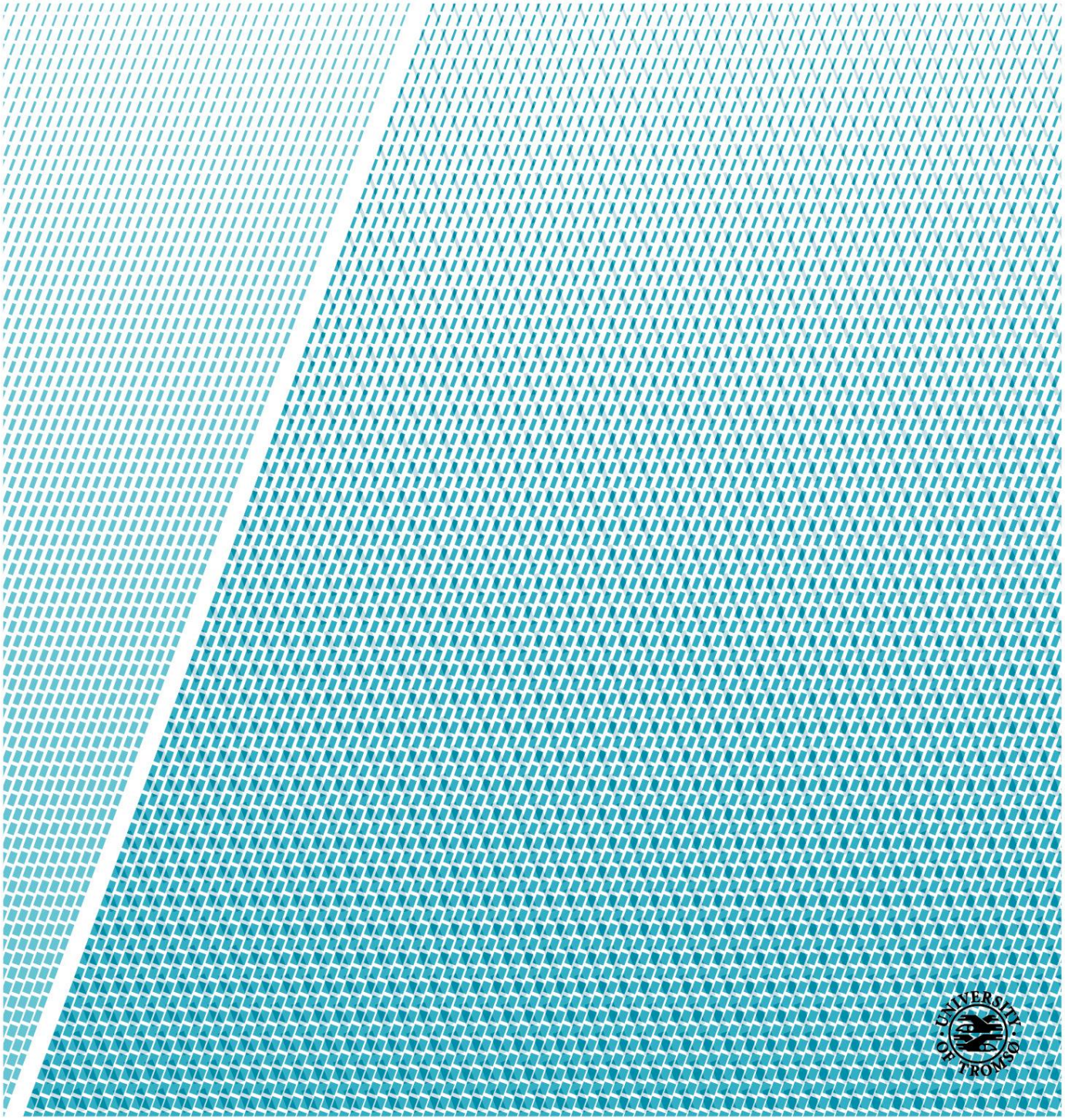
UiT
THE ARCTIC UNIVERSITY OF NORWAY

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

Investigating population genomic structure of salmon lice (*Lepeophtheirus salmonis*) in northern Norway

Kana Banno

BIO-3950 Master's thesis in Biology, May 2019



Main supervisor: Dr. Kim Præbel (NFH, UiT)

Co-supervisors: Dr. Shripathi Bhat (NFH, UiT）
Dr. Roy A. Dalmo (NFH, UiT)

Abstract

Salmon louse (*Lepeophtheirus salmonis*) is a parasitic copepod that causes significant damage to Atlantic salmon (*Salmo salar*) and incurs significant costs to the salmonid aquaculture industry. In Norway, the industry has been coping with salmon lice issues with various methods such as chemical treatments, but it has resulted in an acquisition of resistance in salmon lice towards the existing delousing treatments. In order to maximize the effectiveness of the treatments, it is important to understand population structuring of salmon lice, and factors that are driving the structure, since recent studies suggested that several Atlantic salmon lice populations are sharing genetic materials coding for resistance toward delousing treatments. The objective of this study was to investigate whether salmon lice in northern Norway display population structure and differentiation at small geographical scales. We identified 25,795 robustly supported SNPs among salmon lice from 8 different locations in northern Norway. While only very weak structure was observed based on the full SNP dataset (25,795 SNPs), relatively weak, but highly significant, population structure was observed using 303 important SNPs identified using a random forest classification approach. No significant correlation between genetic differentiation and geographical distance was observed. To our knowledge, it was the first study to reveal the significant population structure of salmon lice at such small spatial scales. Our results suggested that the observed population structure was a result of dispersal and mixing of salmon lice, as well as selective pressure such as delousing treatments and local environment. Our results are potentially helpful to make population specific delousing treatments, where the treatments are designed depending on the genetic characteristics of the targeted salmon lice population.

Keywords: salmon lice, northern Norway, population genetics, SNPs, RAD sequencing, random forest machine learning

Table of Contents

Preface

This Master's thesis project was a contribution to LICE-MAP project - Mapping genomic structure, adaptation, and resistance of salmon lice populations. Time went so fast since I first got informed from my supervisor, Kim, that I would be able to work on this salmon lice genome mapping project in northern Norway. Since then, this project has been a main part of my everyday life. I sometimes even dreamt of pipetting in the lab, but now, this relationship with my thesis is almost ending. I am happy to accomplish the work, and at the same time, I find myself missing my student life full of enjoyment and struggles. I initially came to Norway with a hope to do research that could potentially contribute to the development of the aquaculture industry. I am glad that my hope was fulfilled through being involved in the LICE-MAP project, whose outcomes will potentially be helpful to make more efficient delousing treatments.

I would like to start my thesis with some acknowledgements. I have so many people that I would love to mention here, for their warm support from close and distant places, and amazing supervision. I thank you so much for my main supervisor Kim Præbel, for giving me this great opportunity to be involved in the LICE-MAP project, and guiding my development as a scientist, and as a person. A very big thank you is given to my co-supervisor Shripathi Bhat for helping me especially with data lab part. I was almost a beginner in terms of data analysis and asked him a lot of questions, but he was always so patient and kind. I also would like to thank another co-supervisor Roy A. Dalmo for supporting me with encouraging words.

And I want to say thank you to Julie Bitz-Thorsen, who kindly assisted me in sampling in Kirkenes. She was my lab supervisor when I was working on DNA extraction and RAD library building. Thanks to her supervision, I succeed in finishing all the wet lab part in time, and my work became much more enjoyable. Special thanks are given to Magnus Nygaard, who kindly drove me to Arnøy Laks AS, helped me with the actual sampling work, and also provided salmon lice from several locations. I thank Filipe

Figueiredo for giving me helpful advice regarding salmon lice DNA extraction, and a fellow master student Simon Schmidt, for his nice company in the Population Genetics Lab. I also thank Kamilla Bentsen for checking my English grammar. I am grateful to all the members of the Research Group for Genetics at NFH, for giving me good advice regarding my lab work and thesis. I am so glad that I have been a part of this group of amazing people. Having lunch with them almost every day at 13 o'clock gave me good stimulations and motivation to do a good job. Finally, I thank my family in Japan for caring about my well-being, and my best friend Salih Burma for always cheering me up with that warm smile. This thesis could not have been done without support from all those people around me.

Soon I am starting a new chapter of my life -working as a biologist at a land-based rainbow trout farming company in Japan. My main work will not be related to genetics or salmon lice, but knowledge and skills I acquired through working on this project will for sure a big help to me. It is heartbreaking to leave the university and people I really like, but whenever I look back on my two years in Tromsø, it will be recalled as warm and happy memories.

May 2019, Tromsø

Kana Banno

坂野 加奈

Introduction

Norwegian salmon aquaculture industry and salmon lice – a threat to the industry

Norway has been a world leader in salmon aquaculture since the production technique was pioneered in the late 1960s, and by 2015, the Norwegian production constituted 53% of the world's Atlantic salmon, *Salmo salar* (Linnaeus) production (Olaussen, 2018). Salmon louse, *Lepeophtheirus salmonis* (Krøyer) is a parasitic copepod that causes severe damage to Atlantic salmon and incurs significant costs to the Atlantic salmon aquaculture industry (Carmichael et al., 2013). In this Master's thesis, *L. salmonis* will be referred to as salmon lice. Since the 1960s, the Norwegian salmon aquaculture industry has been battling with the threat of the salmon lice with various control and preventive methods within farms (Overton et al., 2018). Depending on the severity, salmon lice infestations can cause mechanical damage such as skin and fin lesions (Wootten et al., 1982; Jones et al., 1990; Bjørn & Finstad, 1998), osmotic problems and secondary infections (Nolan et al., 1999; Pike & Wadsworth, 1999; Bowers et al., 2000; Finstad et al., 2000; Tully & Nolan, 2002; Heuch et al., 2005), and in severe cases, mortality (Grimnes & Jakobsen, 1996; Bjørn & Finstad, 1998). In Norway, economical losses within industries attributed to salmon lice infestations and treatments were estimated to 5 to 8 billion NOK annually (Liu & Bjelland, 2014). Moreover, a dramatic increase in salmonid production has increased the abundance of salmon lice in the wild, since salmonids are cultured in open net cages, and salmon lice can be dispersed from host farmed fish to coastal water (Pike & Wadsworth, 1999; Thorstad et al., 2015). Salmon farms are often located near wild salmonid migration routes, with smolts being particularly vulnerable to salmon lice infestation (Aaen et al., 2015). In Norwegian waters, the high abundance of salmon lice is thought to be responsible for stock decrease of wild Atlantic salmon and seatrout (*Salmo trutta* L.) (Krkosek et al., 2013; Thorstad et al., 2015), raising issues about loss of biodiversity and socioeconomic challenges.

How is the salmon aquaculture industry fighting against salmon lice issues?

The salmonid aquaculture industry has been handling the challenge of salmon lice infestations by a combination of mechanical, biological, and medical/chemical treatments (Costello, 1993; Aaen et al., 2015). Mechanical treatments include e.g. freshwater baths (Powell et al., 2015; Wright et al., 2016), and warm water treatment (Ljungfeldt et al., 2017). Biological treatments refer to a use of cleaner fish such as lumpfish (*Cyclopterus lumps*) (Imsland et al., 2014; Powell et al., 2018), and wrasse, e.g. corkwing wrasse (*Symphodus melops*) (Gonzalez & de Boer, 2017). Medical treatments have historically been the most predictable measures to prevent the occurrence of the high salmon lice abundance, but extensive use of medicines has resulted in an acquisition of resistance (Jones et al., 1992; Treasurer et al., 2000; Aaen et al., 2015), which makes the delousing treatments less effective and more expensive.

Simple account for the biology of salmon lice

The life cycle of salmon lice consists of non-feeding planktonic larvae (nauplii), infective planktonic copepodites, immature chalimus embedded on the host skin, and mobile pre-adults and adults that can move freely over the host skin (Jones & Beamish, 2011; Torrissen et al., 2013, Thorstad et al., 2015) (Figure 1). Their free-swimming and non-feeding planktonic nauplii stage is relatively long (Johnson & Albright, 1991; Stien et al., 2005), and at this stage the nauplius can be passively dispersed by water currents (Johnsen et al., 2016). During the host associated phases of their life cycle, salmon lice feed on mucus, skin, and blood of their host fish (Kabata, 1974; Pike, 1989; Dawson et al., 1998; Boxaspen, 2006). The lifespan of salmon lice under natural conditions has not been determined, but under laboratory conditions, females can live up to 210 days. Their generation time is around 8-9 weeks at 6°C, and 6 weeks at 9°C and 4 weeks at 18°C (Marine Institute, https://www.marine.ie/Home/site-area/areas-activity/aquaculture/sea-lice/life-cycle-salmon-louse). Females can produce 10 to 11 pairs of egg strings over their life cycle and mean egg numbers per string have been recorded as 152 at 7.2°C (Heuch et al., 2000). Their relatively fast life cycles with high level of replication, dispersal, and obligate host-association give this species a high potential for evolving new adaptations toward the environment and delousing treatments (Denholm et al., 2002; Boxaspen, 2006).

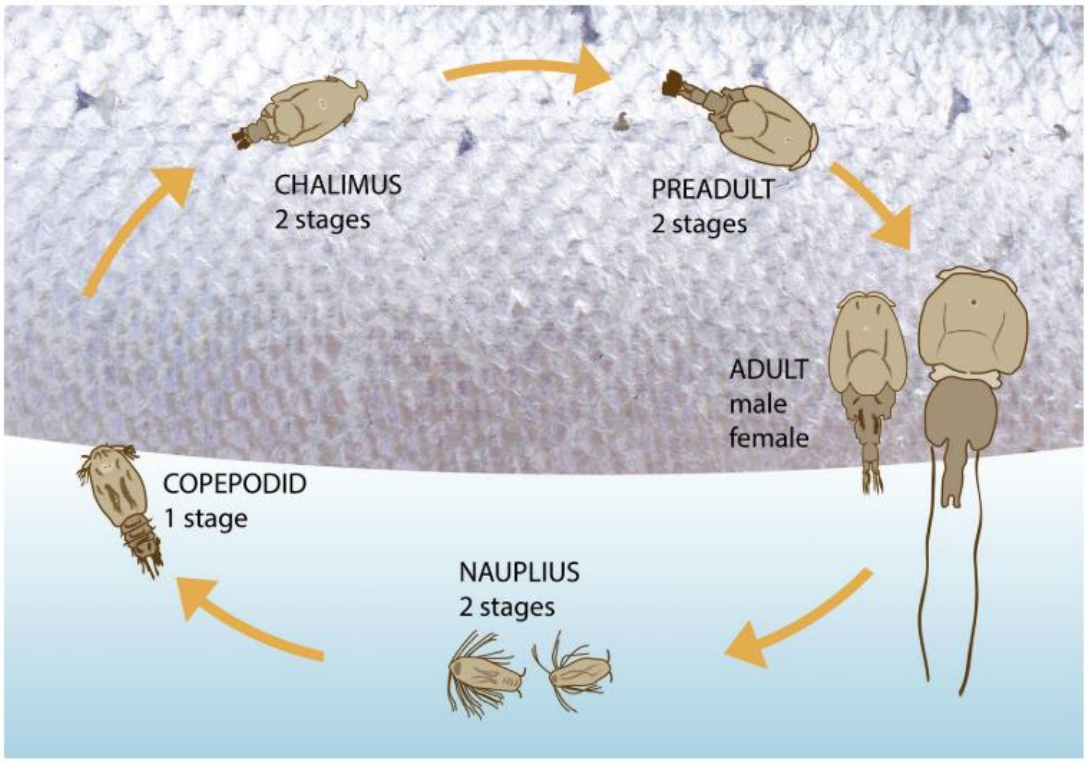


Figure 1. *The 5 phases of the salmon louse life cycle. Each phase comprises 1 or 2 life stages. The sizes of salmon lice on different life stages do not correspond to real size scale. Nauplius are typically ~0.5-0.6 mm, copepodids 0.7 mm, chalimus 1.1-2.3 mm, preadults 3.4-5.2 mm and adults 5-6 mm (males) and 8-12 mm (females). Graphic design: Kari Sivertsen, NINA. The figure and description are cited from Thorstad et al. (2015).*

Salmon lice genetics and use of genetics in the fight against salmon lice

Salmon lice in Norway is thought to be part of a larger panmictic North-east Atlantic population (Todd et al., 2004; Tully & Nolan, 2002), and this assumption has been the basis for the management and actions against salmon lice for the last decades. However, previous studies on salmon lice population genetics used classical markers such as microsatellite loci (e.g. Todd et al., 2004), which lacked the resolution and power to identify genes and alleles involved in genomic differentiation. Modern genome sequencing methods, such as Restriction-site Associated DNA sequencing (RADseq), is a promising avenue for obtaining high density coverage of genetic markers across the genome, which is needed for studies of species with low differentiation, high dispersal rate and rapid evolving adaptations (Andrews et al., 2016). Until now, only a few, but successful attempts have been made to investigate salmon lice population genetic

structure using these modern techniques (e.g. Jacobs et al., 2018). Current evidence supports the theory that several Atlantic salmon lice populations are sharing genetic material coding for resistance towards pesticides (Aaen et al., 2015), for instance towards emamectin benzoate (EMB) (Besnier et al., 2014; Messmer et al., 2018). Therefore, having a clear understanding of gene flow and connectivity of salmon lice populations is essential in terms of health maximization of salmonids (Todd et al., 2004). However, little is known about connectivity and gene flow among salmon lice populations in Norway, despite the importance of knowing whether they are genetically structured. Such knowledge would allow the industry to make population specific delousing treatments based on the genetics of salmon lice, which would maximize the effectiveness of delousing treatments.

Use of RADseq and random forest machine learning in population genetic studies

Genotype-by-sequencing (GBS) is a novel application of next-generation sequencing (NGS) (or high-throughput sequencing) protocols, which has been successfully used for discovering and genotyping single nucleotide polymorphisms (SNPs), which are point mutations spread throughout the genomes (Morin et al., 2004) of any organisms (He et al., 2014). GBS can generate a large number of SNPs for genetic analyses and genotyping (Beissinger et al., 2013), and RADseq is a form of the GBS method. RADseq (Etter et al., 2011; Andrews et al., 2016) is a method that uses a restriction enzyme and samples at restriction sites across the targeted genome (Herrera et al., 2015). By sequencing the restriction sites, this technique can rapidly discover thousands of SNPs for any organisms including non-model organisms at reasonable costs (Davey & Blaxter, 2011; Andrews et al., 2016). RADseq has been successfully used in population genetic studies, especially for species with high potential for gene flow and huge population size, e.g. for American lobster (*Homarus americanus*) (Benestan et al., 2015), European eel (*Anguilla anguilla*) (Laporte et al., 2016), and eastern oyster (*Crassostrea virginica*) (Bernatchez et al., 2018). However, in order to find population structure of such species, RADseq sometimes needs to be used coupled with powerful tools to identify important markers. One of the most promising approaches to identify genetic markers associated with, e.g., environmental variables or adaptive traits, in genomic data is the use of random forest (Breiman, 2001; Chen & Ishwaran, 2012).

Random forest is a machine learning approach that considers a subset of features or predictive variables such as SNPs at each node to grow a series of decision trees (Breiman, 2001). The random forest algorithm considers loci in various combinations of subsets, improving the power of the algorithm to rank these features or loci based on its importance ("important markers"). The use of random forest also allows us to identify markers that are involved in polygenic selection. The increasing use of random forest in biological research is indicating its potential for successful use in population genetic studies (Sylvester et al., 2018). In a population genomics study of American eels (*Anguilla rostrata*) using random forest classification approach on RAD- SNP data, it was shown that ecotypes of American eels, which were thought to be panmictic, were genetically distinct（Pavey et al., 2015). In a study of North Atlantic eel (*Anguilla anguilla*), the use of random forest successfully detected signals of polygenic selection (Laporte et al., 2016). Another study using random forest revealed the presence of significant genetic structure differentiation among four salmon lice populations from Ireland, Scotland, and northern Norway (Jacobs et al., 2018). In this study, the authors only found a weak structure when they used the dataset consisting of all SNPs. However, population discriminatory SNPs identified through the random forest machine learning classification approach revealed significant genetic differentiation among the study locations across North East Atlantic, which was a much wider geographical scale than this study.

Research question and hypothesis

The objective of this study was to investigate the extent of genomic differentiation among the salmon lice from different locations at small geographical scales in northern Norway. It was hypothesized that salmon lice display relatively weak, but significant population genomic structure in northern Norway, and that population structure is driven by a few loci with polygenic background. The results of the study are potentially helpful to make population specific treatment, for example by adjusting the amount of delousing chemicals or changing delousing methods depending on the genetic background of the targeted salmon lice population. This could improve the cost effectiveness of delousing treatments, maximize fish health, and mitigate impacts on wild salmonids stocks.

Materials and Methods

Sampling locations and collections

Salmon lice were collected from 8 different locations in northern Norway (Figure 2, Table 1). Out of 8 sampling locations, samples from Danielsvika, Fartøyvika, Klubben, Russelva, Skjevøy vest, and Uløybukt were sampled at a salmon slaughter house belonging to Arnøy Laks AS (Lauksletta, Troms). Samples from Latvika were collected at an anonymous salmon slaughter house in Finnmark. Samples from Skulgambukt were collected at a salmon rearing facility belonging to UiT's marine research station (Kårvik, Troms). Sampling was done during November to December 2017, and October to November 2018. At the location Uløybukt, salmon lice were collected from 5 different net pens. For population structure analysis, we only retained samples from one net pen (UB7) randomly (Table 1). UB7 will be referred to as UB in later chapters. Salmon lice were picked directly from the skin of the Atlantic salmon. The majority of salmon lice sampled were female since it was easier to find and pick larger female louse than smaller male louse (out of 376 samples, female 90.16%, male 9.84%). All salmon lice were preserved immediately in 96% EtOH upon collection and stored in -20℃ freezer. Individuals were examined under a dissecting microscope to ensure that they were *L. salmonis* because in the case of small males, they appear morphologically similar to *Caligus elongatus* (Nordmann) to the naked eye.

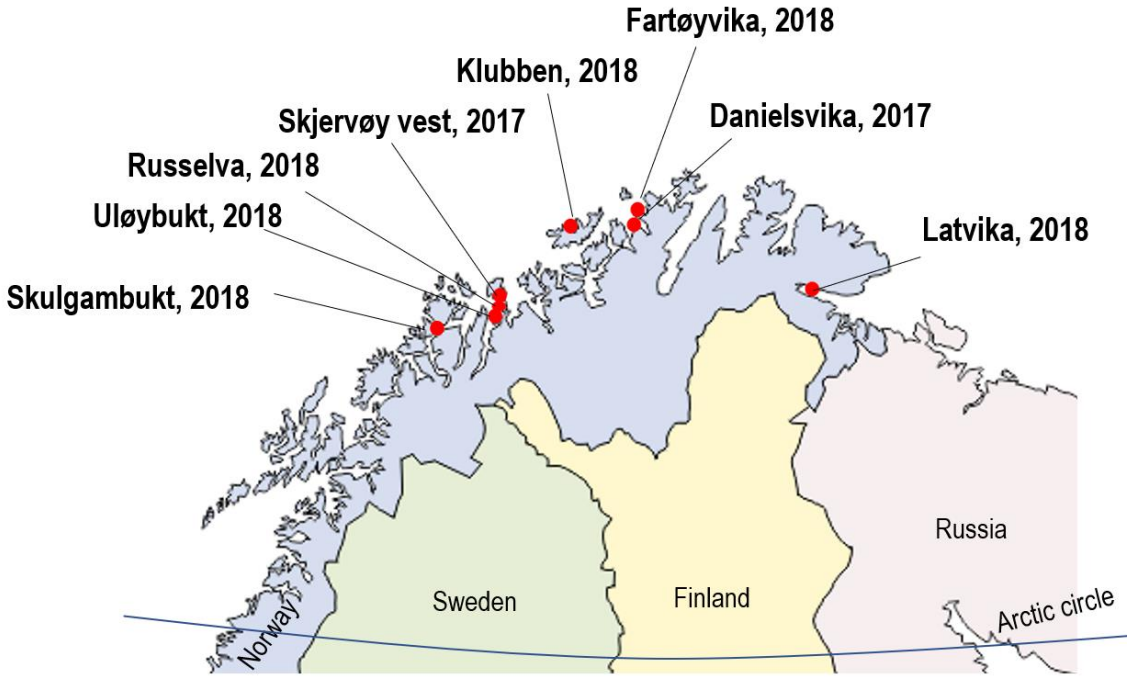


***Figure 2**. Origin of the 376 salmon lice samples in northern Norway. The year where the samples were collected is given next to the location name. Confer to Table 1 for exact locations, and number of sampled individuals per location.*

DNA extraction

Approximately 8-12 mg of tissue per individual salmon lice was used for DNA extraction. When females were used for DNA extraction, only the cephalothorax was used. Utmost care was taken to avoid contamination of tissues with genital segments and egg strings by removing these parts with dissecting scissors and tweezers. However, for males, the whole body was used due to their small size. DNA was extracted using QIAGEN DNeasy Blood & Tissue Kit by following the manufacturer's protocol. To assess the quality, DNA was visualized using electrophoresis on a 0.8% agarose gel (see Appendix 1 for the gel electrophoresis results). Concentration of double-stranded DNA of each sample was measured using either Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher Scientific) (see Appendix 2 for its protocol) or Qubit4 dsDNA HS Assay Kit (Thermo Fisher Scientific).

RAD library preparation and sequencing

RAD libraries for 288 samples (276 samples and 12 replicates) (Table 1) were constructed using a modified version of RAD library preparation protocol from Benestan et al. (2015) (see Appendix 3 for more details about the protocol). Briefly, 500ng of DNA was digested with the restriction enzyme *PstI-HF (#R3140L)* (restriction site 5'…CTGCA*G…3', 3'…G*ACGTC…5'). This enzyme was chosen based on information from a previous study (Carmichael et al., 2013). Digested DNA from each sample was ligated with the P1 adapter, which contained a flow-cell adapter sequence, an Illumina sequencing primer binding site, and unique barcodes of 6-8bp for sample identification post demultiplexing. 12 individuals were combined to form individual pools. Each pool was then sonicated using Bioruptor (Diagenode) to obtain DNA fragments with average size of 500bp. DNA fragments between 300-800bp were selected for downstream treatment. Downstream treatment included adding the P2 adapter (flow-cell adapter + PE sequencing primer binding site), PCR enrichment, and fragment size selection using Blue Pippin (Sage Science Inc). Aimed fragment size of the final product was 350-600bp. Mean fragment size of each pool was assessed with Agilent 2100 Bioanalyzer system (Agilent Technologies, Inc.). Then, 8 pools of libraries (96 individuals) were further pooled in equimolar proportions for each sequencing lane. A total of 3 lanes were sequenced paired-end, with 150 bp read length, on a HiSeq 4000 at Novogene (HK) Company Limited (Hong Kong).

Table 1. *Overview of the salmon lice samples. It includes names of sample origins, locality codes consisting of 2 alphabets taken from the location names, sampled location, map positions, sampled dates, number of individuals collected per location, and number of samples included for the RAD library preparation. Latitudinal and longitudinal information for each location were obtained from Barents Watch website (https://www.barentswatch.no). The names of sample origins are in alphabetical order.*

Origin of samples	Locality code	Sampled Location	Position	Date	Sampled individuals	RAD library
Danielsvika	DL	Arnøy Laks AS	70°40'N 24°29'E	24-27.11.17	19	19
Fartøyvika	FT	Arnøy Laks AS	70°49'N 24°27'E	12.10.18	40	29
Klubben	KB	Arnøy Laks AS	70°39'N 22°36'E	05.10.18	36	28
Latvika	LV	Anonymous	Anonymous	09.11.18	63	29
Russelva	RE	Arnøy Laks AS	69°53'N 20°46'E	04.10.18	38	28
Skjervøy vest	SV	Arnøy Laks AS	70°02'N 20°57'E	11.12.17	23	23
Skulgambukt	SK	UiT Research station	69°47'N 19°09'E	14.11.18	27	27
Uløybukt M2	UB2	Arnøy Laks AS	69°51'N 20°42'E	09.10.18	50	23
Uløybukt M3	UB3	Arnøy Laks AS	69°51'N 20°43'E	10.10.18	14	14
Uløybukt M4	UB4	Arnøy Laks AS	69°51'N 20°44'E	10.10.18	10	10
Uløybukt M6	UB6	Arnøy Laks AS	69°51'N 20°45'E	10.10.18	25	23
Uløybukt M7	UB7	Arnøy Laks AS	69°51'N 20°46'E	10.10.18	31	23
Total number					376	276

Bioinformatic analysis

Most bioinformatic analysis (genotype calling using *STACKS*) was performed by Dr. Shripathi Bhat, NFH (see Figure 3 for contribution). First, raw sequence data were demultiplexed into individual samples using *process rad_tags* module from *STACKS* v.2.1 (Catchen et al., 2011; Catchen et al., 2013; Rochette et al., 2019) using default options, and reads were truncated to final length of 130 bp (because base quality went down after 130 bp). Further, the demultiplexed short paired-end reads were mapped to the latest version of the salmon lice genome (https://licebase.org/organism/Lepeophtheirus/salmonis) using GSNAP aligner (Wu & Nacu, 2010). The reference mapped and sorted bam files were used as input for SNP calling program *STACKS* v.2.1. For the reference aligned analyses, *gstacks*, and *population* module were used as suggested by Rochette et al. (2019). *gstacks* was used to build loci from the single and/or paired-end reads, and then SNPs were called using the *marukilow* SNP calling model. *population* module was used to analyze a population of individual samples by calculating various population genetic statistics and exporting genotypes into various file formats. Along with these, *population* module also enables a conservative SNP filtering option. In the data set we retained for the current study, SNPs were filtered to VCF files if it was present in at least 65% of individuals per location, genotyped in all 8 locations, and had global minor allele frequency of >1%, observed heterozygosity of <0.5. We also retained a single SNP per RAD-tag to reduce

obvious linkage disequilibrium between multiple SNP loci. *PGDSpider* (Lischer & Excoffier, 2012) was used to convert the VCF file format into format required for all downstream analyses. *VCFtools* (Danecek et al., 2011) and *BCFtools* (Narasimhan et al., 2016) were used to manipulate VCF files. We used two data sets in the current study. One dataset consisted of all the SNPs filtered following above described filtering steps, and it will be referred to as "full SNP dataset" in later chapters. From this dataset, another dataset was created by the random forest classification approach, which will be referred to as "important SNP dataset" or "important SNPs" in later chapters.

Estimation of basic population genetic parameters

In total, 206 samples from 8 different locations (DL, FT, KB, LV, RE, SK, SV, and UB) were included for the population structure analysis after removing the 12 replicates and unused populations (70 samples from the 4 net pens in Uløybukt). Summary statistics (H_o: observed heterozygosity, H_e: expected heterozygosity, F_{IS}: inbreeding coefficient) were calculated using *genodive* v.2.01 (Meirmans & Van Tienderen, 2004).

Test for loci under selection

To determine loci putatively under selection, we used the *Fdist* (Beaumont & Nichols, 1996) based approach implemented in *Bayescan* (Foll & Gaggiotti, 2008), with prior odds of 100. *Bayescan* is known to be a more conservative approach and its estimates have less type I errors (Narum & Hess, 2011). *Bayescan* outputs were interpreted and plotted in R (R Core Team, 2018, https://www.R-project.org/). Furthermore, we performed a second outlier analysis using *Arlequin* ver. 3.5.2.2 (Excoffier & Lischer, 2010) to compare with the *Bayescan* results. We performed Arlequin analysis with 50,000 simulations and 100 demes. SNPs were considered putatively under selection at $p<0.05$.

Population differentiation and population structure analysis

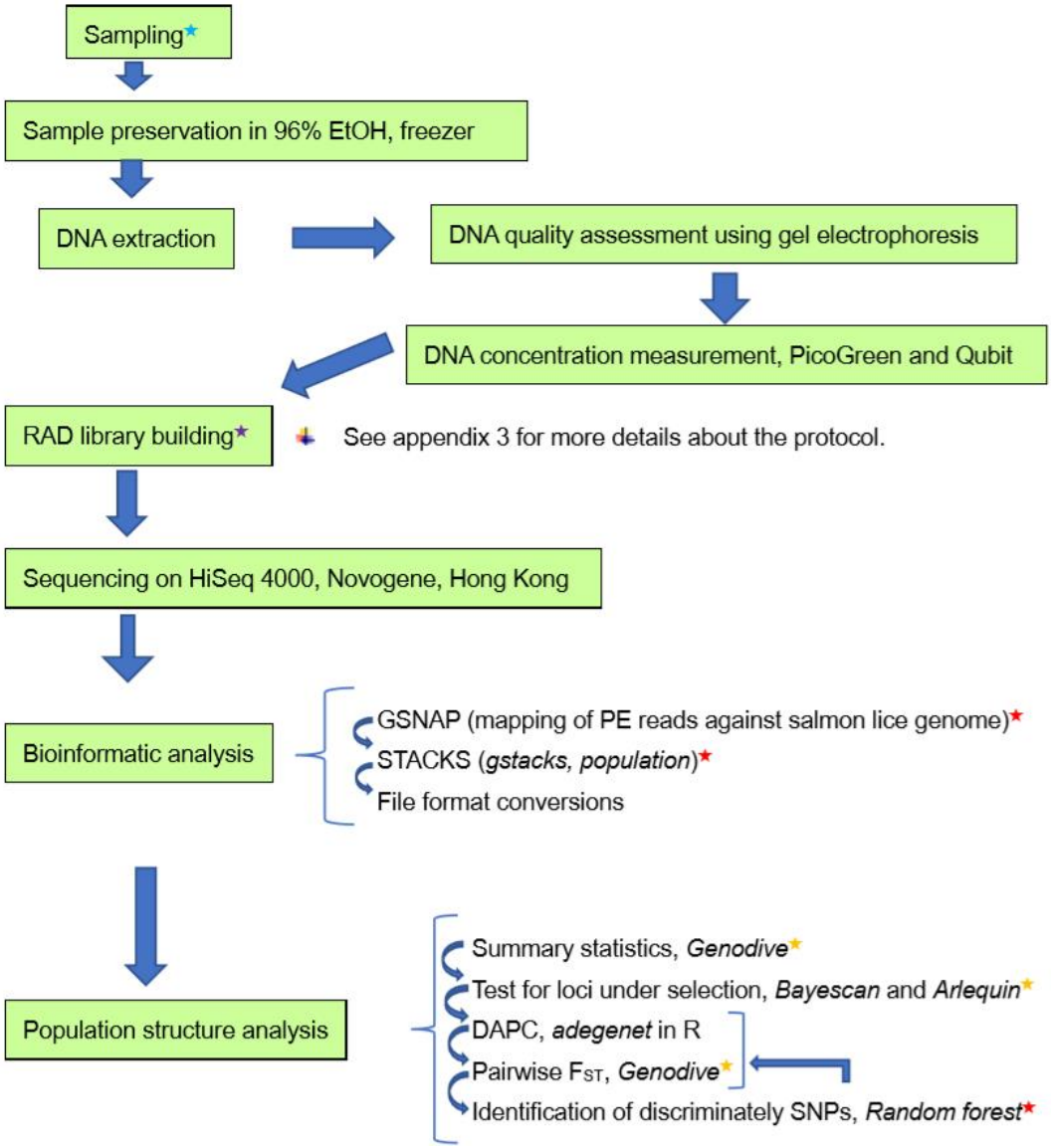
a) Using full SNP dataset

Discriminant analysis of principal components (DAPC) was performed in *adegenet* (Jombart & Ahmed, 2011) in R. *adegenet* is a R package used for handling and analysis of genome-wide SNP data, and it enables the analysis of large genome-wide SNPs datasets by using standard personal computers (Jombart & Ahmed, 2011). To choose the number of principal components (40 in our case) required for DAPC, we used *xval* approach recommended by the authors of *adegenet*, as the use of overestimated PCs (too many PCs) could result in overfitting. *xval* runs were performed 5 times to test the convergence of runs. Pairwise F_{ST} between each sampling location was calculated in *genodive* v.2.01. Associated p-values were obtained with bootstrap 1000, and the p-values were adjusted for multiple comparisons with the BH (Benjamini-Hochberg) method using *p.adjust* function in R, to control the expected proportion of false discoveries (type II errors). Isolation-by-distance (IBD) was evaluated with mantel test of all pairwise F_{ST} comparison as a function of geographical distances between the sampling locations using *mantel.randtest* function in R. Distance between the locations was calculated using Latitude/Longitude Distance Calculator (National Hurricane Center, USA, https://www.nhc.noaa.gov/gccalc.shtml) based on the position information obtained from Barents Watch website (https://www.barentswatch.no).

b) Using "important SNPs" detected by using random forest

Detection of the important SNPs using the random forest classification approach was performed by Shripathi Bhat. In order to detect the important SNPs, we used a tree-based machine learning approach, *randomForest* (Liaw & Wiener, 2002) in R. *randomForest* is an ensemble of decision trees that perform the classification and regression task, where each decision tree is constructed using different bootstrap samples from the training data. Herein, the SNP data were subdivided randomly into a training data set consisting of 2/3 of the data, which was used to determine the association between the SNPs and the origin of samples. 1/3 of the SNP data (Out-of-Bag, OOB) were used to evaluate the predictive power of the SNPs to correctly classify the individuals into the right origin. A similar approach has successfully been used in several recent population genomic studies (Pavey et al., 2015; Laporte et al., 2016;

Jacobs et al., 2018). We converted the VCF file to a *plink2* (Chang et al., 2015) friendly format, using *BCFtools*. *plink2* was further used to numerically code all the genotypic and population relevant information. Missing genotypes were imputed using the *rfImpute* function in the *randomForest* package. We ran each random forest 3 times with 100,000 trees and checked for the convergence between runs by calculating the Pearson's correlation coefficient (r^2) among the SNP importance values (measured as Mean Decrease in Accuracy, MDA). In the first random forest run, we removed the SNPs with a MDA score of less than 0. In the subsequent backward purging approaches, we removed the SNPs having the lowest OOB error rate as suggested in Pavey et al. (2015), Laporte et al. (2016), and Jacobs et al. (2018). We continued the backward purging approach until we found a significant increase in OOB as compare to the previous OOB rate. Once we observed the significant increase in OOB rate, we chose a dataset which showed the lowest OOB in previous purging steps (i.e. elbow point in elbow curve). This dataset was designated as "important SNPs". The population genomic analyses were repeated based on this reduced SNP dataset. To determine the population discriminatory power of these important SNPs, we first performed DAPC using *adegenet* package in R. To estimate the number of PCs to be used for the analysis (60 in our case), we used the *xval* function in the *adegenet* package 5 times. We also calculated pairwise F_{ST} between sampling locations, and corresponding p-values were calculated using *genodive* v.2.01 with bootstrap of 1000. Again, p-values were adjusted for multiple comparisons with the BH (Benjamini-Hochberg) method using *p.adjust* function in R.



★Done by Shripathi Bhat, NFH.

★Done with a help from S. Bhat.

★Done with a help from Julie Bitz-Thorsen, NFH.

★Done with a help from Kim Præbel, NFH, S. Bhat, J. Bitz-Thorsen, and Magnus Nygaard, a previous master student at NFH.

Overall supervision by Kim Præbel.

All the steps except for sampling and sequencing were performed at NFH, UiT.

***Figure 3**. Work flow of this study. It briefly shows the entire work flow of this study.*

Results

Sequencing and genotyping results

The three lanes of HiSeq 4000 (206 samples + 10 replicates) sequencing yielded a total of 1,849,398,680 paired-end reads (median number of reads 7,682,849 reads). After filtering for intact RAD-tags and barcodes we obtained 1,787,556,510 (median number of reads 7,403,310 reads) paired-end reads, which corresponded to a median sequencing depth of 16 per sample. After the *gstacks* analysis, the catalogue file contained 313,812 loci. After applying all of the quality filters available in *population* module, 25,795 SNPs were retained. This data set formed the basis for all of the downstream analysis.

Population genetic summary statistics

H_o (observed heterozygosity) ranged from 0.058 to 0.111, while H_e (expected heterozygosity) ranged from 0.097 to 0.120 (Table 2). Observed heterozygosity was found to be lower than expected heterozygosity for all the locations. F_{IS} value varied among the sample locations and had positive values suggesting individuals within a sampling location were more closely related than what was expected under a model of random mating.

***Table 2**. Summary table. Sample sizes, observed heterozygosity (H_o), expected heterozygosity (H_e), and coefficient of inbreeding (F_{IS}) ($p = 0.001$) for each location are shown in the Table. Location names are as in Table 1. Refer Table 1 for the full name of sampling locations.*

Location	Number of samples	Ho	He	FIS
DL	19	0.094	0.118	0.198
FT	29	0.098	0.113	0.139
KB	28	0.083	0.106	0.217
LV	29	0.087	0.107	0.182
RE	28	0.111	0.120	0.081
SK	27	0.084	0.117	0.288
SV	23	0.070	0.120	0.420
UB	23	0.058	0.097	0.398

Identifying SNPs putatively under selection

Bayescan identified 2 SNPs (0.008%) and 13,077 SNPs (50.696%) out of 25,795 SNPs, putatively under divergent and balancing selection, respectively. The remaining 12,716 SNPs (49.296%) were retained as putatively neutral (Figure 4). In contrast to *Bayescan*, *Arlequin* identified 7 SNPs (0.027%) and 3,343 SNPs (12.960%) putatively under divergent and balancing selection, respectively. The remaining 22,445 SNPs (87.013%) were retained as putatively neutral (Figure 5). One SNP (LSalAtl2s1118_153362) was identified as "putatively under selection" by both programs.

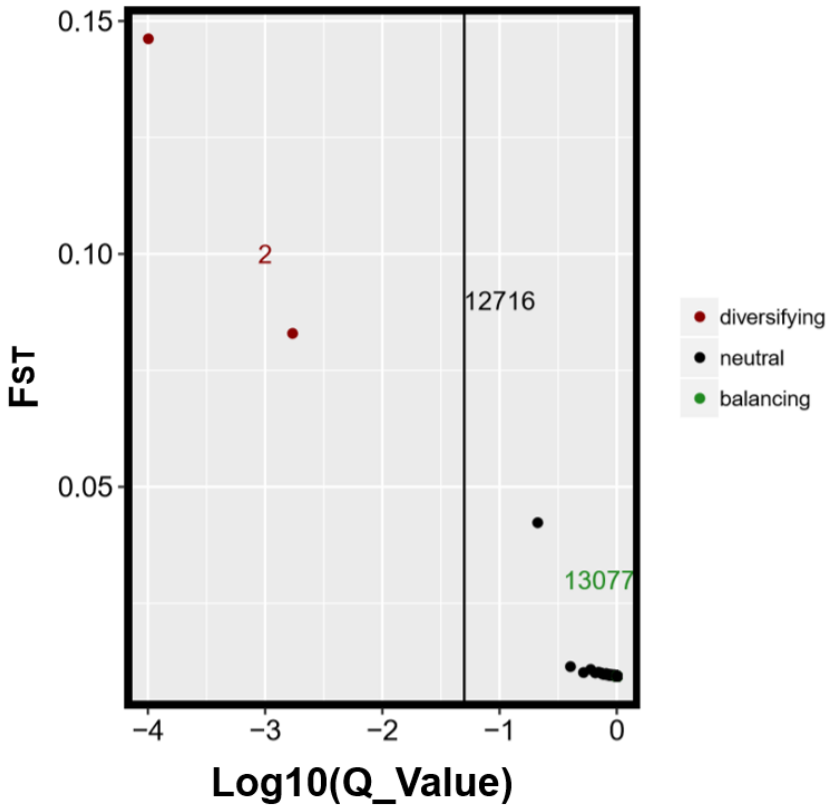


Figure 4. *Bayescan output plot. Red dots, black dots, and green dots represent diversifying SNPs, neutral SNPs, and balancing SNPs, respectively. Each dot represents one SNP. Green dots for balancing SNPs are under the clouds of neutral SNPs (black dots), so it is impossible to see the green dots clearly. Numbers on the plot suggests number of SNPs under respective selection classes (diversifying, no selection (neutral), and balancing).*

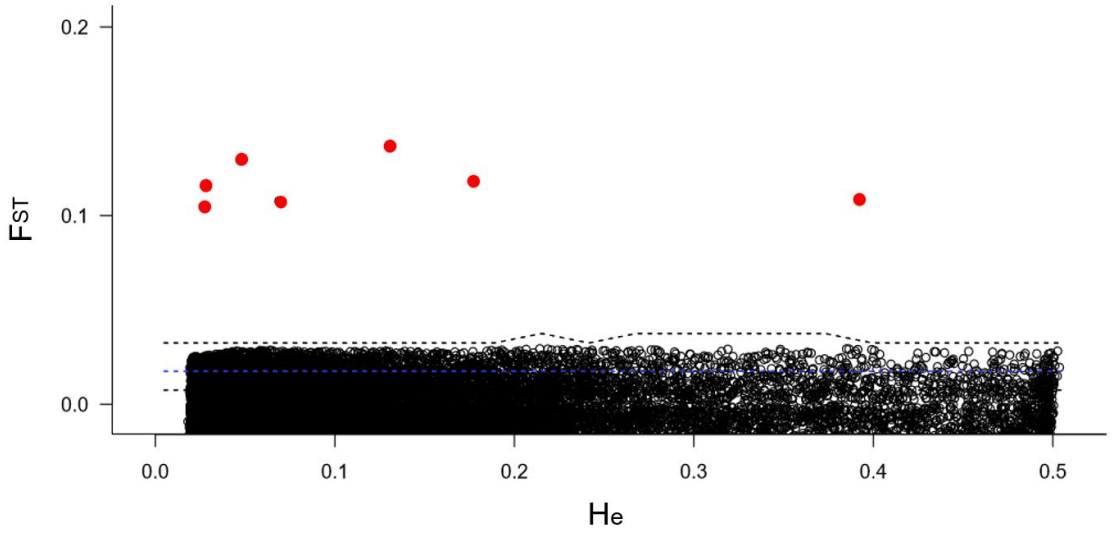


***Figure 5**. Detection of loci under selection based on F_{ST} simulations implemented in Arlequin. Red dots represent loci putatively under diversifying selection. Dotted blue line indicates the 50% quantile. Dotted black lines indicate the 95% quantile.*

Population differentiation and population structure analysis

a) Based on the full SNP dataset

Pairwise F_{ST}´s between sampling locations were found to be weak and ranged from 0.000 to 0.007. Out of 28 F_{ST} comparisons, 21 of them were found to be significant ($p<0.05$) after correcting for multiple testing (Table 3). Regarding DAPC performed based on the full SNP dataset (25,795 SNPs), using *find.cluster* suggested K=1, which indicated lack of population structure in the dataset. When we used sampling locations as priors, none of the sampling locations were shown to be a genetically isolated group in a DAPC scatter plot (Figure 6). However, the salmon lice from Skjervøy vest (SV) looked more genetically distinct from the salmon lice sampled in the rest of the sampling locations. Mantel test for isolation by distance revealed no significant correlation between genetic differentiation and geographical distance (Figure 7) (Mantel r = 0.011, p = 0.401).

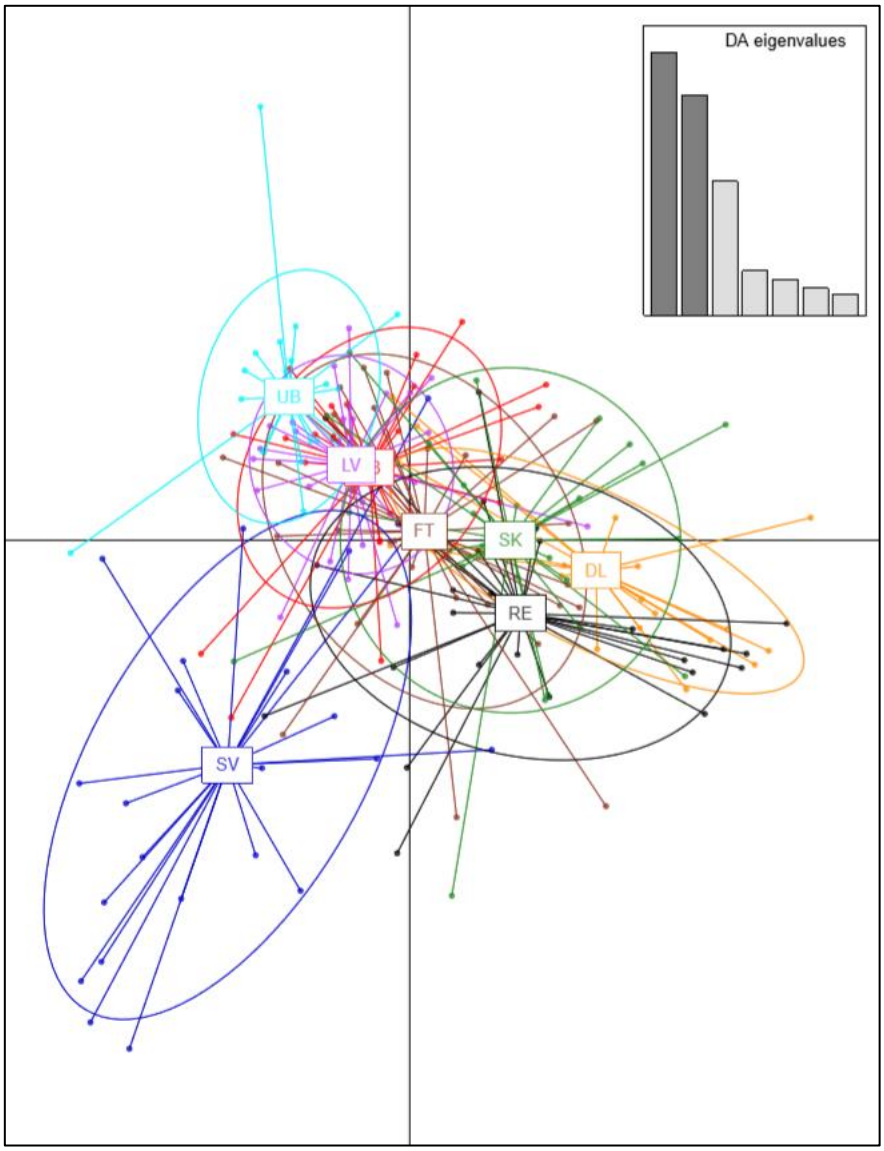


***Figure 6**. DAPC scatter plot based on the full SNP dataset (25,795 SNPs). The scatter plot shows the clustering of individual salmon louse (colored dots) into the groups represented by 95% inertia eclipses. Name of sample locations are represented as locality codes. For full name of the locations, refer to Table 1. Please note that label KB overlaps with label LV. Eigenvalues are shown in the inset.*

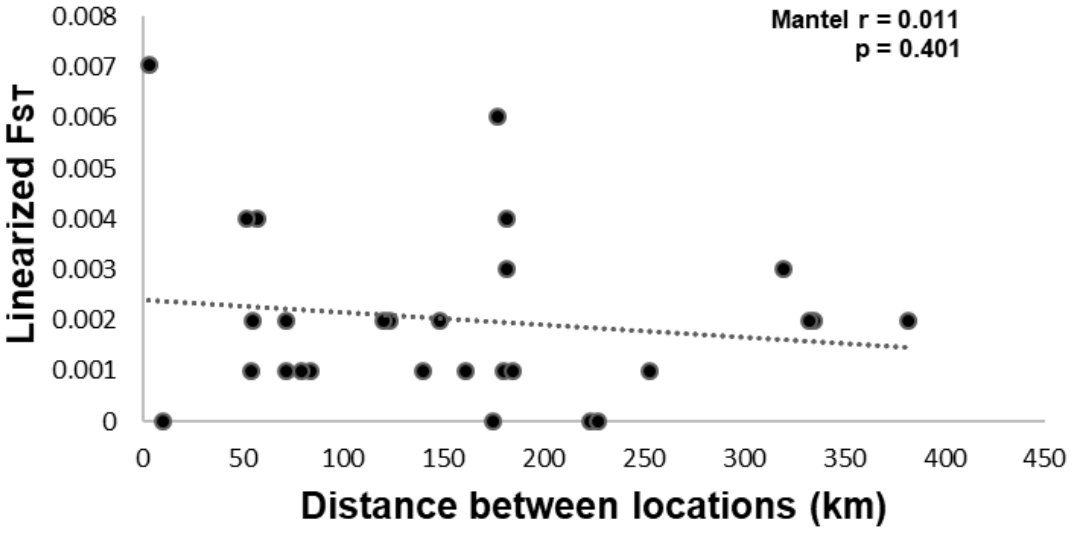


Figure 7*. Results of mantel test for isolation by distance. Genetic differentiation between sample locations was represented as linearized F_{ST} (i.e. $F_{ST}/1-F_{ST}$). Isolation by distance was tested for distance between sample locations measured as kilometers (km).*

b) Based on the "important SNP dataset" detected by using random forest

We detected a subset of 303 SNPs that minimized the OOB error rate to 27.6% as compare to the OOB rate of 84.5% for all the 25,795 SNPs. DAPC analysis based on the important SNP dataset suggested presence of 4 clusters among sampling locations as compared to the 2 main clusters found in the analysis based on the full SNP data set. The 4 clusters consisted of samples from Fartøyvika, Skulgambukt, Uløybukt, and the rest (Figure 8). All pairwise comparisons of F_{ST} showed highly significant ($p<0.05$) values ranging from 0.014 to 0.080 (Table 3).

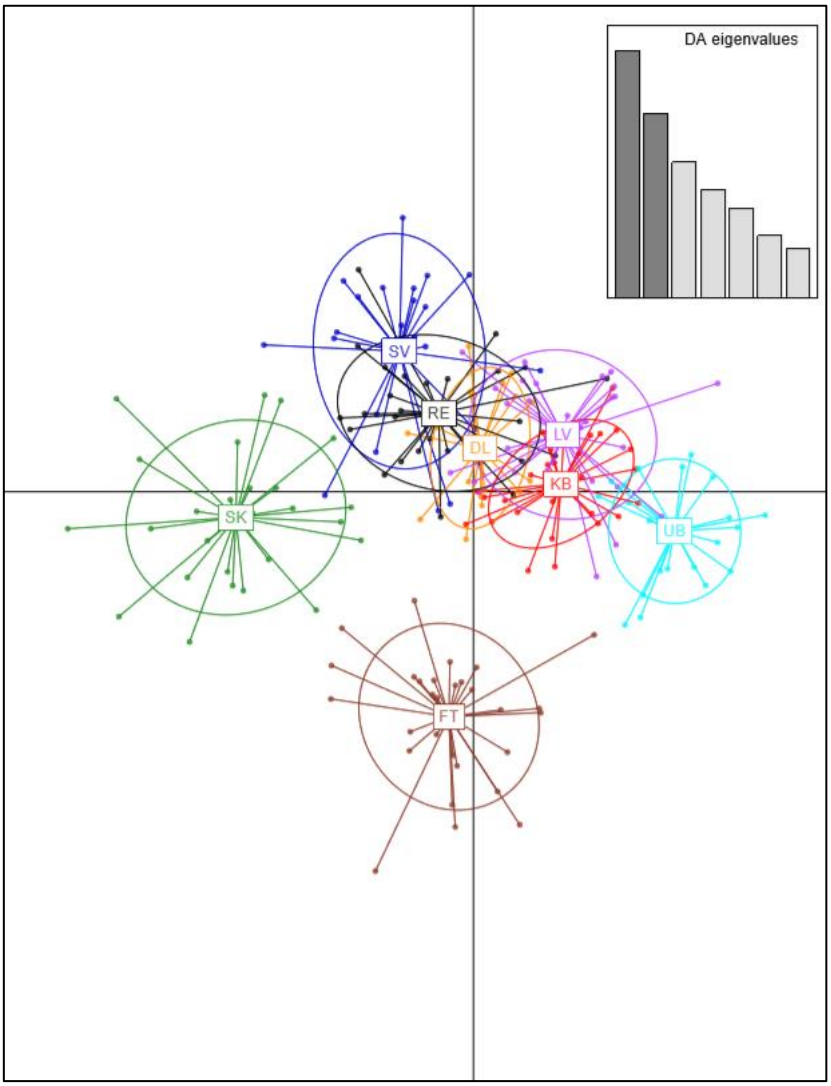


Figure 8. *DAPC scatter plot based on the important SNP dataset (303 SNPs). The scatter plot shows the clustering of individual salmon louse (colored dots) into the groups represented by 95% inertia eclipses. Name of sample locations are represented as locality codes. For full name of the locations, refer Table 1. Eigenvalues are shown in the inset.*

***Table 3**. Heatmap showing pairwise F_{ST}´s between sampling locations, based on the full SNP dataset (below diagonal) and based on the important SNP dataset (above diagonal). Significant values ($p<0.05$) are marked with asterisks.*

	SV	DL	RE	KB	FT	LV	SK	UB
SV		0.031*	0.031*	0.039*	0.031*	0.036*	0.038*	0.080*
DL	0.001		0.014*	0.028*	0.021*	0.023*	0.019*	0.061*
RE	0.002*	0.000		0.038*	0.029*	0.027*	0.028*	0.070*
KB	0.001	0.002*	0.002*		0.037*	0.036*	0.036*	0.050*
FT	0.002*	0.000	0.001*	0.001*		0.028*	0.027*	0.061*
LV	0.003*	0.003*	0.002*	0.001	0.001*		0.036*	0.055*
SK	0.001*	0.000	0.001*	0.001*	0.000	0.002*		0.069*
UB	0.004*	0.006*	0.007*	0.002*	0.004*	0.002*	0.004*	

F_{ST}				
<0.010	0.010-0.030	0.031-0.040	0.041-0.050	>0.050

Discussion

Only a few population genetic studies have been conducted to assess the role of population genetics in management of salmon lice at small geographical scales, despite the significant impacts of salmon lice infestation on the salmonid aquaculture industry and the observed resistance in salmon lice toward existing delousing treatments. To understand that role, we first need to reveal the genetic background of "naïve" salmon lice in areas which are important for the salmonid aquaculture industry, and where many open cages are located (i.e. northern Norway). Hence, investigating population genomic structure of salmon lice in northern Norway formed the main research question of this thesis. Clustering analysis based on the full SNP dataset (25,795 SNPs) did not find signatures of strong grouping among the 8 sampling locations, however, the values of pairwise F_{ST}´s suggested presence of low but significant genetic differentiation among some of the sampling locations. However, using the random forest machine learning classification approach to identify important SNPs (303 SNPs), increased the resolution of population clustering and revealed significant genetic differentiation among the sampling locations.

In this study, the values of pairwise F_{ST}´s based on the full SNP dataset ranged from 0.000 to 0.007, which were lower than the ones observed in the salmon lice population genetic structure study of Jacobs et al. (2018), where the authors found pairwise F_{ST} ranging from 0.002 to 0.01 among 4 locations from Ireland, Scotland, and Norway. The low values of pairwise F_{ST}´s indicated that there was only weak overall genetic differentiation among salmon lice from the 8 sampling locations in this study. In addition, coefficient of inbreeding (F_{IS}) had positive and significant values ranging from 0.081 to 0.420, suggesting that individuals within one location are more related than what was expected under a model of random mating. Combined, these results suggested high potential of salmon lice dispersal and mixing, which was also reported in previous studies, since salmon lice has planktonic life stage and can be passively dispersed by the water current (Salama et al., 2013; Asplin et al., 2014; Johnsen et al., 2016). Salmon

lice could also be dispersed by transportation of host salmonid fish by industrial well boats, and by migration of host fish (Morton et al., 2011), increasing potential for mixing salmon lice from different locations, which leads to lower genetic differentiation among them. Previous works also suggested that marine organisms usually display weak to non-existent population genetic structure over broad spatial scales, due to their high dispersal potential, high fecundity, and large population sizes (Ward et al., 1994; Waples, 1998; Palumbi, 2003; Hedgecock et al., 2007; Gagnaire & Gaggiotti, 2016).

There has been a few attempts to investigate population genomic structure of salmon lice in North Atlantic ocean using classical markers such as microsatellite loci (Bruford & Wayne, 1993), or RAPD (random amplified polymorphic DNA) markers (Lynch & Milljgan, 1994), however, none of the studies have shown a significant population genomic structure of salmon lice within this area (Todd et al., 1997; Todd et al., 2004; Tjensvoll et al., 2006; Nolan & Powell, 2009; Glover et al., 2011). The recent study of Jacobs et al. (2018) found a significant population genomic structure of salmon lice among 4 locations from Ireland, Scotland, and Norway by using RAD sequencing coupled with the random forest classification approach. However, the observed significant population structure was at much wider geographical scales than in this study, which targeted northern Norway as study area. Therefore, to our knowledge, this was the first study which revealed significant population structure of salmon lice at such small spatial scales.

In this study, no significant correlation between genetic differentiation and geographical distance was observed (Mantel $r = 0.011$, $p = 0.401$), suggesting that there was little effect of genetic drift on the observed salmon lice population structure. Similarly, no evidence of correlation between genetic differentiation and geographical distance was reported in salmon lice population genetic studies among farms along the coasts of Ireland (Nolan & Powell, 2009), and 4 locations from Ireland, Scotland, and northern Norway (Jacobs et al., 2018). Since genetic drift does not appear to be driving the population structure, an alternative factor could be gene flow due to salmon lice's high potential for dispersal and mixing (Salama et al., 2013; Asplin et al., 2014; Johnsen et al., 2016), as discussed earlier. Since the life cycle of salmon lice includes rapid

generation time (Costello, 2006), natural selection may also have effects on population structuring of salmon lice. In the aquaculture industry of Atlantic salmon, infestations of salmon lice are treated with anti-parasitic drugs. Recent studies have shown that drug treatment may be linked to several genomic regions in the salmon lice (Besnier et al., 2014; Carmona-Antoñanzas et al., 2017; Messmer et al., 2018), suggesting that those drug treatments have selective pressure on salmon lice. Alternative anti-parasitic treatments, such as freshwater baths (Powell et al., 2015; Wright et al., 2016), and warm water treatment (Ljungfeldt et al., 2017), as well as local environmental condition such as water temperature (Heuch et al., 2000; Samsing et al., 2016), and salinity (Bricknell et al., 2006), may also have selective pressure on salmon lice, as it affects the salmon lice survival and development.

Herein, the random forest classification approach successfully increased the resolution of population structure of salmon lice in northern Norway. This study proved the effectiveness of the random forest classification approach in population genetic studies of species which are known to have characteristics of high gene flow and low genetic differentiation, as shown in previous studies (Brieuc et al., 2015; Laporte et al., 2016; Jacobs et al., 2018; Sylvester et al., 2018). The random forest classification approach is known for its effectiveness in identifying markers that are involved in polygenic selection, and previous theoretical and empirical works suggested that quantitative traits of marine organisms have a polygenic genomic basis (Turelli & Barton, 1990; Berg & Coop, 2014). Both our results and these previous works suggest that adaptive traits of salmon lice toward chemical treatments and local environment have a polygenic genomic basis, and the markers that are involved in polygenic selection were driving the observed population structure of salmon lice in this study. However, it was not only adaptive genomic regions that were involved in population structuring. There are two principal types of genetic diversity: adaptive and neutral. A large part of an organism's DNA is known to be neutral, which refers to a gene or locus that has no or almost no effect on fitness (Holderegger et al., 2006). Only a small proportion of SNPs (< 0.01%) identified in this study were putatively under selection. This suggested the importance of neutral genomic regions in determining the population structuring of salmon lice observed in this study.

Future perspective

This study found a weak but significant population structure of salmon lice in northern Norway. Important SNPs in terms of power of finding the population structure were successfully identified by using the random forest classification approach. In order to utilize our results in the salmonid aquaculture industries, further studies are needed. As next steps, we need to identify specific alleles/genes that are under positive selection, and reveal the function of these alleles/genes, so that we can obtain an understanding of which selective pressures are contributing to salmon lice population structure in northern Norway. We also need to assess if identified populations react differently to delousing treatment by exposing them to delousing chemicals. This may allow us to provide necessary information for making population specific treatment, for example, by adjusting the amount/concentration of delousing chemicals, or changing the method of delousing based on difference in tolerance of salmon lice toward delousing treatments.

Acknowledgements

This project was a part of the first work-package of the LICE-MAP project and it was funded by RFFNOR, the regional office of the Norwegian Research Council in northern Norway. Part of traveling costs for the sampling trips was funded by Faculty of Biosciences Fisheries and Economics (BFE), UiT The Arctic University of Norway (UiT). Expenses of the author's student life in Norway was partly covered by a grant from the Norway Japan Society (Tokyo, Japan). I would like to show my great appreciation to all of the funders. I also thank the Norwegian government for providing me free education, and UiT for giving me an opportunity to study in such a great and international environment. Special thanks are given to Arnøy Laks AS for the collaboration in this project. An anonymous salmon aquaculture company in northern Norway and the UiT's research station in Kårvik are also thanked for supporting us with access to salmon lice samples.

References

Aaen, S. M., Helgesen, K. O., Bakke, M. J., et al., 2015. Drug resistance in sea lice: A threat to salmonid aquaculture. Trends in Parasitology, 31(2), 72-81.

Andrews, K. R., Good, J. M., Miller, M. R., et al., 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics, 17, 81-92.

Asplin, L., Johnsen, I. A., Sandvik, A. D., et al., 2014. Dispersion of salmon lice in the Hardangerfjord. Marine Biology Research, 10(3), 216-225.

Beaumont, M. A., and Nichols, R. A., 1996. Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society of London. Series B, 263, 1619-1626.

Beissinger, T. M., Hirsch, C. N., and Sekhon, R. S., 2013. Marker density and read depth for genotyping populations using genotyping-by-sequencing. Genetics, 193, 1073-1081.

Benestan, L., Gosselin, T., Perrier, C., et al., 2015. RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). Molecular Ecology, 24, 3299-3315.

Berg, J. J., and Coop, G., 2014. A population genetic signal of polygenic adaptation. PLoS Genetics, 10(8): e1004412.

Bernatchez, S., Xuereb, A., Laporte, M., et al., 2018. Seascape genomics of eastern oyster (*Crassostrea virginica*) along the Atlantic coast of Canada. Evolutionary Applications, 12, 587-609.

Besnier, F., Kent, M., Skern-Mauritzen, R., et al., 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, 15:937.

Bjørn, P. A., and 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, 970-977.

Bowers, J. M., Mustafa, A., Speare, D. J., et al., 2000. The physiological responses of Atlantic salmon, *Salmo salar* L., to a single experimental challenge with sea lice, *Lepeophtheirus salmonis*. Journal of Fish Diseases, 23(3), 165-172.

Boxaspen, K., 2006. A review of the biology and genetics of sea lice. ICES Journal of Marine Science, 63, 1304-1316.

Breiman, L., 2001. Random forests. Machine Learning, 45, 5-32.

Bricknell, I. R., Delesman, S. J., O'Shea, B., et al., 2006. Effect of environmental salinity on sea lice *Lepeophtheirus salmonis* settlement success. Diseases of Aquatic Organisms, 71, 201-212.

Brieuc, M. S. O., Ono, K., Drinan, D. P., et al., 2015. Integration of Random Forest with population-based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (*Oncorhynchus tshawytscha*). Molecular Ecology, 24, 2729-2746.

Bruford, M. W., and Wayne, R. K., 1993. Microsatellites and their application to population genetic studies. Current Opinion in Genetics and Development, 3, 939-943.

Carmichael, S. N., Bekaert, M., Taggart, J. B., et al., 2013. Identification of a sex-linked SNP marker in the salmon louse (*Lepeophtheirus salmonis*) using RAD sequencing. PLoS ONE, 8(10): e77832.

Carmona-Antoñanzas, G., Bekaert, M., Humble, J. L., et al., 2017. Maternal inheritance of deltamethrin resistance in the salmon louse *Lepeophtheirus salmonis* (Krøyer) is associated with unique mtDNA haplotypes. PLoS ONE, 12(7): e0180625

Catchen, J., Amores, A., Hohenlohe, P., et al., 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes, Genomes, Genetics, 1, 171-182.

Catchen, J., Hohenlohe, P., Bassham, S., et al., 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology, 22, 3124-3240.

Chang, C. C., Chow, C. C., Tellier, L. C., et al., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience, 4, 1-16.

Chen, X., and Ishwaran, H., 2012. Random forests for genomic data analysis. Genomics, 99, 323-329.

Costello, M. J., 1993. Reviews of methods to control sea lice (Caligidae: Crustacea) infestations on salmon (*Salmo salar*) farms. In Boxshall, G. A., and Defaye, D. (Eds.), *Pathogens on wild and farmed fish: Sea lice,* (pp. 219-254), West Sussex: Ellis Horwood Limited.

Costello, M. J., 2006. Ecology of sea lice parasitic on farmed and wild fish. Trends in Parasitology, 22(10), 475-483.

Danecek, P., Auton, A., Abecasis, G., et al., 2011. The variant call format and VCFtools. Bioinformatics, 27, 2156-2158.

Davey, J. L., and Blaxter, M. W., 2011. RADseq: Next-generation population genetics. Briefings in Functional Genomics, 9, 416–423.

Dawson, L. H. J., Pike A. W., Houlihan, D. F., et al., 1998. Effects of salmon lice *Lepeophtheirus salmonis* on sea trout *Salmo trutta* at different times after seawater transfer. Disease of Aquatic Organisms, 33, 179-186.

Denholm, I., Devine, G. J., Horsberg, T. E., et al., 2002. Analysis and management of resistance to chemotherapeutants in salmon lice, *Lepeophtherius salmonis* (Copepoda: Caligidae). Pest Management Science, 58, 528-536.

Etter, P. D., Bassham, S., Hohenlohe, P. A., et al., 2011. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Molecular Methods for Evolutionary Genetics, 772, 157-178.

Excoffier, L., and Lischer, H. E. L., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources, 10, 564-567.

Finstad, B., Bjørn, P. A., Grimnes, A., et al., 2000. Laboratory and field investigations of salmon lice [*Lepeophtheirus salmonis* (Krøyer)] infestation on Atlantic salmon (*Salmo salar* L.) post-smolts. Aquaculture Research, 31, 795-803.

Foll, M., and Gaggiotti, O. E., 2008. A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics, 180, 977-993

Gagnaire, P. A., and Gaggiotti, O. E., 2016. Detecting polygenic selection in marine populations by combining population genomics and quantitative genetics approaches. Current Zoology, 62(6), 603-616.

Glover, K. A., Stølen, Å. B., Messmer, A., et al., 2011. Population genetic structure of the parasitic copepod *Lepeophtheirus salmonis* throughout the Atlantic. Marine Ecology Progress Series, 427, 161-172.

Gonzalez, E. B., and de Boer, F., 2017. The development of the Norwegian wrasse fishery and the use of wrasses as cleaner fish in the salmon aquaculture industry. Japanese Society of Fisheries Science, 83, 661-670.

Grimnes, A., and Jakobsen, P. J., 1996. The physiological effects of salmon lice infestation on post-smolt of Atlantic salmon. Journal of Fish Biology, 48, 1179-1194.

He, J., Zhao, X., Laroche, A., et al., 2014. Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. Frontiers in Plant Science, 5, 1-8.

Hedgecock, D., Barber, P. H., and Edmands, S., 2007. Genetic approaches to measuring connectivity. Oceanography, 20, 70–79.

Herrera, S., Reyes-Herrera, P. H., and Shank, T. M., 2015. Predicting RAD-seq marker numbers across the eukaryotic tree of life. Genome Biology and Evolution, 7(12), 3207-3225.

Heuch, P. A., Bjørn, P. A., Finstad, B., et al., 2005. A review of the Norwegian ‘‘National action plan against salmon lice on salmonids’’: the effect on wild salmonids. Aquaculture, 246, 79-92.

Heuch, P. A., Nordhagen, J. R., and Schram T. A., 2000. Egg production in the salmon louse [*Lepeophtheirus salmonis* (Krøyer)] in relation to origin and water temperature. Aquaculture Research, 31, 805-814.

Holderegger, R., Kamm, U., and Gugerli, F., 2006. Adaptive vs. neutral genetic diversity: implications for landscape genetics. Landscape Ecology, 21, 797-807.

Imsland, A. K., Reynolds, P., Eliassen, G., et al., 2014. The use of lumpfish (*Cyclopterus lumpus* L.) to control sea lice (*Lepeophtheirus salmonis* Krøyer) infestations in intensively farmed Atlantic salmon (*Salmo salar* L.), Aquaculture, 424, 18-23.

Jacobs, A., De Noia, M., Præbel, K., et al., 2018. Genetic fingerprinting of salmon louse (*Lepeophtheirus salmonis*) populations in the North-East Atlantic using a random forest classification approach. Scientific Reports, 8:1203.

Johnsen, I. A., Asplin, L. C., Sandvik, A. D., et al., 2016. Salmon lice dispersion in a northern Norwegian fjord system and the impact of vertical movements. Aquaculture Environment Interactions, 8, 99-116.

Johnson, S. C., and Albright, L. J., 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.

Jombart, T., and Ahmed, I., 2011. Adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics, 27, 3070-3071.

Jones, A., and Beamish, R., 2011. *An Integrated Approach to Understanding Parasite Abundance and Distribution*. West Sussex: John Wiley & Sons, Inc.

Jones, M. W., Sommerville, C., and Bron, J., 1990. The histopathology associated with the juvenile stages of *Lepeophtheirus salmonis* on the Atlantic salmon, *Salmo salar* L. Journal of Fish Diseases, 13(4), 303-310.

Jones, M. W., Sommerville, C., and Wootten, R., 1992. Reduced sensitivity of the salmon louse, *Lepeophtheirus salmonis*, to the organophosphate dichlorvos. Journal of Fish Diseases, 15(2), 197-202.

Kabata, Z., 1974. Mouth and mode of feeding of Caligidae (Copepoda), parasites of fishes, as determined by light and scanning microscopy. Journal of the Fisheries Research Board of Canada, 31(10), 1583-1588.

Krkosek, M., Revie, C. W., Gargan, P. G., et al., 2013. Impact of parasites on salmon recruitment in the Northeast Atlantic Ocean. Proceedings of the Royal Society B, 280:20122359

Laporte, M., Pavey, S. A., Rougeux, C., et al., 2016. RAD sequencing reveals within-generation polygenic selection in response to anthropogenic organic and metal contamination in North Atlantic Eels. Molecular Ecology, 25, 219-237.

Liaw, A., and Wiener, M., 2002. Classification and Regression by *randomForest*. R News 2(3), 18-22.

Lischer, H. E. L., and Excoffier, L., 2012. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics, 28, 298-299.

Liu, Y., and Bjelland, H. V., 2014. Estimating costs of sea lice control strategy in Norway. Preventive Veterinary Medicine, 117, 469-477.

Ljungfeldt, L. E. R., Quintela, M., Besnier, F., et al., 2017. A pedigree-based experiment reveals variation in salinity and thermal tolerance in the salmon louse, *Lepeophtheirus salmonis*. Evolutionary Applications, 10, 1007-1019.

Lynch, M., and Milljgan, B. G., 1994. Analysis of population genetic structure with RAPD markers. Molecular Ecology, 3, 91-99.

Meirmans, P. G., and Van Tienderen, P. H., 2004. Genotype and genodive: two programs for the analysis of genetic diversity of asexual organisms. Molecular Ecology Notes, 4, 792-794.

Messmer, A. M., Leong, J. S., Rondeau, E. B., et al., 2018. A 200K SNP chip reveals a novel Pacific salmon louse genotype linked to differential efficacy of emamectin benzoate. Marine Genomics, 40, 45-57.

Morin, P. A., Luikart, G., Wayne, R. K., et al., 2004. SNPs in ecology, evolution and conservation. Trends in Ecology and Evolution, 19(4), 208-216.

Morton, A., Routledge, R., McConnel, A., et al., 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.

Narasimhan, V., Danecek, P., Scally, A., et al., 2016. BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics, 32(11), 1749-1751.

Narum, S. R., and Hess, J. E., 2011. Comparison of F_{ST} outlier tests for SNP loci under selection. Molecular Ecology Resources, 11, 184-194.

Nolan, D. T., Reilly, P., Bonga, S. E. W., 1999. Infection with low numbers of the sea louse *Lepeophtheirus salmonis* induces stress-related effects in postsmolt Atlantic salmon (*Salmo salar*). Canadian Journal of Fisheries and Aquaculture Sciences, 56(6), 947-959.

Nolan, D. V., and 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.

Olaussen, J. O., 2018. Environmental problems and regulation in the aquaculture industry. Insights from Norway. Marine Policy, 98, 158-163.

Overton, K., Dempster, T., Oppeldal, F., et al., 2018. Salmon lice treatment and salmon mortality in Norwegian aquaculture: a review. Reviews in Aquaculture, 1-20.

Palumbi, S. R., 2003. Population genetics, demographic connectivity, and the design of marine reserves. Ecological Applications, 13(1), 146-158.

Pavey, S. A., Gaudin, J., Normandeau, E., et al., 2015. RAD sequencing highlights polygenic discrimination of habitat ecotypes in the panmictic American eel. Current Biology, 25, 1666-1671.

Pike, A. W., 1989. Sea Lice – major pathogens of farmed Atlantic salmon. Parasitology Today, 5, 291-297.

Pike, A. W., and Wadsworth, S. L., 1999. Sealice on salmonids: their biology and control. Advances in Parasitology, 44, 233-337.

Powell, A., Treasurer, J. W., Pooley, C. L., et al., 2018. Use of lumpfish for sea-lice control in salmon farming: challenges and opportunities. Reviews in Aquaculture, 10(3), 683-702.

Powell, M. D., Reynolds, P., and Kristensen, T., 2015. Freshwater treatment of amoebic gill disease and sea-lice in seawater salmon production: Considerations of water chemistry and fish welfare in Norway. Aquaculture, 448, 18-28.

Rochette, N. C., Rivera-Colón, A. G., and Catchen, J. M., 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. bioRxiv The Preprint Server for Biology, retrieved from https://doi.org/10.1101/615385

Salama, N., Collins, C., Fraser, J., et al., 2013. Development and assessment of a biophysical dispersal model for sea lice. Journal of Fish Diseases, 36(3), 323-337.

Samsing, F., Oppedal, F., Dalvin, S., et al., 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, 1841-1851.

Stien, A., Bjørn, P. A., Heuch, A. H., et al., 2005. Population dynamics of salmon lice *Lepeophtheirus salmonis* on Atlantic salmon and sea trout. Marine Ecology Progress Series, 290, 263-275.

Sylvester, E. V. A., Bentzen, P., Bradbury, I. R., et al., 2018. Applications of random forest feature selection for fine-scale genetic population assignment. Evolutionary Applications, 11, 153-165.

Thorstad, E. B., Todd, C. D., Uglem, I., et al., 2015. Effects of salmon lice *Lepeophtheirus salmonis* on wild sea trout *Salmo trutta* - A literature review. Aquaculture Environment Interactions, 7, 91-113.

Tjensvoll, K., Glover, K. A., and Nylund, A., 2006. Sequence variation in four mitochondrial genes of the salmon louse *Lepeophtheirus salmonis.* Disease of Aquatic Organisms, 68, 251-259.

Todd, C. D., Walker, A. M., Ritchie, M. G., et al., 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, 1176-1190.

Todd, C. D., Walker, A. M., Wolff, K., et al., 1997. Genetic differentiation of populations of the copepod sealouse *Lepeophtheirus salmonis* (Krøyer) ectoparasitic on wild and farmed salmonids around the coasts of Scotland: Evidence from RAPD markers. Journal of Experimental Marine Biology and Ecology, 210, 251-274.

Torrissen, O., Jones, S., Asche, F., et al., 2013. Salmon lice – impact on wild salmonids and salmon aquaculture. Journal of Fish Diseases, 36, 171-194.

Treasurer, J. W., Wadsworth, S., and Grant, A., 2000. Resistance of sea lice, *Lepeophtheirus salmonis* (Krøyer), to hydrogen peroxide on farmed Atlantic salmon, *Salmo salar*. Aquaculture Research, 31, 855-860.

Tully, O., and Nolan, D. T., 2002. A review of the population biology and host–parasite interactions of the sea louse *Lepeophtheirus salmonis* (Copepoda: Caligidae). Parasitology, 124, 165-182.

Turelli, M., and Barton, N. H., 1990. Dynamics of polygenic characters under selection. Theoretical Population Biology, 38, 1-57.

Waples, R. S., 1998. Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. Journal of Heredity, 89, 438-450.

Ward, R. D., Woodwark, M., and Skibinski, D. O. F., 1994. A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. Journal of Fish Biology, 44, 213-232

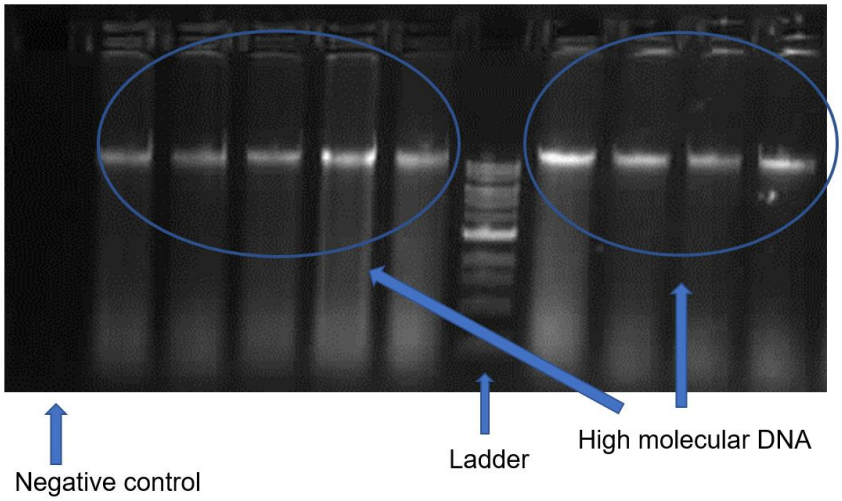
Wootten, R., Smith, J. W., and Needham, E. A., 1982. Aspects of the biology of the parasitic copepods *Lepeophtheirus salmonis* and *Caligus elongatus* on farmed salmonids, and their treatment. Marine Fish Diseases, 81(3), 185-197.

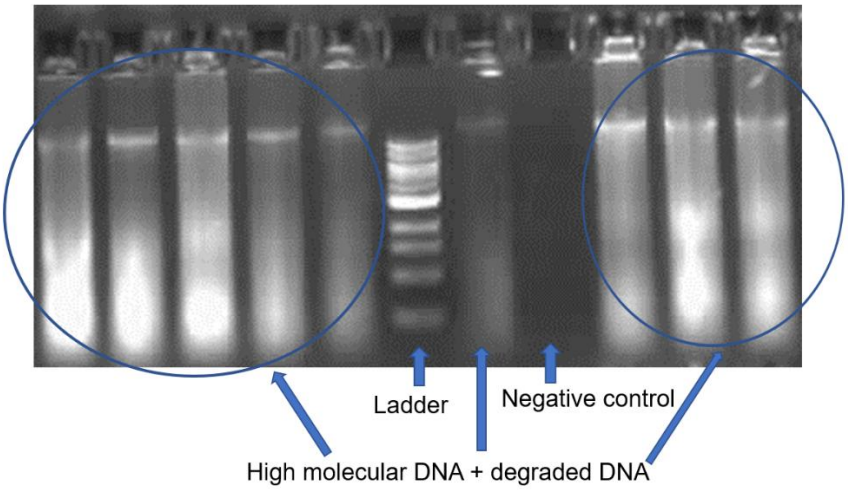
Wright, D. W., Oppedal, F., and Dempster, T., 2016. Early-stage sea lice recruits on Atlantic salmon are freshwater sensitive. Journal of Fish Diseases, 39, 1179-1186.

Wu, T. D., and Nacu, S., 2010. Fast and SNP tolerant detection of complex variants and splicing in short reads. Bioinformatics, 26, 873-881.

Appendices

Appendix 1. Results of DNA quality check using gel electrophoresis.





DNA extracted from salmon lice for this study was either HM (high molecular, good quality DNA: above gel picture) or HM smear (high molecular, good quality DNA containing some degraded DNA: below gel picture). The gel pictures include ladder (channels with multiple bands) and negative control (channels that are all black).

Appendix 2. PicoGreen dsDNA concentration assay, user manual.

This manual is based on the manual provided by Thermo Fisher Scientific, and especially designed for the use at laboratories at NFH. Edited by Julie Bitz-Thorsen, NFH.

PicoGreen dsDNA Concentration Assay

Using this kit will provide you with the concentrations of double-stranded DNA for your samples. Knowing the concentrations are vital for the dilution step of library building.

What is in a kit?

- PicoGreen dsDNA quantification reagent in DMSO. NB! Protect from light always.
- Lambda DNA standard (100 µg/ml stock solution)
- 20xTE Buffer

Preparations

The solutions prepared must be freshly made each time for this assay. Standard, blank, and sample concentrations must be measured using the same solutions.

1) Prepare and fill in a plate sheet with your samples to keep an overview ('1 PicoGreen plate sheet'. (This can be prepared the day before). Remember that you need duplicates or triplicates for this assay.
2) Start the spectrofluorometer on the 4th floor (C458) so it is ready to go!
    a. Open the SoftMax Pro software (version 4.8) and create Assay. Choose Nucleic Acids and PicoGreen Fluorescence. The working window is now be called PicoGreen dsDNA Quantification. Choose plate setup, read type as Fluorescence and top read (tick off bottom read). Excitation 480nm, emission 520 nm.
3) Use the excel sheet '2 PicoGreen calculus' to calculate the amount of 1xTE Buffer and PicoGreen work solution you need (Please download before editing). The solutions need to be freshly prepared each time you run a PicoGreen.
    a. Prepare the 1xTE Buffer. NB! Be sure to have enough buffer because it must be the same TE Buffer you use for/in standards, blanks and all samples.
    b. Prepare PicoGreen working solution. Cover with tin foil to keep away from light!
    c. Prepare the lambda standard 2 µg/ml stock solution and dilute the standards A, B, C and D.

Let's go!

1) Add 3x100 µl per standard in the plate.
2) Add 3x100 µl 1xTE Buffer as blanks.
3) Add 99 µl 1xTE Buffer to all sample wells.
4) Add 1 µl of the sample to the corresponding well.
5) Add 100 µl PicoGreen working solution to all wells. NB! Add only to one plate at a time.
6) Incubate 2-5 min (more or less the time it takes to walk to the 4th floor).
7) Go the spectrofluorometer and read the concentrations of your plate.
8) Save the read as a Pro Data File (.pda) named YYMMDD_picogreen_species/projectname_yourname and also export as a text file.
9) Repeat step 6 and 7 for your other plates.

Data treatments

Use the excel sheet '3 PicoGreen concentrations' to obtain the standard curve and to subtract the blank from your measured sample concentrations. From this you can calculate the DNA concentrations in ng/µl for each sample.

Appendix 3. Simple description of RAD library building and sequencing protocol.
(The original protocol of RAD library building is from Benestan et al. (2015), and it was modified for the use in the Population Genetic Lab at NFH, UiT.)

1) Dilution

Start with approximately 500ng of DNA from each sample (96 samples for 1 RAD lane). Bring the volume of each sample to 40µl with RNAse free water, and transfer each sample into a 96-well plate. Save the sample and its concentration information in a RAD lane form, as an Excel file.

2) Digestion with *PstI* enzyme

For each sample, add 10µl of master mix (includes H_2O 4.5µl, Cutsmart Buffer (10X) 5.0µl, *PstI* enzyme 0.5µl). Put a plastic tape lid on the well plate and incubate the plate in PCR machine: 37°C for 60min and at 65°C for 20min to inactivate the enzyme.

3) Barcoding

Add 2µl of the appropriate barcoded P1 RAD adapter (50nM) to each well in the sample plate using multi-channel pipette. Each sample has a different adaptor. Add adapter information to the RAD lane form.

4) Ligation

Add 8µl of master mix (includes H_2O 5.9µl, Cutsmart Buffer (10X) 1.0µl, rATP (100mM) 0.6µl, and T4 DNA Ligase 0.5µl) to each well. Give quick spin and vortex. Incubate the well plate in PCR machine: 20°C for 60min, and at 65°C for 20min to inactivate the enzyme.

5) Pooling

Multiplex the 12 samples that are to be sequenced together in the same library. Perform a quick vortex. You will get a 720µl final volume for each pool (60µl x 12 samples).

6) Sonication

Sonicate the multiplexed samples to produce an average fragment size of 500bp. Bioruptor needs to be cooled down to 4°C prior to use. Divide each sample into 3 x 100µl. Sonication setting: ON time 30sec, OFF time 90sec, 6 cycles. After sonication, combine the 3 x 100µl for each sample into new 1.5 ml Eppendorf tubes.

7) Drying

Dry samples for approximately 2h10min using vacuum centrifuge until there is only a small drop of liquid left. Do not over dry it as it will create problem in re-suspension in Elution Buffer (QIAGEN).

8) Volume adjusting

Complete volume to 100µl with Elution Buffer.

9) Fragment size selection 400-600bp

Fragment size selection using magnetic beads.

10) Cut and blunt the fragment

Add 6µl of master mix (includes Blunting Buffer (10X) 2.5µl, dNTP mix (1mM) 2.5µl, Blunting Enzyme Mix 1.0µl) to 19µl of sample. Incubate it at 20℃ (room temperature) for 60min.

11) Volume adjusting

Add 25µl of Elution Buffer to the reaction products.

12) Purification with beads

Purification using magnetic beads.

13) Add A-overhangs to the fragments

Add 8µl of master mix (includes NE Buffer 2 (10X) 5.0µl, dATP (10mM) 1.0µl, and Klenow Fragment 2.0µl to 42.0µl of sample. Incubate at 37℃ for 60min. Use Thermomixer with lid.

14) Purification with beads

Purification using magnetic beads.

15) Ligation of the P2 adapter to fragments

Add 7µl of master mix (includes NE Buffer 2 (10X) 5.0µl, P2 RAD adapter (10µM) 1.0µl, rATP (100mM) 0.5µl, and T4 DNA Ligase 0.5µl to 43.0µl of sample. Incubate at 20℃ for 30min.

16) Purification with beads

Purification using magnetic beads. After the purification, measure DNA concentration with Qubit and use approximately 40ng of DNA for PCR.

17) PCR

Make PCR mix containing; P1 Adapter Primer (10µM) 4.0µl, P2 Adapter Primer (10µM) 4.0µl, Phusion Master Mix (X2) 50.0µl, and sample with DNA concentration of 40ng/42.0µl. Use RNAse free water to adjust the volume and DNA concentration.

PCR cycling conditions

Step 1: 98℃ for 30sec

Step 2: {98℃ for 10sec, 65℃ for 30sec, 72℃ for 30sec} 21X

Step 3: 72℃ for 5min, hold at 10℃

18) Fragment size selection by Blue Pippin

Final fragment size selection to get fragment size of 350-600bp, using Blue Pippin. Use the manual provided by Sage Science, Inc.

19) Average fragment size assessment

Average fragment size assessment using Agilent 2100 Bioanalyzer system. Use the manual provided by Agilent Technologies, Inc.

20) RAD Sequencing

Sequence the prepared RAD library lanes. For this study, 3 lanes were sequenced on a HiSeq 4000 using Illumina system at Novogene (HK) Company Limited (Hong Kong).

Web links

Agilent 2100 Bioanalyzer, user's guide for molecular Assays
https://ipmb.sinica.edu.tw/microarray/index.files/Agilent%202100%20Bioanalyzer%20user%20guide.pdf

Barents Watch
https://www.barentswatch.no

Diagenode Bioruptor Pico Sonication System, user manuals
https://www.diagenode.com/files/products/shearing_technology/bioruptor/Bioruptor_pico_cooler_manual.pdf

LiceBase, salmon lice genomic database
https://licebase.org/organism/Lepeophtheirus/salmonis

Marine Institute, life cycle of the salmon louse
https://www.marine.ie/Home/site-area/areas-activity/aquaculture/sea-lice/life-cycle-salmon-louse

National Hurricane Center, USA, Latitude/Longitude Distance Calculator
https://www.nhc.noaa.gov/gccalc.shtml

QIAGEN DNeasy Blood and Tissue Kits, user manuals
https://www.qiagen.com/no/shop/sample-technologies/dna/genomic-dna/dneasy-blood-and-tissue-kit/#resources

R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria
https://www.R-project.org/

Sage Science Blue Pippin DNA Size Selection System, operations manuals
http://www.sagescience.com/wp-content/uploads/2018/10/BluePippin-Operations-Manual-460013-Rev-G.pdf

Thermo Fisher Scientific Qubit 4 Fluorometer, user manuals
https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0017209_Qubit_4_Fluorometer_UG.pdf