a review with special reference to fish genetics. *Aquaculture*, *255*(1-4), 1-29. https://doi.org/10.1016/j.aquaculture.2005.11.031

Clucas, G., Kerr, L., Cadrin, S., Zemeckis, D., Sherwood, G., Goethel, D., Whitener, Z., & Kovach, A. (2019a). Adaptive genetic variation underlies biocomplexity of Atlantic Cod in the Gulf of Maine and on Georges Bank. *PLoS One*, *14*(5), e0216992.

Clucas, G., Lou, R. N., Therkildsen, N., & Kovach, A. (2019b). Novel signals of adaptive genetic variation in northwestern Atlantic cod revealed by whole-genome sequencing. *Evolutionary applications*, *12*(10), 1971-1987. https://doi.org/10.1111/eva.12861

Colonna, V., Pistis, G., Bomba, L., Mona, S., Matullo, G., Boano, R., Sala, C., Vigano, F., Torroni, A., & Achilli, A. (2013). Small effective population size and genetic homogeneity in the Val Borbera isolate. *European journal of human genetics*, *21*(1), 89-94. https://doi.org/10.1038/ejhg.2012.113

Conte, F. (2004). Stress and the welfare of cultured fish. *Applied Animal Behaviour Science*, *86*(3-4), 205-223. https://doi.org/10.1016/j.applanim.2004.02.003

Dalvin, S., & Oppedal, F. (2019). Utvikling av lakselus ved ulik temperatur og lys *Havforskningsinstituttet*. https://www.hi.no/templates/reporteditor/report-pdf?id=14235&04683779

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., & Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158. https://doi.org/10.1093/bioinformatics/btr330

David, P., Perdieu, M.-A., Pernot, A.-F., & Jarne, P. (1997). Fine-grained spatial and temporal population genetic structure in the marine bivalve *Spisula ovalis*. *Evolution*, 1318-1322. https://doi.org/10.2307/2411061

Dawson, L. H. J. (1998). The physiological effects of salmon lice (Lepeophtheirus salmonis) infections on returning post-smolt sea trout (Salmo trutta L.) in western Ireland, 1996. *ICES Journal of Marine Science*, *55*(2), 193-200. https://doi.org/10.1006/jmsc.1997.0358

Eckert, C. G., Samis, K. E., & Lougheed, S. C. (2008). Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular Ecology*, *17*(5), 1170-1188.

Espedal, P. G., Glover, K. A., Horsberg, T. E., & Nilsen, F. (2013). Emamectin benzoate resistance and fitness in laboratory reared salmon lice (*Lepeophtheirus salmonis*). *Aquaculture*, *416*, 111-118. https://doi.org/10.1016/j.aquaculture.2013.09.001

Etter, P. D., Bassham, S., Hohenlohe, P. A., Johnson, E. A., & Cresko, W. A. (2011). SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Methods Mol Biol*, *772*, 157-178. https://doi.org/10.1007/978-1-61779-228-1_9

Fiskeridirektoratet. (2022). *Akvakulturstatistikk: matfiskproduksjon av laks, regnbueørret og ørret* https://www.fiskeridir.no/Akvakultur/Tall-og-analyse/Akvakulturstatistikk-tidsserier/Laks-regnbueoerret-og-oerret/Matfiskproduksjon

Fiskeridirektoratet. (2023). *Akvakultur*. https://portal.fiskeridir.no/portal/apps/webappviewer/index.html?id=87d862c458774397a8466b148e3dd147

Forskrift om bekjempelse av lakselus i akvakulturanlegg, (2012). https://lovdata.no/dokument/SF/forskrift/2012-12-05-1140/

Glover, K. A., Stølen, Å. B., Messmer, A., Koop, B. F., Torrissen, O., & Nilsen, F. (2011). Population genetic structure of the parasitic copepod *Lepeophtheirus salmonis* throughout the Atlantic. *Marine Ecology Progress Series*, *427*, 161-172. https://doi.org/10.3354/meps09045

Godwin, S. C., Bateman, A. W., Kuparinen, A., Johnson, R., Powell, J., Speck, K., & Hutchings, J. A. (2022). Salmon lice in the Pacific Ocean show evidence of evolved resistance to parasiticide treatment. *Scientific Reports*, *12*(1), 4775. https://doi.org/10.1038/s41598-022-07464-1

Goldstein, D. B., Ruiz Linares, A., Cavalli-Sforza, L. L., & Feldman, M. W. (1995). An evaluation of genetic distances for use with microsatellite loci. *Genetics*, *139*(1), 463-471. https://doi.org/10.1093/genetics/139.1.463

Guragain, P., Båtnes, A. S., Zobolas, J., Olsen, Y., Bones, A. M., & Winge, P. (2022). IIb‐RAD‐sequencing coupled with random forest classification indicates regional

population structuring and sex‐specific differentiation in salmon lice (*Lepeophtheirus salmonis*). *Ecology and Evolution*, *12*(4), e8809. https://doi.org/10.1002/ece3.8809

Hamre, L. A., Bui, S., Oppedal, F., Skern-Mauritzen, R., & Dalvin, S. (2019). Development of the salmon louse *Lepeophtheirus salmonis* parasitic stages in temperatures ranging from 3 to 24 C. *Aquaculture Environment Interactions*, *11*, 429-443. https://doi.org/10.3354/aei00320

Hamre, L. A., Eichner, C., Caipang, C. M. A., Dalvin, S. T., Bron, J. E., Nilsen, F., Boxshall, G., & Skern-Mauritzen, R. (2013). The salmon louse *Lepeophtheirus salmonis* (Copepoda: Caligidae) life cycle has only two chalimus stages. *PLoS One*, *8*(9), e73539. https://doi.org/10.1371/journal.pone.0073539

Harðardóttir, H. M., Male, R., Nilsen, F., & Dalvin, S. (2021). Chitin synthases are critical for reproduction, molting, and digestion in the salmon louse (*Lepeophtheirus salmonis*). *Life*, *11*(1), 47.

Hartl, D. L., Clark, A. G., & Clark, A. G. (1997). *Principles of population genetics* (Vol. 116). Sinauer associates Sunderland.

Hastein, T., Scarfe, A., & Lund, V. (2005). Science-based assessment of welfare: aquatic animals. *Revue Scientifique et Technique-Office International des Epizooties*, *24*(2), 529. https://www.ncbi.nlm.nih.gov/pubmed/16358506

Helgesen, K. O., Horsberg, T. E., Stige, L. C., & Tarpai, A. (2022). The surveillance programme for resistance in salmon lice (*Lepeophtheirus salmonis*) in Norway 2021. *Norwegian Veterinary Institute 2022*, *Report 8/2022*.

Helmstetter, N., Chybowska, A. D., Delaney, C., Da Silva Dantas, A., Gifford, H., Wacker, T., Munro, C., Warris, A., Jones, B., & Cuomo, C. A. (2022). Population genetics and microevolution of clinical *Candida glabrata* reveals recombinant sequence types and hyper-variation within mitochondrial genomes, virulence genes, and drug targets. *Genetics*, *221*(1), iyac031. https://doi.org/10.1093/genetics/iyac031

Heuch, P., Nordhagen, J., & Schram, T. (2000). Egg production in the salmon louse [*Lepeophtheirus salmonis* (Krøyer)] in relation to origin and water temperature. *Aquaculture research*, *31*(11), 805-814. https://doi.org/10.1046/j.1365-2109.2000.00512.x

Heuch, P. A., Knutsen, J. A., Knutsen, H., & Schram, T. A. (2002). Salinity and temperature effects on sea lice over-wintering on sea trout (Salmo trutta) in coastal areas of the Skagerrak. *Journal of the Marine Biological Association of the United Kingdom*, *82*(5), 887-892. https://doi.org/10.1017/S0025315402006306

Heuch, P. A., & Mo, T. A. (2001). A model of salmon louse production in Norway: effects of increasing salmon production and public management measures. *Diseases of aquatic organisms*, *45*(2), 145-152. https://doi.org/10.3354/dao045145

Holliday, J. A., Wang, T., & Aitken, S. (2012). Predicting adaptive phenotypes from multilocus genotypes in Sitka spruce (*Picea sitchensis*) using random forest. *G3: Genes| genomes| genetics*, *2*(9), 1085-1093. https://doi.org/10.1534/g3.112.002733

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nature reviews genetics*, *10*(9), 639-650. https://doi.org/10.1038/nrg2611

Igboeli, O. O., Burka, J. F., & Fast, M. D. (2014). Sea lice population and sex differences in P‐glycoprotein expression and emamectin benzoate resistance on salmon farms in the Bay of Fundy, New Brunswick, Canada. *Pest management science*, *70*(6), 905-914. https://doi.org/10.1002/ps.3620

Iversen, A., Asche, F., Hermansen, Ø., & Nystøyl, R. (2020). Production cost and competitiveness in major salmon farming countries 2003–2018. *Aquaculture*, *522*, 735089. https://doi.org/10.1016/j.aquaculture.2020.735089

Jacobs, A., De Noia, M., Praebel, K., Kanstad-Hanssen, Ø., Paterno, M., Jackson, D., McGinnity, P., Sturm, A., Elmer, K. R., & Llewellyn, M. S. (2018). Genetic fingerprinting of salmon louse (*Lepeophtheirus salmonis*) populations in the North-East Atlantic using a random forest classification approach. *Scientific Reports*, *8*(1), 1203. https://doi.org/10.1038/s41598-018-19323-z

Jansson, E., Besnier, F., Malde, K., André, C., Dahle, G., & Glover, K. A. (2020). Genome wide analysis reveals genetic divergence between Goldsinny wrasse populations. *BMC genetics*, *21*, 1-15.

Johnsen, I. A., Harvey, A., Sævik, P. N., Sandvik, A. D., Ugedal, O., Ådlandsvik, B., Wennevik, V., Glover, K. A., & Karlsen, Ø. (2020). Salmon lice-induced mortality of Atlantic salmon during post-smolt migration in Norway. *ICES Journal of Marine Science*, *78*(1), 142-154. https://doi.org/10.1093/icesjms/fsaa202

Johnsen, I. A., Harvey, A., Sævik, P. N., Sandvik, A. D., Ugedal, O., Ådlandsvik, B., Wennevik, V., Glover, K. A., & Karlsen, Ø. (2021). Salmon lice-induced mortality of Atlantic salmon during post-smolt migration in Norway. *ICES Journal of marine Science*, *78*(1), 142-154. https://doi.org/10.1093/icesjms/fsaa202

Johnson, S., & Albright, L. (1991). Development, growth, and survival of *Lepeophtheirus salmonis* (Copepoda: Caligidae) under laboratory conditions. *Journal of the Marine Biological Association of the United Kingdom*, *71*(2), 425-436. https://doi.org/10.1017/S0025315400051687

Jombart, T., & Ahmed, I. (2011). Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070-3071. https://doi.org/10.1093/bioinformatics/btr521

Karlsen, Ø., Dalvin, S. T., Sandvik, A. D., & Serra Llinares, R. M. (2023). Lakselus–risikovurdering og kunnskapsstatus 2023—Dødelighet og negative effekter hos laksefisk som følge av lakselus. *Havforskningsinstituttet*

Keats, B. J. B., & Sherman, S. L. (2013). Chapter 13 - Population Genetics. In D. Rimoin, R. Pyeritz, & B. Korf (Eds.), *Emery and Rimoin's Principles and Practice of Medical Genetics (Sixth Edition)* (pp. 1-12). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-383834-6.00015-X

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC bioinformatics*, *15*(1), 1-13. https://doi.org/10.1186/s12859-014-0356-4

Larson, A., Wake, D. B., & Yanev, K. P. (1984). Measuring gene flow among populations having high levels of genetic fragmentation. *Genetics*, *106*(2), 293-308.

Larson, R. J., & Julian, R. M. (1999). Spatial and temporal genetic patchiness in marine populations and their implications for fisheries management. *Department of Biology, San Francitco State University*, 94-99.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987-2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*. https://doi.org/10.48550/arXiv.1303.3997

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Subgroup, G. P. D. P. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079. https://doi.org/10.1093/bioinformatics/btp352

Li, M.-H., & Merilä, J. (2010). Sex-specific population structure, natural selection, and linkage disequilibrium in a wild bird population as revealed by genome-wide microsatellite analyses. *BMC evolutionary biology*, *10*, 1-11. https://doi.org/10.1186/1471-2148-10-66

Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R news*, *2*(3), 18-22.

Long, A., Liti, G., Luptak, A., & Tenaillon, O. (2015). Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nature reviews genetics*, *16*(10), 567-582. https://doi.org/10.1038/nrg3937

Lou, R. N., Jacobs, A., Wilder, A. P., & Therkildsen, N. O. (2021). A beginner's guide to low‐coverage whole genome sequencing for population genomics. *Molecular Ecology*, *30*(23), 5966-5993. https://doi.org/10.1111/mec.16077

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature reviews genetics*, *4*(12), 981-994. https://doi.org/10.1038/nrg1226

Magwire, M. M., Fabian, D. K., Schweyen, H., Cao, C., Longdon, B., Bayer, F., & Jiggins, F. M. (2012). Genome-wide association studies reveal a simple genetic basis of resistance to naturally coevolving viruses in *Drosophila melanogaster*. *PLoS genetics*, *8*(11), e1003057.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, *17*(1), 10-12.

Matute, D. R. (2013). The role of founder effects on the evolution of reproductive isolation. *Journal of evolutionary biology*, *26*(11), 2299-2311. https://doi.org/10.1111/jeb.12246

Mayo, O. (2008). A century of Hardy–Weinberg equilibrium. *Twin Research and Human Genetics*, *11*(3), 249-256. https://doi.org/10.1375/twin.11.3.249

Morton, A., Routledge, R., McConnell, A., & Krkošek, M. (2011). Sea lice dispersion and salmon survival in relation to salmon farm activity in the Broughton Archipelago. *ICES Journal of marine Science*, *68*(1), 144-156. https://doi.org/10.1093/icesjms/fsq146

Muneer, A. (2014). Application of microsatellite markers in conservation genetics and fisheries management: recent advances in population structure analysis and

conservation strategies. *Genetics research international*, *2014*. https://doi.org/10.1155/2014/691759

Myksvoll, M., Sandvik, A., Johnsen, I., Skardhamar, J., & Albretsen, J. (2020). The impact of variable physical conditions and future increased aquaculture production on lice infestation pressure and sustainability in Norway. *Aquaculture Environment Interactions*, *12*. https://doi.org/10.3354/aei00359

Nadler, S. A. (1995). Microevolution and the genetic structure of parasite populations. *The Journal of parasitology*, 395-403. https://www.ncbi.nlm.nih.gov/pubmed/7776124

Nei, M., Maruyama, T., & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, 1-10. https://doi.org/10.1111/j.1558-5646.1975.tb00807.x

Nolan, D., Martin, S. A. M., Kelly, Y., Glennon, K., Palmer, R., Smith, T., McCormack, G., & Powell, R. (2000). Development of microsatellite PCR typing methodology for the sea louse, *Lepeophtheirus salmonis* (Krøyer). *Aquaculture research*, *31*(11), 815-822. https://doi.org/10.1046/j.1365-2109.2000.00514.x

Nolan, D. V., & Powell, R. (2009). Geographic and temporal genetic structure in *Lepeophtheirus salmonis* from four salmon farms along the northwest and west coasts of Ireland: results from a microsatellite analysis. *Hydrobiologia*, *617*, 55-63. https://doi.org/10.1007/s10750-008-9525-7

Norwegian Ministry of Trade, I. a. F. (2015). Norwegian salmon farming: predictable and environmentally sustainable growth. *Regjeringen.no*.

NOU. (2019). *Skattlegging av havbruksvirksomhet* (2019:18). https://www.regjeringen.no/contentassets/207ae51e0f6a44b6b65a2cec192105ed/no/pdfs/nou201920190018000dddpdfs.pdf

Olaussen, J. O. (2018). Environmental problems and regulation in the aquaculture industry. Insights from Norway. *Marine Policy*, *98*, 158-163. https://doi.org/10.1016/j.marpol.2018.08.005

Overton, K., Dempster, T., Oppedal, F., Kristiansen, T. S., Gismervik, K., & Stien, L. H. (2019). Salmon lice treatments and salmon mortality in Norwegian aquaculture: a review. *Reviews in Aquaculture*, *11*(4), 1398-1417. https://doi.org/10.1111/raq.12299

Overton, K., Samsing, F., Oppedal, F., Dalvin, S., Stien, L. H., & Dempster, T. (2018). The use and effects of hydrogen peroxide on salmon lice and post-smolt Atlantic salmon. *Aquaculture*, *486*, 246-252.

Penston, M., & Davies, I. (2009). An assessment of salmon farms and wild salmonids as sources of *Lepeophtheirus salmonis* (Krøyer) copepodids in the water column in Loch Torridon, Scotland. *Journal of Fish Diseases*, *32*(1), 75-88. https://doi.org/10.1111/j.1365-2761.2008.00986.x

Penston, M., Millar, C., Zuur, A., & Davies, I. (2008). Spatial and temporal distribution of *Lepeophtheirus salmonis* (Krøyer) larvae in a sea loch containing Atlantic salmon, *Salmo salar L.*, farms on the north‐west coast of Scotland. *Journal of Fish Diseases*, *31*(5), 361-371. https://doi.org/10.1111/j.1365-2761.2008.00915.x

Pinceel, T., Brendonck, L., Larmuseau, M. H., Vanhove, M. P., Timms, B. V., & Vanschoenwinkel, B. (2013). Environmental change as a driver of diversification in temporary aquatic habitats: does the genetic structure of extant fairy shrimp

populations reflect historic aridification? *Freshwater Biology*, *58*(8), 1556-1572. https://doi.org/10.1111/fwb.12137

Poley, J. D., Sutherland, B. J., Jones, S. R., Koop, B. F., & Fast, M. D. (2016). Sex-biased gene expression and sequence conservation in Atlantic and Pacific salmon lice *(Lepeophtheirus salmonis)*. *BMC genomics*, *17*(1), 1-16. https://doi.org/10.1186/s12864-016-2835-7

Rae, G. H. (2002). Sea louse control in Scotland, past and present. *Pest Management Science: formerly Pesticide Science*, *58*(6), 515-520.

Ritchie, G. (1997). The host transfer ability of *Lepeophtheirus salmonis* (Copepoda: Caligidae) from farmed Atlantic salmon, *Salmo salar L*. *Journal of Fish Diseases*, *20*(2), 153-157. https://doi.org/10.1046/j.1365-2761.1997.00285.x

Samsing, F., Oppedal, F., Dalvin, S., Johnsen, I., Vågseth, T., & Dempster, T. (2016). Salmon lice (*Lepeophtheirus salmonis*) development times, body size, and reproductive outputs follow universal models of temperature dependence. *Canadian Journal of Fisheries and Aquatic Sciences*, *73*(12), 1841-1851. https://doi.org/10.1139/cjfas-2016-0050

Sandvik, A. D., Myksvoll, M., & Ådlandsvik, B. (2020). Modellering av smittsomme lakselusslarver. *Havforskningsinstituttet*. https://www.hi.no/hi/nettrapporter/rapport-fra-havforskningen-2020-54

Sommerset, I., Walde, C., Bang Jensen, B., Wiik-Nielsen, J., Bornø, B., Oliveira, V., & Brun, E. (2022). Fiskehelserapporten 2021 *Veterinærinstituttet*.

Spieth, H. T. (1974). Courtship behavior in *Drosophila*. *Annual review of entomology*, *19*(1), 385-405.

Stien, A., Bjørn, P. A., Heuch, P. A., & Elston, D. A. (2005). Population dynamics of salmon lice *Lepeophtheirus salmonis* on Atlantic salmon and sea trout. *Marine Ecology Progress Series*, *290*, 263-275. https://doi.org/10.3354/meps290263

Straume, H.-M., Anderson, J. L., Asche, F., & Gaasland, I. (2020). Delivering the goods: the determinants of Norwegian seafood exports. *Marine Resource Economics*, *35*(1), 83-96. https://doi.org/10.1086/707067

Sutherland, B. J., Poley, J. D., Igboeli, O. O., Jantzen, J. R., Fast, M. D., Koop, B. F., & Jones, S. R. (2015). Transcriptomic responses to emamectin benzoate in Pacific and Atlantic Canada salmon lice *L epeophtheirus salmonis* with differing levels of drug resistance. *Evolutionary applications*, *8*(2), 133-148. https://doi.org/10.1111/eva.12237

Sylvester, E. V., Bentzen, P., Bradbury, I. R., Clément, M., Pearce, J., Horne, J., & Beiko, R. G. (2018). Applications of random forest feature selection for fine‐scale genetic population assignment. *Evolutionary applications*, *11*(2), 153-165. https://doi.org/10.1111/eva.12524

Tautz, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic acids research*, *17*(16), 6463-6471. https://doi.org/10.1093/nar/17.16.6463

ThermoFisherScientific. (2010). *NanoDrop 1000 spectrophotometer V3.8 user's manual* https://tools.thermofisher.com/content/sfs/manuals/nd-1000-v3.8-users-manual-8%205x11.pdf

Thompson, C., Madaro, A., Oppedal, F., Stien, L. H., & Bui, S. (2023). Delousing Efficacy and Physiological Impacts on Atlantic Salmon of Freshwater and Hyposaline Bath Treatments. *Rapport fra havforskningen*.

Thorstad, E. B., Todd, C. D., Uglem, I., Bjørn, P. A., Gargan, P. G., Vollset, K. W., Halttunen, E., Kålås, S., Berg, M., & Finstad, B. (2015). Effects of salmon lice *Lepeophtheirus salmonis* on wild sea trout *Salmo trutta* a literature review. *Aquaculture Environment Interactions*, *7*(2), 91-113. https://doi.org/10.3354/aei00142

Todd, C., Walker, A., Ritchie, M., Graves, J., & Walker, A. (2004). Population genetic differentiation of sea lice (*Lepeophtheirus salmonis*) parasitic on Atlantic and Pacific salmonids: analyses of microsatellite DNA variation among wild and farmed hosts. *Canadian Journal of Fisheries and Aquatic Sciences*, *61*(7), 1176-1190. https://doi.org/10.1139/F04-069

Torrissen, O., Jones, S., Asche, F., Guttormsen, A., Skilbrei, O. T., Nilsen, F., Horsberg, T. E., & Jackson, D. (2013). Salmon lice–impact on wild salmonids and salmon aquaculture. *Journal of Fish Diseases*, *36*(3), 171-194. https://doi.org/10.1111/jfd.12061

Tully, O., & Whelan, K. (1993). Production of nauplii of *Lepeophtheirus salmonis* (Krøyer)(Copepoda: Caligidae) from farmed and wild salmon and its relation to the infestation of wild sea trout (*Salmo trutta* L.) off the west coast of Ireland in 1991. *Fisheries Research*, *17*(1-2), 187-200.

Ugelvik, M. S., Mæhle, S., & Dalvin, S. (2022). Temperature affects settlement success of ectoparasitic salmon lice (*Lepeophtheirus salmonis*) and impacts the immune and stress response of Atlantic salmon (*Salmo salar*). *Journal of Fish Diseases*, *45*(7), 975-990.

Versace, E., Nolte, V., Pandey, R. V., Tobler, R., & Schlötterer, C. (2014). Experimental evolution reveals habitat‐specific fitness dynamics among *Wolbachia clades* in *Drosophila melanogaster*. *Molecular Ecology*, *23*(4), 802-814.

Wade, M. J., & Kalisz, S. (1990 ). The causes of natural selection. *Evolution*, *44*(8), 1947-1955. https://doi.org/10.1111/j.1558-5646.1990.tb04301.x

Walde, C. S., Stormoen, M., Pettersen, J. M., Persson, D., Røsæg, M. V., & Jensen, B. B. (2022). How delousing affects the short-term growth of Atlantic salmon *(Salmo salar)*. *Aquaculture*, *561*, 738720. https://doi.org/10.1016/j.aquaculture.2022.738720

Waples, R. S. (1998). Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, *89*(5), 438-450. https://doi.org/10.1093/jhered/89.5.438wa

Ward, R., Woodwark, M., & Skibinski, D. (1994). A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *Journal of fish biology*, *44*(2), 213-232. https://doi.org/10.1111j.1095-8649.1994.tb01200.x

Whyte, S. K., Westcott, J. D., Elmoslemany, A., Hammell, K. L., & Revie, C. W. (2013). A fixed‐dose approach to conducting emamectin benzoate tolerance assessments on field‐collected sea lice, *Lepeophtheirus salmonis*. *Journal of Fish Diseases*, *36*(3), 283-292. https://doi.org/10.1111/jfd.12055

Wilding, C., Butlin, R., & Grahame, J. (2001). Differential gene exchange between parapatric morphs of Littorina saxatilis detected using AFLP markers. *Journal of evolutionary biology*, *14*(4), 611-619. https://doi.org/10.1046/j.1420-9101.2001.00304.x

Wright, A. F. (2001). Genetic variation: polymorphisms and mutations. *Encyclopedia of Life Sciences*. https://doi.org/10.1038/npg.els.0005005

Yue, G., & Wang, L. (2017). Current status of genome sequencing and its applications in aquaculture. *Aquaculture*, *468*, 337-347. https://doi.org/10.1016/j.aquaculture.2016.10.036

Zimmerman, S. J., Aldridge, C. L., & Oyler-McCance, S. J. (2020). An empirical comparison of population genetic analyses using microsatellite and SNP data for a species of conservation concern. *BMC genomics*, *21*, 1-16. https://doi.org/10.1186/s12864-020-06783-9

Aaen, S. M., Helgesen, K. O., Bakke, M. J., Kaur, K., & Horsberg, T. E. (2015). Drug resistance in sea lice: a threat to salmonid aquaculture. *Trends in parasitology*, *31*(2), 72-81. https://doi.org/10.1016/j.pt.2014.12.006

# Appendix

## Appendix A. Spin-Column Protocol

### Protocol: Purification of Total DNA from Animal Tissues (Spin-Column Protocol)

This protocol is designed for purification of total DNA from animal tissues, including rodent tails.

#### Important points before starting

- If using the DNeasy Blood & Tissue Kit for the first time, read "Important Notes" (page 15).
- For fixed tissues, refer to the pretreatment protocols "Pretreatment for Paraffin Embedded Tissue", page 46, and "Pretreatment for Formalin-Fixed Tissue", page 48.
- All centrifugation steps are carried out at room temperature (15–25°C) in a microcentrifuge.
- Vortexing should be performed by pulse-vortexing for 5–10 s.
- **Optional**: RNase A may be used to digest RNA during the procedure. RNase A is not provided in the DNeasy Blood & Tissue Kit (see "Copurification of RNA", page 20).

#### Things to do before starting

- Buffer ATL and Buffer AL may form precipitates upon storage. If necessary, warm to 56°C until the precipitates have fully dissolved.
- Buffer AW1 and Buffer AW2 are supplied as concentrates. Before using for the first time, add the appropriate amount of ethanol (96–100%) as indicated on the bottle to obtain a working solution.
- Preheat a thermomixer, shaking water bath or rocking platform to 56°C for use in step 2. If using frozen tissue, equilibrate the sample to room temperature (15–25°C).
- Avoid repeated thawing and freezing of samples, because this will lead to reduced DNA size.

## Procedure

1. Cut up to 25 mg tissue (up to 10 mg spleen) into small pieces, and place in a 1.5 ml microcentrifuge tube. For rodent tails, place one (rat) or two (mouse) 0.4–0.6 cm lengths of tail into a 1.5 ml microcentrifuge tube. Add 180 µl Buffer ATL. Earmark the animal appropriately.

   Ensure that the correct amount of starting material is used (see "Starting amounts of samples", page 15). For tissues, such as spleen, with a very high number of cells for a given mass of tissue, no more than 10 mg starting material should be used.

   We strongly recommend cutting the tissue into small pieces to enable more efficient lysis. If desired, lysis time can be reduced by grinding the sample in liquid nitrogen* before addition of Buffer ATL and Proteinase K. Alternatively, tissue samples can be effectively disrupted before Proteinase K digestion using a rotor–stator homogenizer, such as the TissueRuptor II, or a bead mill, such as the TissueLyser II (see ordering information starting on page 59). A supplementary protocol for simultaneous disruption of up to 48 tissue samples using the TissueLyser II can be obtained by contacting QIAGEN Technical Services (see back cover). For rodent tails, a maximum of 1.2 cm (mouse) or 0.6 cm (rat) tail should be used. When purifying DNA from the tail of an adult mouse or rat, it is recommended to use only 0.4–0.6 cm.

2. Add 20 µl Proteinase K. Mix thoroughly by vortexing, and incubate at 56°C until the tissue is completely lysed. Vortex occasionally during incubation to disperse the sample or place in a thermomixer, shaking water bath or on a rocking platform.

   Lysis time varies depending on the type of tissue processed. Lysis is usually complete in 1–3 h or, for rodent tails, 6–8 h. If it is more convenient, samples can be lysed overnight; this will not affect them adversely.

   After incubation the lysate may appear viscous, but should not be gelatinous as it may clog the DNeasy Mini spin column. If the lysate appears very gelatinous, see the "Troubleshooting Guide", page 52, for recommendations.

* When working with chemicals, always wear a suitable lab coat, disposable gloves, and protective goggles. For more information, consult the appropriate safety data sheets (SDSs), available from the product supplier.

**Optional**: If RNA-free genomic DNA is required, add 4 µl RNase A (100 mg/ml), mix by vortexing, and incubate for 2 min at room temperature (15–25°C) before continuing with step 3.

Transcriptionally active tissues, such as liver and kidney, contain high levels of RNA, which will copurify with genomic DNA. For tissues that contain low levels of RNA, such as rodent tails, or, if residual RNA is not a concern, RNase A digestion is not necessary.

3. Vortex for 15 s. Add 200 µl Buffer AL to the sample, and mix thoroughly by vortexing. Then add 200 µl ethanol (96–100%), and mix again thoroughly by vortexing.

   It is essential that the sample, Buffer AL, and ethanol are mixed immediately and thoroughly by vortexing or pipetting to yield a homogeneous solution. Buffer AL and ethanol can be premixed and added together in one step to save time when processing multiple samples.

   A white precipitate may form on addition of Buffer AL and ethanol. This precipitate does not interfere with the DNeasy procedure. Some tissue types (e.g., spleen, lung) may form a gelatinous lysate after addition of Buffer AL and ethanol. In this case, vigorously shaking or vortexing the preparation is recommended.

4. Pipet the mixture from step 3 (including any precipitate) into the DNeasy Mini spin column placed in a 2 ml collection tube (provided). Centrifuge at ≥6000 x *g* (8000 rpm) for 1 min. Discard flow-through and collection tube.*

5. Place the DNeasy Mini spin column in a new 2 ml collection tube (provided), add 500 µl Buffer AW1, and centrifuge for 1 min at ≥6000 x *g* (8000 rpm). Discard flow-through and collection tube.*

* Flow-through contains Buffer AL or Buffer AW1 and is therefore not compatible with bleach. See page 6 for safety information.

6. Place the DNeasy Mini spin column in a new 2 ml collection tube (provided), add 500 µl Buffer AW2, and centrifuge for 3 min at 20,000 x *g* (14,000 rpm) to dry the DNeasy membrane. Discard flow-through and collection tube.

   It is important to dry the membrane of the DNeasy Mini spin column, since residual ethanol may interfere with subsequent reactions. This centrifugation step ensures that no residual ethanol will be carried over during the following elution.

   Following the centrifugation step, remove the DNeasy Mini spin column carefully so that the column does not come into contact with the flow-through, since this will result in carryover of ethanol. If carryover of ethanol occurs, empty the collection tube, then reuse it in another centrifugation for 1 min at 20,000 x *g* (14,000rpm).

7. Place the DNeasy Mini spin column in a clean 1.5 ml or 2 ml microcentrifuge tube (not provided), and pipet 200 µl Buffer AE directly onto the DNeasy membrane. Incubate at room temperature for 1 min, and then centrifuge for 1 min at ≥ 6000 x *g* (8000 rpm) to elute.

   Elution with 100 µl (instead of 200 µl) increases the final DNA concentration in the eluate, but also decreases the overall DNA yield (see Figure 2, page 23).

8. **Recommended**: For maximum DNA yield, repeat elution once as described in step 7.

   This step leads to increased overall DNA yield.

   A new microcentrifuge tube can be used for the second elution step to prevent dilution of the first eluate. Alternatively, to combine the eluates, the microcentrifuge tube from step 7 can be reused for the second elution step.

   **Note**: Do not elute more than 200 µl into a 1.5 ml microcentrifuge tube because the DNeasy Mini spin column will come into contact with the eluate.

## Appendix B. Micro kit Isolation of Genomic DNA

# Protocol: Isolation of Genomic DNA from Tissues

This protocol is for isolation of genomic DNA from less than 10 mg tissue.

**Important points before starting**

- Perform all centrifugation steps at room temperature (15–25°C)
- If isolating DNA from very small amounts of tissue, carrier RNA is required (see page 15).
- Prepare tissue samples on a cold surface (e.g., a glass, steel, or aluminum plate placed on top of a block of dry ice).
- If using frozen tissue, ensure that the sample does not thaw out before addition of Buffer ATL in step 2.

**Things to do before starting**

- Equilibrate Buffer AE or distilled water for elution to room temperature (15–25°C).
- Set a thermomixer or heated orbital incubator to 56°C for use in step 4. If a thermomixer or heated orbital incubator is not available, a heating block or water bath can be used instead.
- If Buffer AL or Buffer ATL contains precipitates, dissolve by heating to 70°C with gentle agitation.
- Ensure that Buffers AW1 and AW2 have been prepared according to the instructions on page 14.

**Procedure**

1. **Transfer a tissue sample of less than 10 mg in weight to a 1.5 ml microcentrifuge tube (not provided).**
2. **Immediately add 180 µl Buffer ATL, and equilibrate to room temperature (15–25°C).**
3. **Add 20 µl proteinase K and mix by pulse-vortexing for 15 s.**
4. **Place the 1.5 ml tube in a thermomixer or heated orbital incubator, and incubate at 56°C overnight until the sample is completely lysed.**

   For small amounts of tissue, lysis is complete in 4–6 h, but best results are achieved after overnight lysis.

Tissues

5. **Add 200 µl Buffer AL, close the lid, and mix by pulse-vortexing for 15 s.**

   To ensure efficient lysis, it is essential that the sample and Buffer AL are thoroughly mixed to yield a homogenous solution.

   **Note:** If carrier RNA is required (see page 13), add 1 µg dissolved carrier RNA to 200 µl Buffer AL. Note that carrier RNA does not dissolve in Buffer AL. It must first be dissolved in Buffer AE and then added to Buffer AL.

6. **Add 200 µl ethanol (96–100%), close the lid, and mix thoroughly by pulse-vortexing for 15 s. Incubate for 5 min at room temperature (15–25°C).**

   **Note:** If room temperature exceeds 25°C, cool the ethanol on ice before adding to the tube.

7. **Briefly centrifuge the 1.5 ml tube to remove drops from inside the lid.**

8. **Carefully transfer the entire lysate from step 7 to the QIAamp MinElute column (in a 2 ml collection tube) without wetting the rim. Close the lid, and centrifuge at 6000 x *g* (8000 rpm) for 1 min. Place the QIAamp MinElute column in a clean 2 ml collection tube, and discard the collection tube containing the flow-through.**

   If the lysate has not completely passed through the membrane after centrifugation, centrifuge again at a higher speed until the QIAamp MinElute column is empty.

9. **Carefully open the QIAamp MinElute column and add 500 µl Buffer AW1 without wetting the rim. Close the lid, and centrifuge at 6000 x *g* (8000 rpm) for 1 min. Place the QIAamp MinElute column in a clean 2 ml collection tube, and discard the collection tube containing the flow-through.**

10. **Carefully open the QIAamp MinElute column and add 500 µl Buffer AW2 without wetting the rim. Close the lid, and centrifuge at 6000 x *g* (8000 rpm) for 1 min. Place the QIAamp MinElute column in a clean 2 ml collection tube, and discard the collection tube containing the flow-through.**

    Contact between the QIAamp MinElute column and the flow-through should be avoided. Some centrifuge rotors may vibrate upon deceleration, resulting in the flow-through, which contains ethanol, coming into contact with the QIAamp MinElute column. Take care when removing the QIAamp MinElute column and collection tube from the rotor, so that flow-through does not come into contact with the QIAamp MinElute column.

11. **Centrifuge at full speed (20,000 x *g*; 14,000 rpm) for 3 min to dry the membrane completely.**

    This step is necessary, since ethanol carryover into the eluate may interfere with some downstream applications.

12. **Place the QIAamp MinElute column in a clean 1.5 ml microcentrifuge tube (not provided) and discard the collection tube containing the flow-through. Carefully open the lid of the QIAamp MinElute column and apply 20–100 µl Buffer AE or distilled water to the center of the membrane.**

    If high pH or EDTA affects sensitive downstream applications, use water for elution (see page 12).

    **Important:** Ensure that Buffer AE or distilled water is equilibrated to room temperature (15–25°C). If using small elution volumes (<50 µl), dispense Buffer AE or distilled water onto the center of the membrane to ensure complete elution of bound DNA.

    QIAamp MinElute columns provide flexibility in the choice of elution volume. Choose a volume according to the requirements of the downstream application. Remember that the volume of eluate will be up to 5 µl less than the volume of the solution applied to the column.

13. **Close the lid and incubate at room temperature (15–25°C) for 1 min. Centrifuge at full speed (20,000 x *g*; 14,000 rpm) for 1 min.**

    Incubating the QIAamp MinElute column loaded with Buffer AE or water for 5 min at room temperature before centrifugation generally increases DNA yield.

## Appendix C. Gel electrophoresis results

The DNA extracted for this project was mainly high molecular DNA, with few individuals with small amount of degraded DNA.

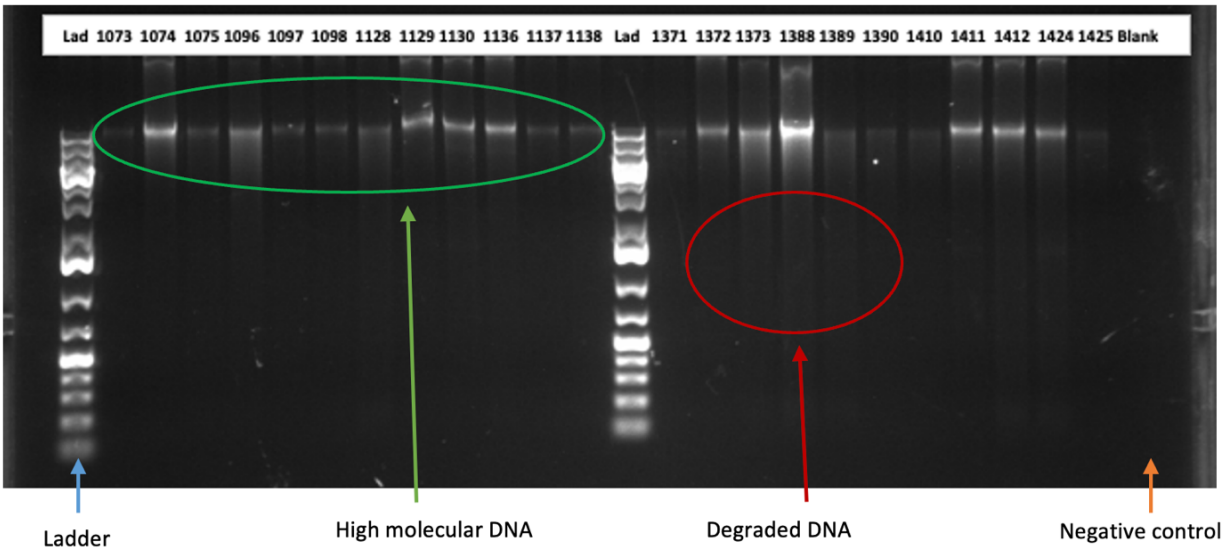



*Figure 18. The gel electrophoresis included 2 ladders (channels with multiple bands) and a negative control (blank with no DNA).*

# Appendix D. Library Prep Kit for Illumina

## Section 1
## Protocol for use with Inputs ≤ 100 ng

**Symbols**

*This is a point where you can safely stop the protocol.*

*This caution sign signifies a step in the protocol that has multiple paths leading to the same end point but is dependent on a user variable, like the amount of input DNA.*

*Colored bullets indicate the cap color of the reagent to be added to a reaction.*

**Note: Follow the protocol in this chapter for inputs ≤ 100 ng, as size selection is not recommended for this input range. Follow the protocol in Chapter 2 for inputs ≥ 100 ng, as size selection is recommended for this input range. Follow the protocol in Chapter 3 for inputs ≥ 100 ng and fragment sizes > 550 bp. For 100 ng inputs, either the no size selection protocol (Chapter 1) or a size selection protocol (Chapter 2 or 3) can be followed.**

**Starting Material:** 100 pg–100 ng purified, genomic DNA. We recommend that the DNA be in 1X TE (10 mM Tris pH 8.0, 1 mM EDTA), however, 10 mM Tris pH 7.5–8, low EDTA TE or $H_2O$ are also acceptable. If the input DNA is less than 26 µl, add TE (provided) to a final volume of 26 µl.

### 1.1. Fragmentation/End Prep

Fragmentation occurs during the 37°C incubation step. Use the chart below to determine the incubation time required to generate the desired fragment sizes. Incubation time may need to be optimized for individual samples. See Figure 1.1 for a typical fragmentation pattern.

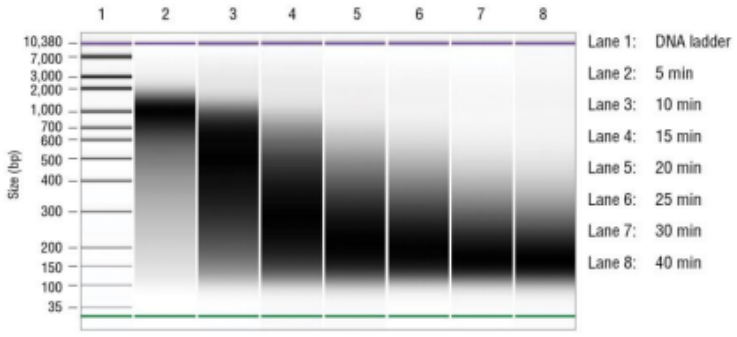
| FRAGMENTATION SIZE | INCUBATION @ 37°C | OPTIMIZATION |
|---|---|---|
| 100 bp–250 bp | 30 min | 30–40 min |
| 150 bp–350 bp | 20 min | 20–30 min |
| 200 bp–450 bp | 15 min | 15–20 min |
| 300 bp–700 bp | 10 min | 5–15 min |
| 500 bp–1 kb | 5 min | 5–10 min |

1.1.1. Ensure that the Ultra II FS Reaction Buffer is completely thawed. If a precipitate is seen in the buffer, pipette up and down several times to break it up, and quickly vortex to mix. Place on ice until use.

1.1.2. Vortex the Ultra II FS Enzyme Mix 5-8 seconds prior to use and place on ice.

**Note: It is important to vortex the enzyme mix prior to use for optimal performance.**

**Figure 1.1: Example of size distribution on a Bioanalyzer®. Human DNA (NA19240) was fragmented for 5-40 min.**



1.1.3. Add the following components to a 0.2 ml thin wall PCR tube on ice:

| COMPONENT | VOLUME PER ONE LIBRARY |
|---|---|
| DNA | 26 µl |
| ● (yellow) NEBNext Ultra II FS Reaction Buffer | 7 µl |
| ● (yellow) NEBNext Ultra II FS Enzyme Mix | 2 µl |
| Total Volume | 35 µl |

1.1.4. Vortex the reaction for 5 seconds and briefly spin in a microcentrifuge.

1.1.5. In a thermal cycler, with the heated lid set to 75°C, run the following program:

**5–30 min @ 37°C**
**30 min @ 65°C**
**Hold @ 4°C**

**If necessary, samples can be stored at –20°C; however, a slight loss in yield (~20%) may be observed. We recommend continuing with adaptor ligation before stopping.**

**1.2. Adaptor Ligation**

Determine whether adaptor dilution is necessary.

*If DNA input is < 100 ng, dilute the ● (red) NEBNext Adaptor for Illumina in 10 mM Tris-HCl, pH 7.5-8.0 with 10 mM NaCl as indicated in Table 1.2.1.*

**Table 1.2.1: Adaptor Dilution**

| INPUT | ADAPTOR DILUTION (VOLUME OF ADAPTOR: TOTAL VOLUME) | WORKING ADAPTOR CONCENTRATION |
|---|---|---|
| 100 ng–500 ng | No Dilution | 15 µM |
| 5 ng–99 ng | 10-Fold (1:10) | 1.5 µM |
| less than 5 ng | 25-Fold (1:25) | 0.6 µM |

**Note: The appropriate adaptor dilution for your sample input and type may need to be optimized experimentally. The dilutions provided here are a general starting point.**

1.2.1. Add the following components directly to the FS Reaction Mixture:

| COMPONENT | VOLUME |
|---|---|
| FS Reaction Mixture (Step 1.1.5) | 35 µl |
| ● (red) NEBNext Ultra II Ligation Master Mix* | 30 µl |
| ● (red) NEBNext Ligation Enhancer | 1 µl |
| ● (red) NEBNext Adaptor for Illumina** | 2.5 µl |
| Total Volume | 68.5 µl |

* Mix the Ultra II Ligation Master Mix by pipetting up and down several times prior to adding to the reaction.

** The NEBNext adaptor is provided in NEBNext Oligo kits. NEB has several oligo options which are supplied separately from the library prep kit. Please see www.neb.com/oligos for additional information.

**Note: The Ligation Master Mix and Ligation Enhancer can be mixed ahead of time and is stable for at least 8 hours @ 4°C. Do not premix the Ligation Master Mix, Ligation Enhancer and adaptor prior to use in the Adaptor Ligation Step.**

1.2.2. Set a 100 µl or 200 µl pipette to 50 µl and then pipette the entire volume up and down at least 10 times to mix thoroughly. Perform a quick spin to collect all liquid from the sides of the tube. **(Caution: The NEBNext Ultra II Ligation Master Mix is very viscous. Care should be taken to ensure adequate mixing of the ligation reaction, as incomplete mixing will result in reduced ligation efficiency. The presence of a small amount of bubbles will not interfere with performance).**

1.2.3. Incubate at 20°C for 15 minutes in a thermal cycler with the heated lid off.

1.2.4. Add 3 µl of ● (red or blue) USER® Enzyme to the ligation mixture from Step 1.2.3.

**Note: Steps 1.2.4. and 1.2.5. are only required for use with non indexed NEBNext Adaptor. USER enzyme can be found in most NEBNext oligo kits. If you are using the indexed UMI adaptor, USER is not needed. Please see corresponding manual for use with UMI on the NEB #E7395 product page under the protocols, manuals, and usage tab.**

1.2.5. Mix well and incubate at 37°C for 15 minutes with the heated lid set to ≥ 47°C.

**Samples can be stored overnight at –20°C.**

### 1.3. Size Selection or Cleanup of Adaptor-ligated DNA

**The following section is for cleanup of the ligation reaction for inputs ≤ 100 ng. If your input DNA is > 100 ng, follow the size selection protocol in Chapter 2, Section 2.3. If you want fragment sizes > 550 bp and your input is ≥ 100 ng, follow the entire protocol in Chapter 3.**

**Note: The volumes of SPRIselect or NEBNext Sample Purification Beads provided here are for use with the sample contained in the exact buffer at this step (71.5 µl; Step 1.2.5.). AMPure XP Beads can be used as well. If using AMPure XP Beads, allow the beads to warm to room temperature for at least 30 minutes before use. These bead volumes may not work properly for a cleanup at a different step in the workflow, or if this is a second cleanup at this step. For cleanups of samples contained in different buffer conditions, the volumes may need to be experimentally determined.**

1.3.1. Vortex SPRIselect or NEBNext Sample Purification Beads to resuspend.

1.3.2. Add 57 µl (0.8X) resuspended beads to the Adaptor Ligation reaction. Mix well by pipetting up and down at least 10 times. Be careful to expel all of the liquid out of the tip during the last mix. Vortexing for 3-5 seconds on high can also be used. If centrifuging samples after mixing, be sure to stop the centrifugation before the beads start to settle out.

1.3.3. Incubate samples at room temperature for at least 5 minutes.

1.3.4. Place the tube/plate on an appropriate magnetic stand to separate the beads from the supernatant. If necessary, quickly spin the sample to collect the liquid from the sides of the tube or plate wells before placing on the magnetic stand.

1.3.5. After 5 minutes (or when the solution is clear), carefully remove and discard the supernatant. Be careful not to disturb the beads that contain DNA targets **(Caution: do not discard the beads)**.

1.3.6. Add 200 µl of 80% freshly prepared ethanol to the tube/plate while in the magnetic stand. Incubate at room temperature for 30 seconds, and then carefully remove and discard the supernatant. Be careful not to disturb the beads that contain DNA targets.

1.3.7. Repeat Step 1.3.6. once for a total of two washes. Be sure to remove all visible liquid after the second wash. If necessary, briefly spin the tube/plate, place back on the magnet and remove traces of ethanol with a p10 pipette tip.

1.3.8. Air dry the beads for up to 5 minutes while the tube/plate is on the magnetic stand with the lid open.

**Caution: Do not over-dry the beads. This may result in lower recovery of DNA. Elute the samples when the beads are still dark brown and glossy looking, but when all visible liquid has evaporated. When the beads turn lighter brown and start to crack, they are too dry.**

1.3.9. Remove the tube/plate from the magnetic stand. Elute the DNA target from the beads by adding 17 µl 0.1X TE (dilute 1X TE Buffer 1:10 in water).

1.3.10. Mix well by pipetting up and down 10 times, or on a vortex mixer. Incubate for at least 2 minutes at room temperature. If necessary, quickly spin the sample to collect the liquid from the sides of the tube or plate wells before placing back on the magnetic stand.

1.3.11. Place the tube/plate on the magnetic stand. After 5 minutes (or when the solution is clear), transfer 15 µl to a new PCR tube.

1.3.12. Proceed to PCR Enrichment of Adaptor-ligated DNA in Section 1.4.

**Samples can be stored at –20°C.**

### 1.4. PCR Enrichment of Adaptor-ligated DNA

Follow Step 1.4.1A. if you are using the following oligos:
Use Option A for any NEBNext Oligo Kit where index primers are supplied in tubes. These kits have the forward and reverse primers supplied in separate tubes. Primers are supplied at 10 µM each.

Follow Step 1.4.1B. if you are using the following oligos:
Use Option B for any NEBNext Oligo Kit where index primers are supplied in a 96-well plate format. These kits have the forward and reverse (i7 and i5) primers combined. Primers are supplied at 10 µM combined (5 µM each).

1.4.1. Add the following components to a sterile strip tube:

#### 1.4.1A. Forward and Reverse Primers not already combined

| COMPONENT | VOLUME |
|---|---|
| Adaptor Ligated DNA Fragments (Step 1.3.12.) | 15 µl |
| • (blue) NEBNext Ultra II Q5 Master Mix | 25 µl |
| • (blue) Index Primer/i7 Primer*,** | 5 µl |
| • (blue) Universal PCR Primer/i5 Primer*,** | 5 µl |
| Total Volume | 50 µl |

* NEBNext Oligos must be purchased separately from the library prep kit. For oligo purchasing options refer to "Required Materials Not Included" section (page 1). Refer to the corresponding NEBNext Oligo kit manual for determining valid barcode combinations.

** Use only one i7 primer/ index primer per sample. Use only one i5 primer (or the universal primer for single index kits) per sample

#### 1.4.1B. Forward and Reverse Primers already combined

| COMPONENT | VOLUME |
|---|---|
| Adaptor Ligated DNA Fragments (Step 1.3.12.) | 15 µl |
| • (blue) NEBNext Ultra II Q5 Master Mix | 25 µl |
| • (blue) Index Primer Mix* | 10 µl |
| Total Volume | 50 µl |

* NEBNext Oligos must be purchased separately from the library prep kit. For oligo purchasing options refer to "Required Materials Not Included" section (page 1).

1.4.2. Set a 100 µl or 200 µl pipette to 40 µl and then pipette the entire volume up and down at least 10 times to mix thoroughly. Perform a quick spin to collect all liquid from the sides of the tube.

1.4.3. Place the tube on a thermal cycler with the heated lid set to 105°C and perform PCR amplification using the following PCR cycling conditions:

| CYCLE STEP | TEMP | TIME | CYCLES |
|---|---|---|---|
| Initial Denaturation | 98°C | 30 seconds | 1 |
| Denaturation | 98°C | 10 seconds | 3-13* |
| Annealing/Extension | 65°C | 75 seconds | |
| Final Extension | 65°C | 5 minutes | 1 |
| Hold | 4°C | ∞ | |

* The number of PCR cycles recommended in Table 1.4.1 are to be seen as a starting point to determine the number of PCR cycles best for standard library prep samples. Use Table 1.4.2 for applications requiring high library yields, such as target enrichment. The number of PCR cycles should be chosen based on input amount and sample type. Thus, samples prepared with a different method prior to library prep may require re-optimization of the number of PCR cycles. The number of cycles should be high enough to provide sufficient library fragments for a successful sequencing run, but low enough to avoid PCR artifacts and over-cycling (high molecular weight fragments on Bioanalyzer).

**1.5. Cleanup of PCR Reaction**

**Note: The volumes of SPRIselect or NEBNext Sample Purification Beads provided here are for use with the sample contained in the exact buffer at this step. AMPure XP beads can be used as well. If using AMPure XP beads, allow the beads to warm to room temperature for at least 30 minutes before use. These volumes may not work properly for a cleanup at a different step in the workflow. For cleanups of samples contained in different buffer conditions, the volumes may need to be experimentally determined.**

1.5.1. Vortex SPRIselect or NEBNext Sample Purification Beads to resuspend.

1.5.2. Add 45 µl (0.9X) resuspended beads to the PCR reaction. Mix well by pipetting up and down at least 10 times. Be careful to expel all of the liquid out of the tip during the last mix. Vortexing for 3-5 seconds on high can also be used. If centrifuging samples after mixing, be sure to stop the centrifugation before the beads start to settle out.

1.5.3. Incubate samples on bench top for at least 5 minutes at room temperature.

1.5.4. Place the tube/plate on an appropriate magnetic stand to separate the beads from the supernatant. If necessary, quickly spin the sample to collect the liquid from the sides of the tube or plate wells before placing on the magnetic stand.

1.5.5. After 5 minutes (or when the solution is clear), carefully remove and discard the supernatant. Be careful not to disturb the beads that contain DNA targets **(Caution: do not discard the beads)**.

1.5.6. Add 200 µl of 80% freshly prepared ethanol to the tube/plate while in the magnetic stand. Incubate at room temperature for 30 seconds, and then carefully remove and discard the supernatant. Be careful not to disturb the beads that contain DNA targets.

1.5.7. Repeat Step 1.5.6. once for a total of two washes. Be sure to remove all visible liquid after the second wash. If necessary, briefly spin the tube/plate, place back on the magnet and remove traces of ethanol with a p10 pipette tip.

1.5.8. Air dry the beads for up to 5 minutes while the tube/plate is on the magnetic stand with the lid open.

**Caution: Do not over-dry the beads. This may result in lower recovery of DNA. Elute the samples when the beads are still dark brown and glossy looking, but when all visible liquid has evaporated. When the beads turn lighter brown and start to crack, they are too dry.**

1.5.9. Remove the tube/plate from the magnetic stand. Elute the DNA target from the beads by adding 33 µl of 0.1X TE (dilute 1X TE Buffer 1:10 in water).

1.5.10. Mix well by pipetting up and down 10 times, or on a vortex mixer. Incubate for at least 2 minutes at room temperature. If necessary, quickly spin the sample to collect the liquid from the sides of the tube or plate wells before placing back on the magnetic stand.

1.5.11. Place the tube/plate on the magnetic stand. After 5 minutes (or when the solution is clear), transfer 30 µl to a new PCR tube and store at –20°C.

# Appendix E. R-scripts of PCA, DAPC and Heatmap with pairwise $F_{ST}$

PCA-script

```r
#load tidyverse
library(tidyverse)
### read individual-and-population information
pop <- read.table("TP2_pop.txt", header = F)

##### only read sencond column of the pop object
pop <-as.factor(pop$V2)

### read eigenvec file generated from plink run
pca <- read_table2("TP2. eigenvec ny.txt", col_names = FALSE)

#### read eigenvec file generated from plink run
eigenval <- scan("TP2.eigenval")

## remove first unnecessary column.
pca <- pca[,-1]

#Hvordan få de ulike individuell og pop.txt til å matche
write.csv(pca, "PC.csv", row.names = FALSE,quote = FALSE)

###########
#combine with our population assignments and remake pca object
pca <-as_tibble(cbind(pop, pca)) #combine with our population assignments

## first convert to percentage variance explained
pve <- data.frame(PC = 1:20, pve = eigenval/sum(eigenval)*100)

## make plot using ggplot
a <- ggplot(pve, aes(PC, pve)) + geom_bar(stat = "identity")
a + ylab("Percentage variance explained") + theme_light()

# calculate the cumulative sum of the percentage variance explained
cumsum(pve$pve)

#plot pca
b <- ggplot(pca, aes(X3, X4, col = pop, shape = pop)) +
     geom_point(size = 3) +
     scale_shape_manual(values=c(0, 1,2,3, 4,5,6,7,8,9,10,11,12))
b <- b + scale_colour_manual(values = c("red","black","green3","blue","cyan","magenta","yellow","gray","firebrick","peru","wheat"
b <- b + coord_equal() + theme_light()
b <- b + coord_fixed(ylim = c(-0.5,0.75), xlim = c(-0.5, 0.6)) + theme_light()
b + xlab(paste0("PC1 (", signif(pve$pve[1], 3), "%)")) + ylab(paste0("PC2 (", signif(pve$pve[2], 3), "%)"))
ggsave(filename = "./pca_plot_1percent.pdf", plot = pca) #change file path if data on your own computer


#Hvordan få de ulike individuell og pop.txt til å matche
write.csv(pca, "PC.csv", row.names = FALSE,quote = FALSE)


#change axis scales in Base R
#create scatterplot with custom axes
(data, aes(x=pca, y=pve)) +
  geom_point() +
  xlim(-0,3, 0,6) +
  ylim(-0,3, 0,75)
```

## DAPC-script

```
library(adegenet)
library(parallel)
# do not use: library(vcfR)
getwd()

Data <- read.structure("139.netpen.str",NA.char = "-9" )
levels(Data$pop)=c("1_11","1_12","1_15","2_11","2_12","2_15","3_11","3_12","3_15","4_11","4_12","4_15")
summary(Data$pop) # give sample size per population

###### xvalidation
mat <- as.matrix(tab(Data, NA.method="mean")) #simulates missing data with mean of the population mean
grp <- pop(Data)
grp
xval <- xvalDapc(mat,grp,parallel = "multicore",ncpus= "8") #(run it 3 times - mean av de tre gangene).
xval[2:6] # choose PCA with least SSE

#####find.cluster: denvo way to predict the clusters represented in the population
grp <- find.clusters(Data)
grp

####### dapc plotting after find.cluster (without priors)
dapc1 <- dapc(Data, grp$grp) #running dapc analysis.
scatter (dapc1,posi.da="bottomright") #making scatter plot for dapc object
dapc1
grp$grp
table.value(table(pop(Data), grp$grp), col.lab=paste("inf", 1:4),row.labels = paste("ori", 1:4)) #assignment plot

####dapc with priors (ex populations, netpens or time points)

dapc2 <- dapc(Data)
mycol=c("darkorange2", "forestgreen", "blue", "black", "deepskyblue1", "darkorchid2","darkcyan","lightsalmon2","brown","grey57","magenta","red")
#mycol=funky(12)
#without ellipse
scatter(dapc2, posi.leg = "bottomright", cell=0, cstar=0, scree.da=FALSE, clab=0, cex=1,5, solid=0,8, bg="white", leg=TRUE, col=mycol)
#with ellipse
assignplot(dapc2)
scatter (dapc2,col = mycol, posi.da="bottomright")
compoplot(dapc2) #Membership
compoplot(dapc2, col.pal = mycol)

percent= dapc2$eig/sum(dapc2$eig)*100 #percentage eigenvalues.
```

## Heatmap-script with pairwise $F_{ST}$

```
if (!require("BiocManager", quietly = TRUE))
install.packages("BiocManager")
BiocManager::install("Heatplus")
getwd()

library("Heatplus")
library(gplots)

### Dowload fst matrix (without NA)
fst= read.table("hm_netpen.txt", header= F, dec=".", sep="\t")
dim(fst)
colnames(fst)= c("1_11","1_12","1_15","2_11","2_12","2_15","3_11","3_12","3_15","4_11","4_12","4_15")
rownames(fst)= c("1_11","1_12","1_15","2_11","2_12","2_15","3_11","3_12","3_15","4_11","4_12","4_15")
#rownames(dlr_fst1)= c("ANT","BOO","BRA","BRO","BON","BUZ","CAN","CAR","GAS","LOB","MAG","MAL","MAR","OFF","RHO","SEA","SID","SJI","TRI")
fst

### Do an heatmap
#pdf("heatmap_fst_neutral.pdf")
heatmap.2(as.matrix(fst), breaks = c(0.000,0.020, 0.04, 0.06,0.08),col= c("yellow", "orange", "red", "darkred"), dendrogram = "none", scale="none",
          key=F, keysize=1,density.info="none",trace="none", labRow=c("1_11","1_12","1_15","2_11","2_12","2_15","3_11","3_12","3_15","4_11","4_12","4_15"), Rowv=F, Colv=F)
legend("bottom",fill = c("yellow", "orange", "red", "darkred"), inset=-0.0, legend = c("<0.020", "0.021-0.040", "0.041-0.060", "0.061-0.080"))
#dev.off()
```

## Appendix F. PCA from TP1, TP2, TP3 and TP4

PCA has been a tool to bring out patterns from biological datasets by using a few principal components. The PCA plot shows variation in allele frequencies explained by the first and the second principal components (PC1 and PC2), which are the two components that convey the most variation in the dataset. TP1 and TP3 show genetic population structure similarities with only one cluster, with three deviant individuals (Figure 19). TP2 and TP4 show genetic population structure similarities with two separate clusters, with a random distribution of individuals from the different net pens.

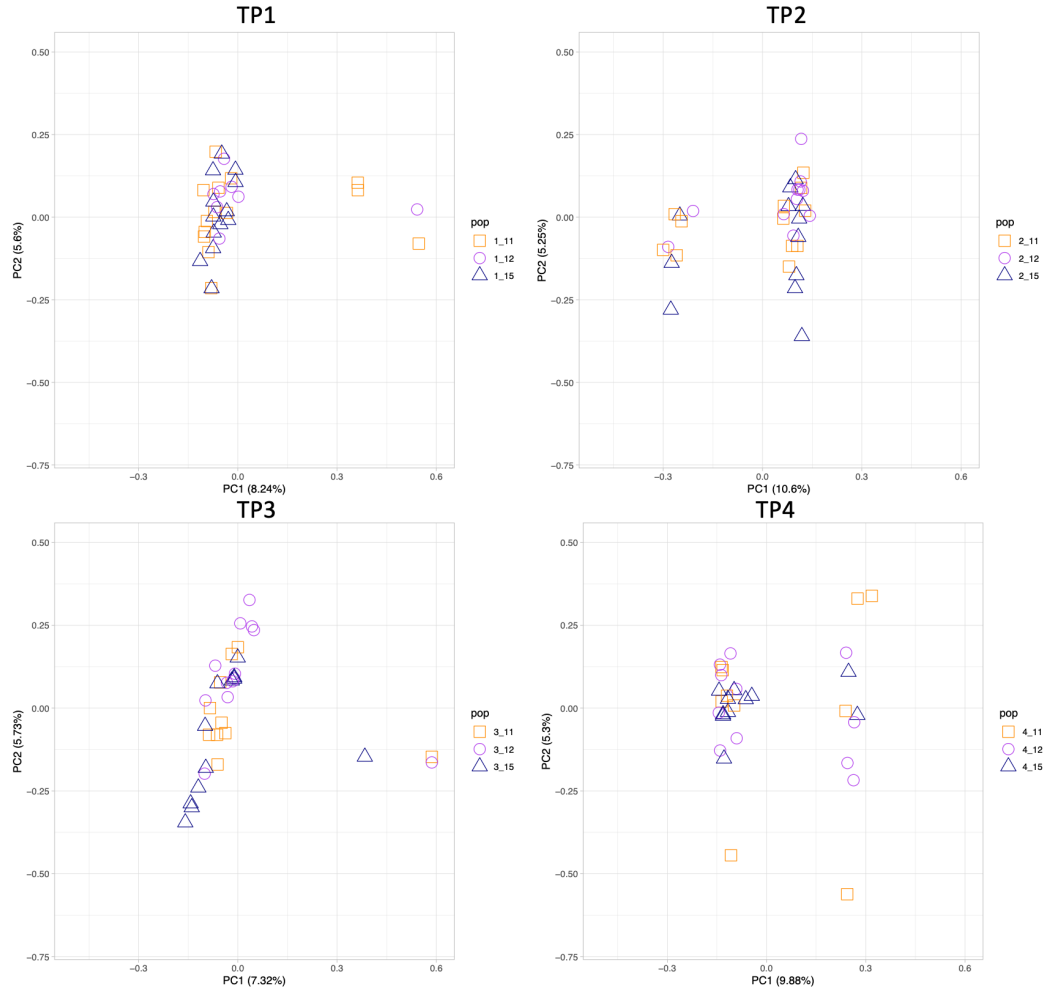



*Figure 19. PCA plots representing timepoints with associated net pens, based on the full dataset of 66,769 SNPs. The markers are represented as the combination of timepoint and net pen number (example: 1_12, where 1 represents TP1 and 12 represents net pen 12). Component PC1 and PC2 explained 14.01% variation in TP1, 15.43% variation in TP2, 12.81% variation in TP3 and 15.69% variation in TP4.*

## Appendix G. Sex differentiation

Pairwise $F_{ST}$ between females and males were found to be 0.01. The PCA plot based on the full dataset of 66,769 SNPs shows two clusters (Figure 20). One cluster represent only males except two female individuals. The other cluster represent a mixture of males and females, with a clear distinction between females and males within the same cluster.

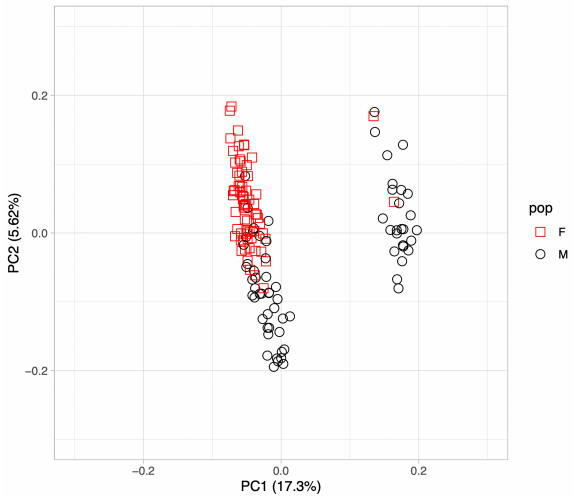



*Figure 20. PCA plot representing female and male lice used in this project, based on the full dataset of 66,769 SNPs. Females are represented by red squares and males are represented by black circles.*

Dot chart representing female and male distribution within the timepoints based on the full dataset of 66,769 SNPs, performed in Excel (Figure 21). TP2 show one cluster with mainly males, except two female individuals. TP4 show one cluster with only male individuals.

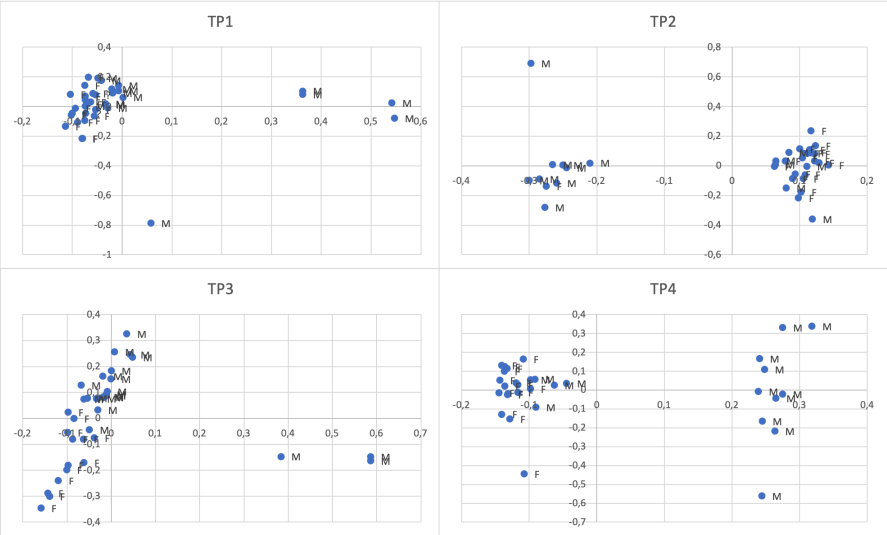



*Figure 21. Dot chart representing female and male distribution within the timepoints, based on the full dataset of 66,769 SNPs. Every dot represents an individual of salmon louse represented by either M (male) or F (female).*

## Appendix H. $F_{ST}$ results by using the full dataset

Heatmap representing $F_{ST}$ between the net pens and the timepoints based on the full dataset of 66,769 SNPs, performed in R. Significant $F_{ST}$ values represented by * and indicated P-values $\leq 0.05$. The pairwise $F_{ST}$ among the timepoints were significant (Table 3).

*Table 2. Pairwise* $F_{ST}$ *between net pens, based on the full dataset of 66,769 SNPs. The net pens are represented as the combination of timepoint and net pen number (example: 1_12, where 1 represents TP1 and 12 represents net pen 12). Fst values with * indicates that the P-values are* $\leq 0.05$.

| | 1_11 | 1_12 | 1_15 | 2_11 | 2_12 | 2_15 | 3_11 | 3_12 | 3_15 | 4_11 | 4_12 | 4_15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1_11 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1_12 | -0,0010 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 1_15 | 0,0008* | -0,0002 | NA | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 2_11 | 0,0003* | -0,0003 | 0,0021* | NA | NA | NA | NA | NA | NA | NA | NA | NA |
| 2_12 | -0,0002 | -0,0011 | 0,0011* | -0,0004 | NA | NA | NA | NA | NA | NA | NA | NA |
| 2_15 | 0,0004* | -0,0007 | 0,0009* | -0,0004 | -3E-04 | NA | NA | NA | NA | NA | NA | NA |
| 3_11 | 0,0016* | 0,0001* | 0,0015* | 0,0004* | 5E-04* | -1E-04 | NA | NA | NA | NA | NA | NA |
| 3_12 | 0,0027* | 0,0008* | 0,0021* | 0,0013* | 1E-03* | 9E-04* | -5E-05 | NA | NA | NA | NA | NA |
| 3_15 | 0,0008* | -0,0003 | 0,0005* | 0,0009* | 8E-04* | 2E-05* | 4E-04* | 0,0009* | NA | NA | NA | NA |
| 4_11 | 0,0022* | 0,0009* | 0,0038* | -0,0003 | 8E-04* | 3E-04* | 1E-03* | 0,0025* | 2E-03* | NA | NA | NA |
| 4_12 | 0,0010* | 0,0002* | 0,0029* | -0,0007 | 2E-05* | 6E-05* | 5E-04* | 0,0007* | 1E-03* | -0,0005 | NA | NA |
| 4_15 | 0,0008* | -0,0005 | 0,0013* | -0,0004 | -3E-04 | -8E-04 | -2E-04 | 0,0003* | -2E-05 | -0,0003 | -0,0004 | NA |

*Table 3. Pairwise* $F_{ST}$ *between timepoints, based on the full dataset of 66,769 SNPs.* $F_{ST}$ *values with * indicates that the P-values are* $\leq 0.05$.

| | TP1 | TP2 | TP3 | TP4 |
|---|---|---|---|---|
| TP1 | NA | NA | NA | NA |
| TP2 | 0,0005* | NA | NA | NA |
| TP3 | 0,0011* | 0,0007* | NA | NA |
| TP4 | 0,0016* | 0,0001* | 0,0008* | NA |

# Appendix I. Seawater temperature

Seawater temperature was measured at the Skogshamn aquaculture facility from every sampling point (TP1, TP2, TP3 and TP4) (Table 4).

*Table 4. Average seawater temperature measurements from TP1, TP2, TP3 and TP4.*

| **Timepoint** | **Week** | **Average seawater temperature (°C)** |
|---|---|---|
| TP1 | 35 – 41 (2021) | 10.4 |
| TP2 | 3+4 (2022) | 4.15 |
| TP3 | 19+20 (2022) | 4.2 |
| TP4 | 32+33 (2022) | 12.7 |