The effect of removing a natural migration barrier on a previously isolated population of Brown trout (*Salmon trutta L.*) – Kvåsfossen, Norway.

SÅVE-MARIA RØRNES SVEINSON

SUPERVISORS

Marte Sodeland
Trond Rafoss

University of Agder, 2023
Faculty of Engineering and Science
Department of Natural Science

Master

University of Agder
Faculty of Engineering and Science
Department of Natural Science
Gimlemoen 25
4604 Kristiansand
http://www.uia.no

Abstract

In the years of 1970s and 1980s acidic rain threatened coastal freshwater fish populations in European rivers. Several fish populations in western and southern Norway were particularly affected including the Norwegian brown trout (*Salmon trutta* L.). Liming of lakes and rivers has had positive effects on re-establishment of fish populations in many Norwegian rivers. Migration barriers, natural or man-made is another challenge for anadromous coastal fish populations. By completely or partially isolating upstream populations from downstream migration, these barriers hinder gene flow and thereby reduce genetic diversity in the upstream populations. In 2014 a fish ladder was built in Kvåsfossen a waterfall in the river Lygna in Lyngdal municipality, Agder, Norway. Kvåsfossen completely isolated the upstream trout population from downstream migration before establishment of the fish ladder. This study has assessed the effects of removing a natural migration barrier on trout in the river Lygna by comparing genomic DNA from 2016 and 2022 using double digest RAD and Illumina sequencing. The results showed decreased genetic structure and genetic differentiation among sample locations in 2022, but no significant changes in genetic diversity and inbreeding. The decreased genetic structure and differentiation confirms migration from downstream to upstream locations after removal of the migration barrier.

Sammendrag (Norwegian)

I 1970 og 1980 årene truet surnedbør Europeiske populasjoner av kystnær ferskvannsfisk til nær utryddelse. Mange fiskepopulasjoner i sør og vest Norge ble særlig berørt inkludert den norske brun ørreten (*Salmon trutta* L.). Kalking av innsjøer og elver har hatt positiv effekt på reetableringen av fiskepopulasjoner i mange norske elever. Migrasjons barrierer både naturlige og menneskeskapte er enda en faktor som truer populasjoner av anadrom kystnær fisk, ved å helt eller delvis isolere populasjoner oppstrøms fra nedstrøms migrasjon. Migrasjons barrierer hindrer gen flyt og reduserer den genetiske diversiteten i populasjoner oppstrøms for barrieren. I 2014 ble det bygget en laksetrapp Kvåsfossen, en foss som er del av elven Lygna i Lyngdal kommune, Agder, Norge. Ørret populasjonen over Kvåsfossen var totalt isolert fra nedstrøms migrasjon før innføringen av laksetrappen. Denne studien vurderer effekten fjerning en naturlig migrasjons barriere har hatt på den tidligere isolerte ørret populasjonen i Lygna. Dette gjøres ved å sammenligne geonomisk DNA fra 2016 og 2022 ved hjelp av double digest RAD og Illumina sekvensering. Resultatene viste tydelig lavere genetisk struktur og genetisk forskjell mellom prøve lokasjonene i 2022, men ingen signifikant forskjell i genetisk diversitet og innavl. Lavere genetisk struktur og genetisk forskjell mellom lokasjonene bekrefter migrasjon av nedstrøms ørret til oppstrøms lokasjoner etter fjerning av migrasjons barrieren.

Table of Contents

Preface

First, I will especially thank my main supervisor Marte Sodeland at UiA for all help when I have been stuck in R, all useful suggestions, answering all my silly questions and for all the helpful feedback during writing this master thesis. Also thank you for handling the SNP detection. A big thank you to Ida Kristine Mellerud at the Marine Institute of research for the DNA extraction and ddRAD library preparations.

Thank you to Trond Rafoss for helping with the preliminary planning of the study and field work. Thank you, Ragnvald Andersen, for providing information on the study area and supervising the electrofishing and sampling.

Thank you to my fellow master writing friends for support and encouragement. Thank you to my boyfriend Ole-Andre Lien for being there when the day is hard, cooking dinner for me and proofreading, you are the best. Thank you to all my wonderful friends.

Finally, thank you mum and dad for always believing in me.

Kristiansand, May 2023 Såve-Maria Rørnes Sveinson

1. Introduction

Population genomics is defined as the simultaneous study of numerous loci and genome regions to better understand evolutionary processes such as mutation, genetic drift gene flow and natural selection that influence variation across genomes and populations (Black IV et al., 2001; Luikart et al., 2003) Population genomics has revolutionized molecular ecology making it possible to identify thousands to millions of single nucleotide polymorphisms (SNPs) that allows for studying both genome-wide and allele-specific patterns of diversity in both model and non-model organisms (Morin et al., 2004; Seeb et al., 2011; Stapley et al., 2010). The revolution in genomics has given biologists the inspiration to support genome analyses in advanced conservation beyond what was possible in the past (Kardos et al., 2021).

Interestingly, freshwater fish have a more profound structure when it comes to genetic diversity and population structure than marine species (McCusker & Bentzen, 2010). This comes from higher level of restriction in gene flow among freshwater species due to the many physical barriers in freshwater systems, compared to marine environments where physical barriers are less profound (Waples, 1998). Genetic diversity has several ecological impacts on populations such as productivity, growth and the populations resilience against both natural and human disturbances (Hughes et al., 2008). Genetic diversity is described as the foundation of biodiversity providing the basic material for evolution to act (Fisher, 1999). Analyzing and understanding the impacts genetic diversity has on populations provides important knowledge about ecology and evolution that can be used to preserve and manage sustainable freshwater populations. Anadromous fishes are species that may utilize both freshwater and marine habitats during their life cycle. These fish spawn in rivers and streams but may migrate to sea to grow and mature (Schiewe, 2013). Migration patterns in anadrome fish are affected by temperature, waterflow and migration barriers (B. Jonsson & Jonsson, 2010; N. Jonsson & Jonsson, 2002). One of the biggest threats to biodiversity is fragmentation which leads to small isolated population of both plants and animals with reduced genetic diversity caused by a combination of reduced gene flow, increased genetic drift and inbreeding. However, this can depend on several factors that are yet poorly understood such as organism group, habitat type, history-life traits and time since fragmentation (Schlaepfer et al., 2018). Migration of anadrome fish can be an important attribute to gene flow among populations and subpopulations increasing genetic diversity and

resilience within small and isolated populations (Hansen & Mensberg K-L.D., 1998; Waples, 1998; Westgaard et al., 2016).

1.1 Resident and anadromous trout

Introducing the polymorphic salmonid species Brown trout (*Salmon trutta* L.), from here on referred to as trout. With natural habitats distributed all over Europe, North Africa, the United Kingdom and western Asia (Klemetsen et al., 2003), trout is of great commercial importance to European economy through fishery resources and tourism (ICES, 2020). Trout's ability to adapt to new environments and its reputation as good food has made trout popular beyond Europe and across the continents (Klemetsen et al., 2003). Trout is part of the salmonid family, which consists of the three sub families the *Salmonidae* (trout, chars, and salmons), the *Thymallinae (*graylings), and the *Coregoninae* (withefish and ciscos). The salmon family diverged from the pike family (*Esocidae*) during 110-150 Mya (Near et al., 2016), and we can date their last common ancestor to about 58-63 Mya (Crête-Lafrenière et al., 2012). The particular event of whole genome duplication of salmonids occurred between 88-93 Mya (Macqueen & Johnston, 2014) (Berthelot et al., 2014) which eventually resulted in the different genotypical and phenotypical salmonid species that are known today.

Natural selection determines the genotype. Genotype is decided by genes inherited from the parents. Phenotype is how the fish looks and acts like and is the sum of both inherited genes and environmental impact. Fish can be genetically similar but phenotypically different. Trout can have great variation in both appearance (colours, spots) and behaviour (resident, migratory) (Heggenes, 2016). This makes trout a phenotypically plastic species and about 50 sub-species of trout have been described (Heggenes, 2016).

Trout have several life strategies and can be classified by three different life history forms: resident land-locked trout, resident trout and anadromous trout (Klemetsen et al., 2003; Pakkasmaa & J., 2001). Land-locked trout is isolated populations of trout with restricted gene flow from other populations resulting in populations with low genetic diversity. This separates the land-locked trout genetically from the two other forms (Hindar et al., 1991). Resident trout and anadromous trout co-exists in the same watercourses and can descended from the same parents but choose different life strategies. Resident trout stays in their natal rivers' trough their whole life, and every year they return to spawn in their natal tributaries (Klemetsen et al., 2003). Anadromous trout commonly known as sea trout, migrate to the sea from their natal

rivers and streams. During this migration they feed in the ocean until sexual maturity, before returning to their natal rivers were spawning happens (Elliott, 1994).

1.2 Migration behaviour

Coexistence between migratory and resident individuals in the same populations is a common life plasticity among fishes (Jonsson & Jonsson, 1993). The urge an individual has to migrate is controlled by genetic and environmental factors, such as competition for food, space and the opportunities for reproduction (Charles et al., 2004). Individuals that choose to migrate typically grow larger and possess a higher reproduction potential at the cost of lower survival than resident individuals. The choice to migrate or not can depend on the correlating processes growth rate and metabolic rate. Fish mature when cost of energy for maintenance is similar to the energy intake. Alternatively, delaying maturation allows fish to migrate to richer feeding habitats increasing their fitness before maturation (Jonsson & Jonsson, 1993). Females migrates at higher rates than males among resident fish. Probably because female fitness and reproduction success increases with body size (Bagenal, 1973; Fleming & Gross, 1991). A Norwegian study (Jonsson & Jonsson, 2021) from Imsa River in Norway shows populations of trout isolated by barriers for a period of 25 years, indicating that offspring of resident trout had a lower probability to become anadromous. Suggesting that not only the space factor but also genetic factors is responsible for the migratory decisions of an individual trout (Bernaś et al., 2021).

1.3 Migration barriers

Migration barriers can occur as natural waterfalls but also in the form of hydroelectric dams that are completely or partially isolating the upstream movement of trout, which is restraining gene flow, reducing genetic diversity and reproductivity of the trout (Prodöhl et al., 2019). In lower parts of rivers one can find both resident and anadromous trout, but due to expansion of artificial barriers prevalence of anadromous trout has been reduced in lower segments of European rivers (Ferguson et al., 2019). Lower costal river streams may not be of great importance for the resident trout populations upstream but serves as significantly important spawning habitat for anadromous trout (King et al., 2020). Continuous dividing of habitat by both natural and man-made barriers impacts rivers by habitat fragmentation (Jones

et al., 2019). Fragmentation of habitats can from a genetic perspective have a range of different effects on fish populations, causing reduction in both count and effective population size, increase inbreeding that will lead to lower levels of genetic diversity, and increased genetic structuring (Frankham et al., 2017; Schlaepfer et al., 2018). Small populations are in general at higher risk of damaging effects from genetic bottlenecks, inbreeding and genetic drift, all leading to loss of genetic diversity and inbreeding depression (Vandewoestijne et al., 2008). In the long run the loss of genetic diversity can lead to poorer fitness among individuals and are increasing the threat of population extinction on a local scale (Vandewoestijne et al., 2008). The risk of genetic drift and inbreeding effects on a population can be prevented by gene flow from individuals from other populations and also mutations (Ho & Larson, 2006). Isolated fish populations have been found to have reduced genetic diversity, e.g., coastal Cutthroat trout (*Oncorhynchus clarki)* populations isolated above untraversable waterfalls were highly different from downstream populations, having substantially lower genetic diversity and heterozygosity (Whiteley et al., 2010). Furthermore, anadromous Rainbow trout (*Oncorhynchus mykiss*) from downstream dams in Colombia River had higher diversity than resident population of rainbow trout upstream (Winans et al., 2018). Because resident and anadrome species have a less profound population structure than marine species due to restricted gene flow (Waples, 1998), extinction of resident or anadrome species results in more total genetic diversity lost. This is due to smaller population sizes and higher degree of genetic divergence among resident and anadrome specie (Ryman et al., 1995).

1.4. Trout in Southern Norway and Lygna

In the 1970 and 1980s acidic rain caused by emissions of nitrogen and sulphur caused great reduction in the biological diversity in European lakes and rivers (Menz & Seip, 2004). Coastal freshwater fish in southern and western Norway were particularly affected, the acidic rain caused several salmon and other river fish populations to go extinct or become severely threatened (Rosseland et al., 1986). Liming of lakes and rivers has had positive effects and made fish populations re-establish in many rivers in both south-and western Norway, salmon have also reappeared in many rivers (Clair & Hindar, 2005; Sandlund & Hesthagen, 2011). Lygna is a river system located in southern Norway and has been limed since 1991. There are now two liming facilities along the watercourse, Rossevatn, limed since 1991 and Gyseland, limed since 2000. Before liming of the watercourse, the trout population were threatened by

extinction due to acidification and the salmon (*Salmon salar* L.) population were already extinct (Vikøyr et al., 1989). Today Lygna inhabits rich salmon and trout populations serving as a popular destination for trout and salmon fishing. Liming is a well-known method for improving environmental condition in streams, yet another problem threatened the trout population in Lygna. The waterfall Kvåsfossen made an impassable barrier for trout and salmon isolating the upstream population from downstream migration. In 2014 a 220-meter-long fish ladder was built inside a tunnel beneath Kvåsfossen for salmon and sea trout. Before the entry of the fish ladder fish from the anadrome stretch were prevented to migrate further up the watercourse. In 2016 several stations upstream and downstream Kvåsfossen were electro fished to gather genetic material on the isolated trout population above Kvåsfossen before downstream sea trout and trout came an interfered. Shortly after the establishment of the fish ladder in Kvåsfossen, it was discovered that there was yet another obstacle in Gysfossen. This restrained trout and salmon in further upward migration except on occasions of particular water levels. The obstacle is now cleared (Miljødrektoratet, 2016).

1.5 Study description

This study will examine how removal of a migration barrier has affected genes of the upstream population of trout. The goal is to assess the effect fish ladders may have on trout populations isolated by barriers. To measure the effect from the fish ladder this study will compare genomic DNA from 2016 and 2022. Genomic DNA is obtained from either whole organisms such as microbes or small invertebrates, or organism material such as hair, feathers fish-scales, leaves or in this case tissue from the adipose-fin (Creer et al., 2016). Genomic DNA is the commonly used DNA in such population studies. For the genetic analyses double digest restriction associated DNA (ddRAD) sequencing will be used. This is an inexpensive method and an important source for both population history and genotype-phenotype association information (Peterson et al., 2012). The ddRAD libraries are sequenced with high throughput Illumina sequencing, allowing for cost-effective genotyping of a high number of single nucleotide polymorphisms SNPs and individuals. SNPs only focuses on the nucleotides that are known to vary across individuals, making it possible to differ between closely related individuals with many invariant nucleotides and relatively few differences (Freeland, 2020).

1.6 Aim

The main purpose of the study will be to assess the effect of removing a migration barrier on trout in the river Lygna. This will be done by comparing genomic data from 2016 with new data from field work late summer 2022, using population genomics approaches to assess gene flow between up- and downstream locations after establishment of the fish ladder.

Hypothesis 1. Genetic differences between upstream and downstream trout in Kvåsfossen have decreased.

Hypothesis 2. Genetic diversity has increased, and inbreeding has decreased in upstream trout.

2. Materials and methods

2.1 Study area

Trout used in this study were caught from six different sampling locations along the river Lygna. Stations: 1-Lislåna, 2-Kvinemonen, 3-Birkeland, 4-Kvåsfidjan, 5-Kvåsfossen and 6-Nothølen both up- and downstream Kvåsfossen (58° 15.8692´N, 007° 11.3974´E), Lyngdal municipality, Agder, Norway (**Fig. 1**). The river Lygna stretches 82km from Fjotland to Lyndgdal before flowing into Lyngdalsfjorden. Station 2, 3, 5 and 6 were sampled in 2016 and 2017, on station 5 and 6 sea trout were also sampled. In 2022 only the upstream stations 1, 2, 3 and 4 were sampled. All sampling stations lies close to spawning hollows with gravel bottom and deeper water levels which are common spawning habitats for trout and salmon (Greeley, 2011), except station 5 where trout was caught in the fish-ladder. The downstream stations were included to compare the genetic difference between the up- and downstream population. Furthermore, an out-group (Nesheim) from a separate coastal stream was included in the analysis to emphasise the genetic distance within the trout population in Lygna. The Nesheim watercourse is located on the peninsula Lista in Farsund municipality (58° 04.6281´N, 006° 40.0170´E), approximately 38km in overhead line from Kvåsfossen. Kvåsfossen is a significant waterfall due to its protected waterfall status and is also the location of one of the Norwegian National Wild Salmon centres (**Fig. 2**).

2.2 Capturing and sampling

A total of 75 individual trout were sampled upstream Kvåsfossen in the late summer of 2022. These were to be compared with 57 individuals sampled in 2016 and 2017, including the outgroup Nesheim. Electrofishing was performed in late summer due to the requirement of low water flow to conduct sampling. Trout was captured by electrofishing a standardized method used for catching fish for release elsewhere, registration of stocks, collection of genetic material and removal of unwanted fish species (Malcolm et al., 2019). In this study the electrofishing was used to collect genetic material in the form of standardised tissue samples for genetic analyses. Electrofishing has been conducted in Lygna by Ragnvald Andersen, Rune Eikeland, Alf Kåre Friestad and Trond Rafoss from 2016 until today. The electrofishing was conducted with a portable electric fishing tool with 1500V voltage and adjustable frequency

and pulse length, stunning the fish. The tool has a positive pol (anode) that consist of several oval steel rings attached to a fiberglass rod. In the handle of the tool there is a magnet switch. The negative pol (cathode) is a wire that is dragged after the person carrying the electrofishing tool. The sample location is fished out and sampled once (qualitative electrofishing) (Bohlin et al., 1989). Fish was collected in a bucket and one by one counted, species determined, and body length was measured before the adipose fin was cut for genetic analyse. The adipose-fin samples were carefully marked by number and stored on small separate test tubes with ethanol for further analysis. After finishing sampling, the fish were released back into the river. All trout sampled upstream were 0+ - 1+ to make sure seaward migration had not occurred. The electrofishing tool used in this study was constructed by engineer Steinar Paulsen, Trondheim.

Figure 1. *Sampling location upstream and downstream Kvåsfossen. (Kartverket).*

Figure 2. *Picture of the fish ladder below Kvåsfossen (left) and Kvåsfossen (right).*

2.3 Ethical statement

Electrofishing is a well-established method for capturing and sampling fish. When used correctly electrofishing can be a low impact technique (Snyder, 2003). The potential of short term and long-term injuries and death by electrofishing, shows the importance of the method being conducted under restricted conditions for minimum harm (Snyder, 2003). Mortality induced by electrofishing sometimes occur during the sampling process due to stress from capture (Bayley & Herendeen, 2000; Portt et al., 2006). Snyder (1995) suggested that when injury from electrofishing is a problem and cannot be adequately avoided the technique must be abandoned or severely limited. Whereas Schill & Beland (1995) stated that in comparison to other sampling techniques, electrofishing is a relatively gentle method, and that the injury and mortality inflicted when sampling small parts of a population does not impose a considerable risk to the population.

The adipose fin was previously believed to have no function (Aiello et al., 2016). Recent studies however have proven the adipose fin to be a significant flow sensor and attraction feature (Koll et al., 2020).

2.4 Genetic Analysis

ddRAD sequencing uses two enzymes for digesting and fragmenting DNA. Precise size selection excludes regions surrounded by close or distant RE recognition sites, creating an enhanced library containing only the target size fragments (Peterson et al., 2012). In comparison, traditional RAD sequencing uses only a single enzymes and secondary random fragmentation with a size selection that result in reduced representation libraries (Peterson et al., 2012). The ddRADseq follows the Institute of Marine research ddRADseq protocol and spreads over two days. Fish DNA is highly suitable for this method due to its generally good quality (Taggart et al., 1992). Before executing the protocol, the following requirements must be fulfilled. DNA must be high-quality, preferably un-degraded but also mildly degraded DNA can be used. DNA quality is verified with gel electrophoresis (Jonkers & Sharkely, 2016). Ratio of absorbance must be at 260/280nm to estimate the purity of the DNA and RNA. DNA is considered pure at >1.7 and RNA at 2.0 (Wilfinger et al., 2018). The 260/230 ratio is used as a control measure for nucleic acid purity and values are usually higher than for 260/280 and is expected to range from 2.0-2.2. Lower ratios may indicate contaminants that absorb at 230nm (Arif et al., 2010).

2.4.1 Double digest RAD Protocol and variant detection

Double digest RAD sequencing library preparation was conducted by Ida K. Mellerud at the Institute of Marine Research (Flødevigen, Arendal, Norway). Before ddRAD library preparation DNA was extracted and dissolved in EDTA free buffer i.e., TE buffer which purpose is to make DNA or RNA more soluble while at the same time protecting it from degradation. All samples normalized to 3 ng/µl to increase the accuracy of the sample data. Accuracy during all steps in this process is crucial due to very small reaction volumes.

DAY 1

Step 1. Double RE digest is making the RE-digest master mix. This master mix digests and cuts the DNA using the reagents CutSmart buffer which separates DNA strands before the restriction site enzymes Pst1-HF and Mse1 cuts the target sites of the DNA. Step 2. Ligation of barcoded P1 and P2 adaptors. This process uses enzymes to connect specialized adapters with barcodes unique to each fish to both ends of the DNA fragment.

Step 3. Pooling after indexing. Qiagen MinElute PCR clean-up kit. Assembling all "samples" before size selection clean up.
DAY 1 of the protocol is repeated up to 5x times creating 5 libraries. DAY 2 can do 5 libraries combined.

DAY 2

Step 1. Size selection using Blue Pippin. This system uses colourless, pre-cast gel cassettes to separate and pull-out different size ranges of DNA fragments using electrophoresis. Gel cassettes selecting for 250bp - 550bp are used for the size selection in this analysis. One cassette can run 5x libraries at once.
Step 2. Ampure bead clean-up is a solution containing small magnetic particles which separates DNA by size. Firstly, the beads bind the larger fragments which then are discharged. Secondly, more beads are added in the solution binding the smaller fragment sizes. This is the final library that is pipetted into a tube and labelled (size select, ddRAD name, date)
Step 3. PCR amplification copies/amplifies the target DNA in the final libraries making sure there are enough target DNA for analysis e.g., sequencing or visualization by gel electrophoresis. PCR amplification can make thousands to millions of copies of a particular DNA fragment.
Step 4. Final Ampure bead clean-up after PCR repeats step 2. Final library.
Step 5. Check the library QC on a Bioanalyzer and a Qubit. The Qubit BR kit calculates the average ddRAd library concentration. Bioanalyzer 2100 Expert Software calculates the average base pairs of your library (**Fig. 3**). In this case we want it to be between 250 and 550bp *(Figure 3)*. Finally calculate the nM of the library based on the results from Qubit and Bioanalyzer.

The final libraries contained 96 individuals each, pooling samples across studies. Each library was sequenced by 150bp pair-end sequencing in a single lane on an Illumina Xten machine at BGI Genomics (ww.bgi.com). The sequencing generated an average of 9.8 million reads per individual fish, with a Q30 of 95.3 percent. After sequencing, sequence reads were aligned to the Brown trout genome with the BWA software (Li & Durbin, 2009). The aligned data were analysed in STACKS 2.0 (Catchen et al., 2013) for variant detection. SNPs were designated to the following filtering requirements: minor allele count should be at least 2, the minimum read depth for each loci should be at least 8 to be designated and loci should be present in at least 10% of the individuals from at least one location sample. Finally,

VCFTOOLS (Danecek et al., 2011) was applied to create the final dataset of 8999 loci for further statistical analyses. Out of the 132 trout sampled 106 trout were genotyped *(Table.1)*.

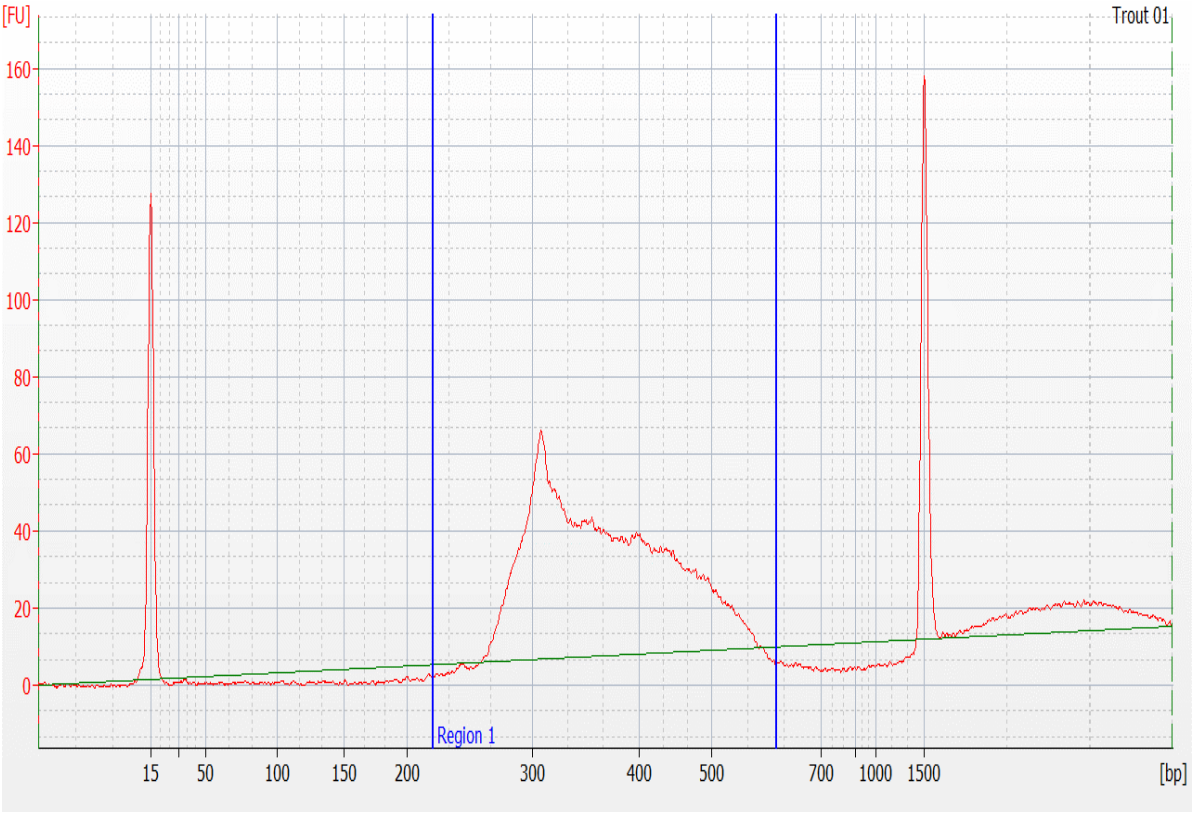


Figure 3. *Region with average base pairs of the library calculated by the Bioanalyzer 2100 Expert Software.*

2.5 Quality control

Before running statistical analyses, it is important to conduct quality control on the data. This is to identify individuals with low quality data and remove them to ensure data of high quality before further analyses (Knaus et al., 2023). The quality control was performed in R Studio (RStudio, 2020) with the SNPRelate (Zheng et al., 2012) and vcfR package (Knaus & Grünwald, 2017). After the quality control 7 individuals was removed (**Appendix 1**). Additionally, 15 individuals were identified as salmon and were removed as well, leaving a final dataset of 84 individual trout.

2.6 Statistical analysis

All data analyses were performed in R Studio (RStudio, 2020), using the Bioconductor software (Huber et al., 2015). Genetic diversity was measured and characterized by heterozygosity H_O and H_E. Heterozygosity within samples were characterized by observed heterozygosity (H_O), and total material of each locus by expected heterozygosity (H_E) using the method of (Nei & Chesser, 1983) and the hierfstat package (Goudet. J. et al., 2022). F_{IS} and pairwise F_{ST} was estimated by the method of (Weir & Cockerham) with the hierfstat package. F_{IS} values were used to estimate the proportion of heterozygosity in each location. To estimate the level of genetic differentiation among sample locations pairwise F_{ST} was applied. F_{ST} values were bootstrapped with a 95% confidence interval to identify significant structuring changes using hierfstat. Inbreeding coefficient F was calculated to measure fixation on alleles in individuals using the SNPRelate package (Zheng et al., 2012). Furthermore, SNPRelate was used to create a principal component analysis (PCA) (Jolliffe & J., 2016). SNPRelate and the Ape package (Paradis et al., 2023) was used to create the phylogenetic tree. Both the PCA and the polygenetic tree visualize population structure and possible increased gene flow between up- and downstream locations. Length distribution of juvenile trout from both 2016 and 2022 upstream locations were calculated and visualized in a boxplot using the ggplot2 geom_boxplot() function and Kvinemonen being the only location tested in both years were tested for significance with a two sampled t-test. All plots were visualized using ggplot2 (Wickham., 2023).

3. Results

A total of 84 individuals sampled at nine different locations including one outgroup (Nesheim) were included in the genetic study. Through ddRAD sequencing genotypes for 8999 SNP loci were retrieved. Genetic diversity was measured by observed heterozygosity (H_O) and compared with expected heterozygosity (H_E) (**Table 1**). The H_O varied from 0.20 to 0.34, where Kvinemonen2022 and Kvinemonen2016 had the lowest estimates. Most locations had negative average F_{IS} values (excess of heterozygotes), also the total average F_{IS} (-0.0345) was negative. Except from Kvinemonen2022 and Kvåsfidjan2022, all sites showed an overall excess of heterozygotes (negative F_{IS}).

Table 1. *Sample locations and number of individual and total trout sampled and analysed including number of individuals in the final dataset. Genetic diversity is estimated by average* H_O *and* H_E *per sample location. Average* F_{IS} *measures the proportion of heterozygotes (excess/deficiency of heterozygotes). Average length in each location is also given (Lislåna22, Kvinemonen22, Kvåsfidjan22, Kvinemonen16, Birkeland16 are juvenile fish and Foss16, Res_Nothølen17(resident), Sea_Nothølen17 (sea trout) are adult fish).*

	Sampled	***Analysed***	***Total***				***Average***
Location	***n***	***n***	***n***	***Ho***	***HE***	***FIS***	***Length***
Lislåna22	41	15	10	0.23	0.22	-0.037	16.5 cm
Kvinemonen22	11	11	10	0.20	0.21	0.012	10.12 cm
Kvåsfidjan22	16	16	5	0.24	0.26	0.002	10.16 cm
Kvinemonen16	15	15	15	0.21	0.20	-0.029	17.0 cm
Birkeland16	5	5	5	0.34	0.29	-0.139	19.68 cm
Foss16	9	9	7	0.29	0.29	-0.011	43.0 cm
Res_Nothølen17	5	5	5	0.28	0.27	-0.051	26.54 cm
Sea_Nothølen17	8	8	8	0.30	0.29	-0.029	41.62 cm
Nesheim	22	22	19	0.27	0.26	-0.029	NA
Total	*132*	*106*	*84*	-	-	-	-
Average	*14.66*	*11.77*	*9.33*	*0.262*	*0.254*	*-0.0345*	*23.07 cm*

The inbreeding coefficient F is visualized as a dotplot (**Fig. 4**). Almost all individuals from Kvinemoen2022 and Kvåsfidjan2022 had an inbreeding coefficient F above 0. One individual from Foss2016 showed a considerable degree of inbreeding in comparison to the others that were close to 0 or below 0. The dotplot indicates little variance from 2016-2022 when comparing Kvinemonen2016 and Kvinemonen2022.

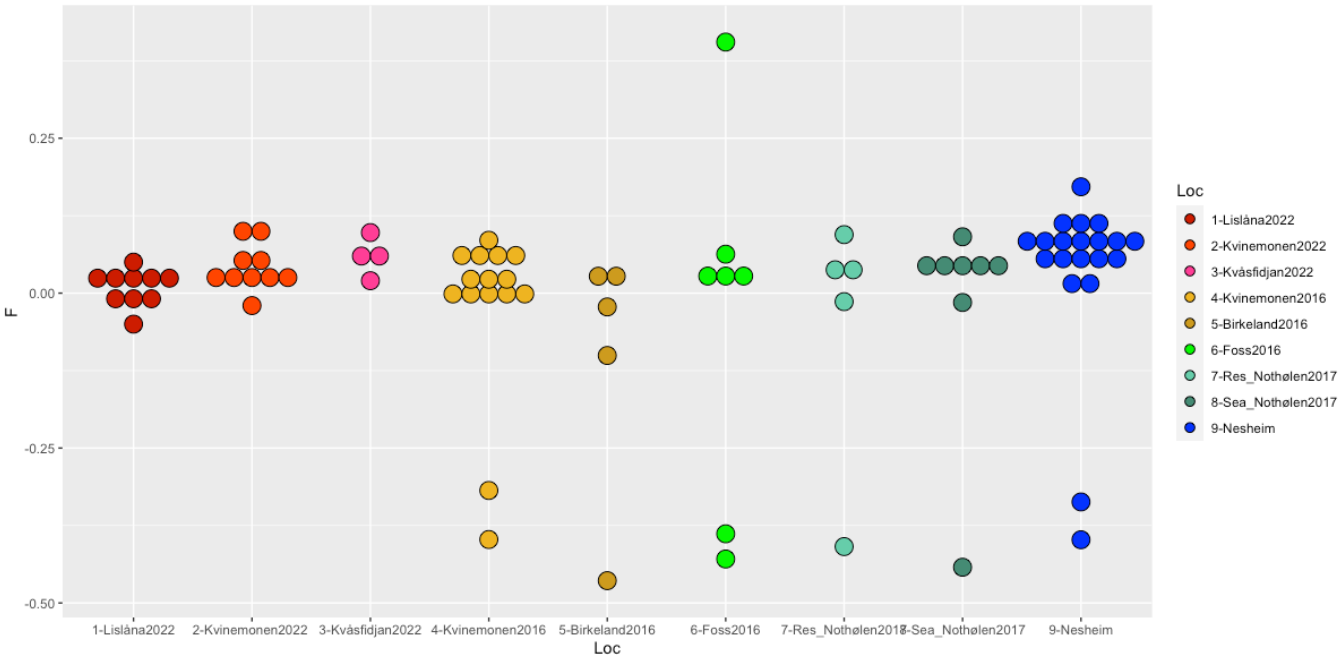


Figure 4. *Dotplot of the inbreeding coefficient F per individual in the different locations signified by a unique colour for both location and year.*

Pairwise F_{ST} values among locations ranged from 0 to 0.103 (**Table 2**) and the pairwise average value among locations were 0.0343 indicating that 3.43% of genetic variance is attributed by genetic difference among sample locations. In addition, 29 of the 36 locations F_{ST} values displayed significance when tested with a 95% bootstrapped confidence interval. Genetic difference has decreased from 2016 to 2022 between Kvinemonen and the downstream locations Foss2016, Res_Nothølen2017 and Sea_Nothølen2017.

Table 2. *Pairwise F_{ST} values calculated of the nine up- and downstream sample locations including one outgroup (Nesheim).*

	Lislåna22	Kvinemonen22	Kvåsfidjan22	Kvinemonen16	Birkeland16	Foss16	Res_Nothølen17	Sea_Nothølen17
Kvinemonen22	0.003*	-	-	-	-	-	-	-
Kvåsfidjan22	0.029*	0.036*	-	-	-	-	-	-
Kvinemonen16	0.005*	0.003*	0.054*	-	-	-	-	-
Birkeland16	0.025*	0.029*	0.006*	0.044*	-	-	-	-
Foss16	0.043*	0.049*	0	0.068*	0	-	-	-
Res_Nothølen17	0.019*	0.025*	0.001	0.036*	0	0	-	-
Sea_Nothølen17	0.047*	0.052*	0	0.070*	0.006*	0	0.01*	-
Nesheim	0.090*	0.092*	0.067*	0.103*	0.053*	0.053*	0.055*	0.042*

**Represent significance based on 95% bootstrapped confidence interval*

The PCA plot shows three prominent clusters, and the geographically close populations from upstream (2016 and 2022) and downstream (2016-2017) Kvåsfossen cluster closer together than the outgroup (Nesheim) (**Fig. 5**). Variation between the populations can be observed by how far apart they cluster in relation to the two principal components PCA1 and PCA2. PCA1 explains 6.39% of the genetic variation in our dataset and exhibits patterns where the upstream populations are clustered in one group to the left side of the plot and the downstream populations and Nesheim are clustered towards the right side of the plot. PCA2 explain 3.55% of the genetic variation and display divergence between the upstream, downstream and Nesheim population. Two individuals from Kvinemonen2022 and Kvåsfidjan2022 are clustered together with the downstream population indicating gene flow between the two populations. Additionally, two individuals from upstream locations 2016 cluster together with the downstream locations.

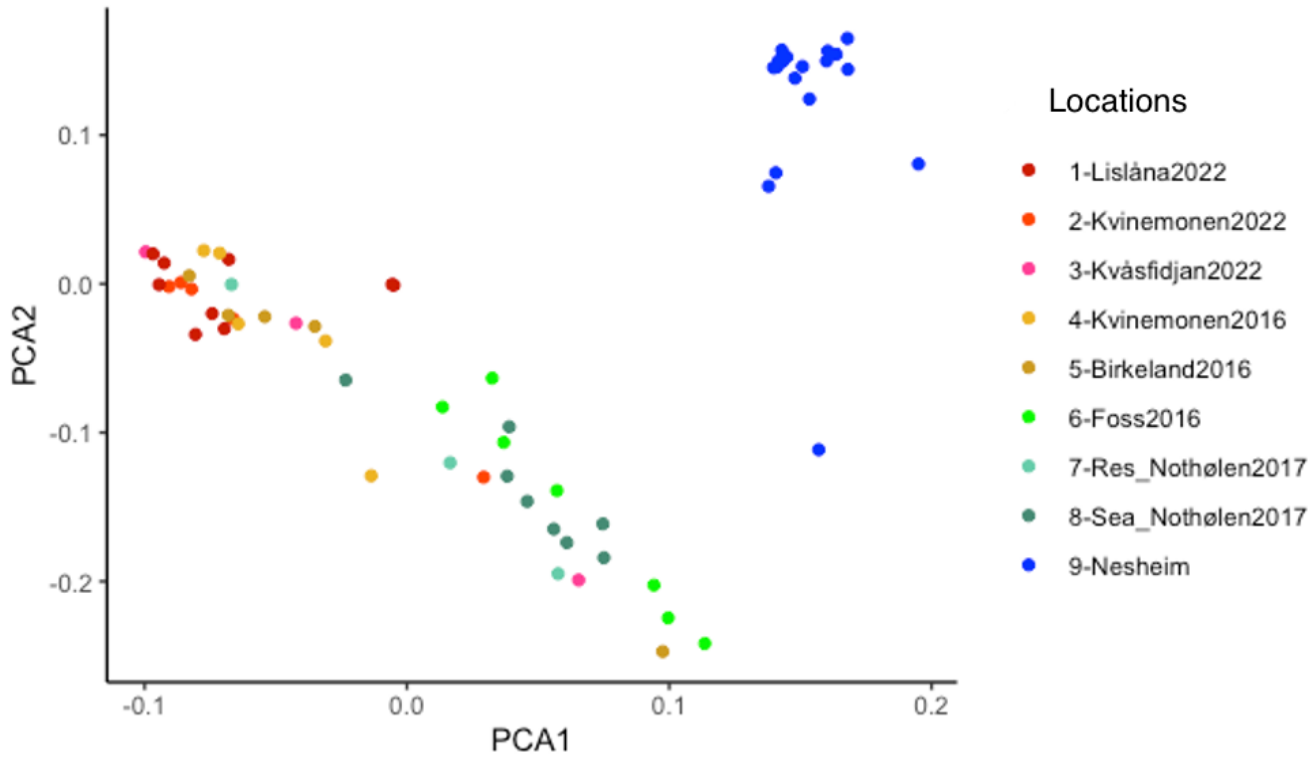


Figure 5. *Principal component analysis (PCA) on three population divided into nine locations. Each dot represents an individual and each location is signified by a unique colour that represents year and sample location.*

The phylogenetic tree also shows a prominent pattern of divergence (**Fig. 6**), clustering upstream- downstream and the Nesheim location into three clusters similar to the PCA plot. Two individuals from Kvåsfidjan2022 and one from Kvinemonen2022 clustered together with the downstream population indicating gene flow. In addition, one individual from Birkeland2016 cluster together with the downstream locations.

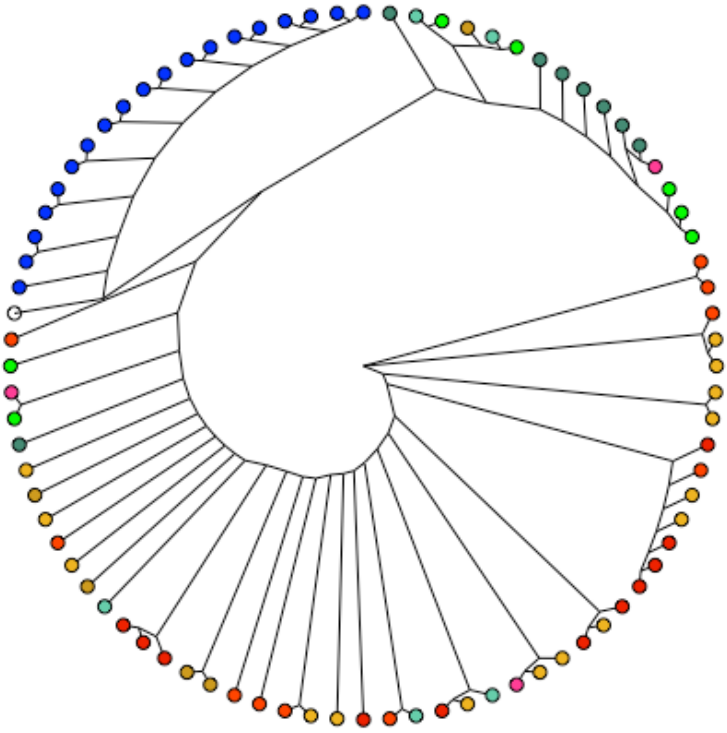


Figure 6. *Neighbouring phylogenetic tree showing three population divided into nine locations presented as a radial cladogram. Each dot represents an individual and each location is signified by a unique colour that represents year and sample location.*

Length comparison between upstream locations 2016 and 2022 were conducted to assess any phenotypic effect of establishment of the fish ladder (**Fig. 7**). T-tests revealed significant differences in length distribution among juvenile trout when comparing Kvinemonen2016 and Kvinemonen2022 ($t(23) = 5.5937$, $P = 1.08e\text{-}05$). Kvinemonen was the only comparable location sampled in both 2016 and 2022.

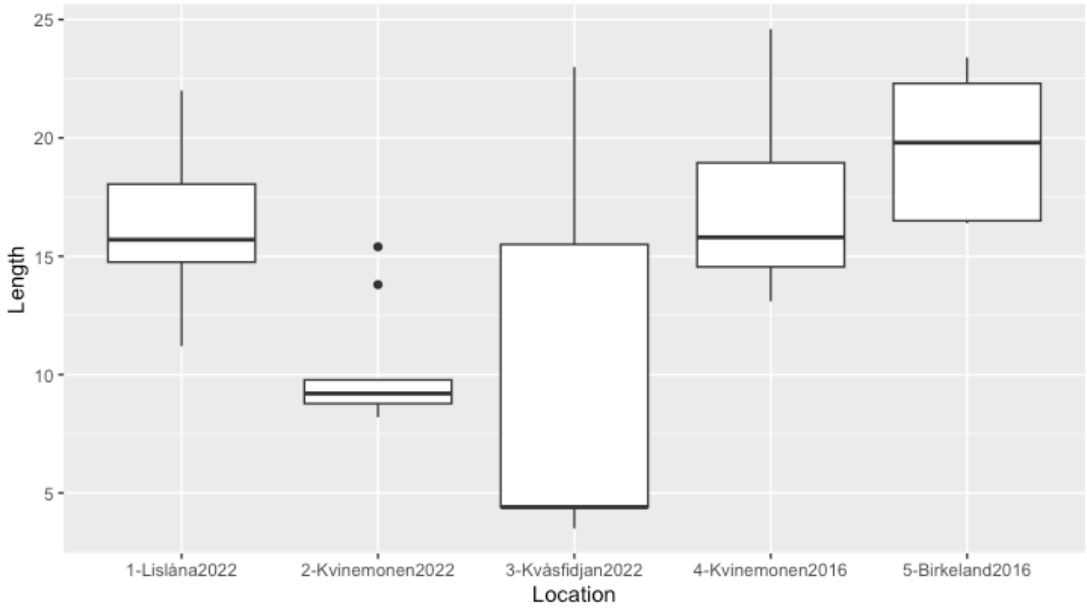


Figure 7. *Boxplot displaying length distribution of juvenile trout in upstream locations from 2016 and 2022.*

4. Discussion

The aim of this study was to assess the effect removal of a migration barrier had on a previously isolated population of trout in the river Lygna. Genomic data from 2016 and 2022 have been compared to evaluate this effect through investigation of genetic differentiation, genetic diversity and inbreeding, as well as trough clustering analyses.

The two hypotheses were: i) That genetic difference between up- and downstream locations had decreased after establishment of the fish ladder and ii) That genetic diversity and inbreeding had decreased in locations upstream the fish ladder.

After establishment of the fish ladder in 2014 trout were able to reach the upstream habitats. Compared with trout sampled in 2016, trout sampled upstream Kvåsfossen in 2022 showed decreased genetic differentiation to trout from sampling locations downstream Kvåsfossen. These results suggests that the fish ladder has eliminated a migration barrier that previously divided the trout population in Lygna. Pair-wise comparison of genetic differentiation (**Table 2**) of trout samples from the nine locations (**Table 1**) in combination with a PCA (**Fig. 5**) and phylogenetic tree (**Fig. 6**) indicates gene flow between the genetically distinct populations from up- and downstream Kvåsfossen in 2022. Observed heterozygosity were higher than expected

in most locations except from Kvinemonen2022 and Kvåsfidjan2022 (**Table 1**) and difference in the level of heterozygosity was observed among locations. Average F_{IS} and the inbreeding coefficient F confirmed inbreeding in two of the upstream locations from 2022 (**Fig. 4**). There were also found an interesting difference when comparing length distribution among locations between 2016 and 2022 (**Fig. 7**) The discussion below will elaborate further on the genetic patterns found within and between the sampled locations and removal of migration barriers as an effective conservation method.

4.1 Genetic effects

The bootstrapped 95% confidence interval revealed that most F_{ST} values for sample locations were significantly different from zero (**Table 2**). The results also showed that there had been a decrease in genetic difference between upstream and downstream trout from 2016 to 2022. The principal component analysis and the phylogenetic tree further confirms this by clustering individual trout caught upstream in 2022 that are genetically similar to downstream trout (**Fig. 5-6**). These results support hypothesis 1 that genetic difference between upstream and downstream locations has decreased after establishment of the fish-ladder. Regardless of the decreased differentiation between the upstream and downstream population, the population is still not totally admixed. This can be the result of behavioural or reproductive barriers that evolved during isolation (Fraik et al., 2022). The PCA plot and the phylogenetic tree also show genetic similarities in trout caught up- and downstream in 2016. Trout caught downstream that were genetically alike the upstream trout in 2016 may have swum down the waterfall. Genetically similar upstream trout to downstream trout can have been caught downstream and put back upstream or be offspring of early migrants from downstream (**Fig. 5-6.)** These findings are similar to those of Bernaś et al. (2021), who found gene flow between an isolated upstream population of trout and downstream trout in a hybridisation zone downstream of a migration barrier. A monitoring study of brown trout and migration patterns on smolt showed that resident parents gave a considerable contribution to anadrome migration. Offspring of resident trout also showed a buffering effect against elevated marine mortality (Duval et al., 2021). This communicates the importance of free migration among resident and anadrome fish.

Genetic differences between almost all the up- and downstream locations were very high (average F_{ST} values 0.19 - 0.103) in 2016. F_{ST} values from 2022 were high but considerably lower than in 2016 (**Table 2**). Downstream locations from 2016 and two upstream

locations Kvåsfidjan2022 and Birkeland2016 showed very low average F_{ST} (0 or close to 0) (**Table 2**). This can be due the fact that Kvåsfidjan and Birkeland is the geographically closest upstream locations to Kvåsfossen (**Fig.1**) and that gene flow has occurred in a higher degree than in other upstream locations from 2022. The low F_{ST} values can also have been caused by the low sample sizes in the two locations. A study on gene flow in a fragmented population of trout (Bernaś et al., 2021) also showed high or very high F_{ST} values between the up- and downstream populations, and that that trout from downstream locations in the basin were overall homogenous most likely because of the possibility for free migration. Selection, genetic drift and gene flow all affect genetic variation both within and among populations. It is the absence of gene flow that increases the genetic differences between fragmented populations, whereas genetic drift and selection (Bernaś et al., 2021).

Relative few studies have investigated the genetic impacts of removal of natural migration barriers. What we do have is a lot of studies on (Bohlin et al., 1989) with free migration that show great genetic distance and differences in genetic diversity (Bernaś et al., 2021; Deiner et al., 2007; Whiteley et al., 2010). Populations isolated by natural barriers show higher genetic differentiation than populations isolated by man-made barriers (Deiner et al., 2007). Lower genetic difference in populations divided by man-made barriers can be due to historical events of gene flow prior to isolation (Fraik et al., 2022). Knowing that gene flow decreases genetic differences and having results that shows decreased difference between locations and supporting results from (Fraik et al., 2022) there is reason to indicate our results are viable.

Genetic diversity in sample locations was measured by heterozygosity (**Table 1**). The average observed heterozygosity within locations is higher than expected suggesting an isolate-breaking effect that can be ascribed to mixing of two previously isolated populations that have resulted in increased genetic diversity (Hartl & Clark, 1997). Increased genetic diversity can have several positive effects on a population such as the ability to endure environmental changes and create new genetic variants that are beneficial in the long run (Bernatchez, 2016). The general assumption is that higher levels of genetic diversity increase species fitness and their ability to endure environmental changes (Teixeira & Huber, 2020). Several studies argue that conservation biology gives conserving genome-vide genetic variation too much focus and that the focus should be on functional genetic variation that is believed to affect fitness (Teixeira & Huber, 2020). For example, small populations show persistence over long time periods in despite of low genetic variation (Robinson et al., 2018; Teixeira & Huber, 2020; Xue, 2015), in addition to the collapse of the Isle Royal wolf population after increased genetic

variation through immigration (Robinson et al., 2019). These incidences are used to argue that genome-wide genetic variation is not essential to conservation. Despite these arguments (Kardos et al., 2021) states that conservation of genome-wide genetic variation is the best approach to prevent inbreeding depression and extinction and that functional genetic diversity is only applicable occasionally. Kvinemonen2022 and Kvåsfidjan2022 were the only locations showing lower observed heterozygosity which can be ascribed to inbreeding in the upstream locations.

Negative F_{IS} values dominated in the majority of locations showing an overall excess of heterozygotes (**Table 1**). Negative F_{IS} shows avoidance of inbreeding. Kvinemonen2022 and Kvåsfidjan2022 being the only two location showing positive F_{IS} indicates inbreeding in upstream locations (deficiency in heterozygotes). Inbreeding in Kvinemonen2022 and Kvåsfidjan2022 was confirmed by the inbreeding coefficient F that showed a degree of inbreeding in almost all of the individuals sampled at these two locations (**Fig. 4**). When comparing F between Kvinemonen2016 and 2022 there is a slight variation. These findings reject hypothesis 2 of increased genetic diversity and decreased inbreeding in upstream locations. These results correlate with a study in the Parsęta River from 2021 (Bernaś et al., 2021), where calculations of genetic diversity indicated very clearly that individuals from upstream locations of barriers had the lowest heterozygosity. What's interesting is that our data show little increase in genetic diversity in upstream locations after the introduction of the fish-ladder. This can be due to low migration of trout upstream, a to short time frame to notice significant increase or a to small sample size but is most likely caused by genetic drift or natural selection against heterozygosity (King et al., 2020). Genetic diversity results from a dam-fragmented population in Elwha River, Washington, showed concordant findings to our study, which showed genetic differentiation but no significant changes in genetic diversity after dam removal (Fraik et al., 2022).

4.2 Phenotypic effects

The body length of all trout was measured during sampling. Length is a phenotypical trait and can tell a lot about the fitness of a population (Dimitriew, 2011; Klemetsen et al., 2003). It is noteworthy to highlight that there was only one comparable location that was sampled both in 2016 and 2022, namely Kvinemonen. This comparison is not enough to make more than assumptions on what may be the reason for the length difference found between 2016 and 2022. Kvinemonen showed significant difference in length distribution from 2016 to

2022 (**Fig. 7**). Average juvenile length in 2016 were 17.0 cm whereas in 2022 average length were 10.12 cm (**Table 1**). Why has the average length decrease from 2016-2022? Firstly, new competition from salmon but also resident and anadromous trout from downstream has moved in. The increased abundance can have negative effects on fish growth and can happen in both streams (Jenkins Jr et al., 1999) and lakes (Jensen, 1977). Potentially making competition for food, space and spawning grounds considerably higher. Barriers are removed to increase habitat availability to restore old or create new migratory routes with the intention of increasing genetic diversity in several fish species, but these isolating barriers can also protect native species from aggressive downstream migrants (Stanley et al., 2007). Since 2000 the salmon population in Lygna has increased while the sea trout population has decreased, indicating that competition from salmon has intensified (Miljødrektoratet, 2016).

Secondly, more of the large females and males can have migrated to the sea and become seatrout and a larger portion of "sneakers" (small male trout) are left to fertilise the eggs resulting in smaller offspring (Bagenal, 1973; Fleming & Gross, 1991). Thirdly, both increasing and decreasing temperatures affect growth. Optimal temperature for growth lies between 13-18 ℃, lower critical temperature is 3-6 ℃ and the upper critical temperature is 25-26% (Elliott & Hurley, 2000; Ojanguren et al., 2001). An example is the apparent correlation between declining trout populations (Clews et al., 2010) and warming of British rivers due to increased summer temperatures (Jonkers & Sharkely, 2016).

4.3 Limitations

A short coming in this study was few comparable locations that were sampled both in 2016 and 2022 from, due to confusion in which locations that had been genotyped in 2016. The possibility for resampling comparable stations was limited due to the short time frame where electrofishing is effective. Including the time needed for genotyping (3-4 months). Additionally, some locations had low number of DNA samples in particularly Kvåsfidjan2022 where a lot of trout were wrongly species determined and had to be removed. These limitations may be ascribed to as human errors.

5. Conclusion

This study characterizes the effects removal of a natural barrier has on the genetic variation on a previously isolated Brown trout population in the river Lygna. Decreased genetic structure and differentiation between the previously genetically different locations confirms gene flow between the up-and downstream populations. Lack of total admixture of up-and downstream populations indicates that behavioural or reproductive barriers has evolved during isolation. No significant change in genetic diversity or inbreeding was detected above Kvåsfossen, which can be ascribed to genetic drift or natural selection. Migration between populations is proven to be a viable conservation strategy to increase population fitness in genetically depleted populations. Limitations in the form of few comparable locations and low sample size in some location were ascribed to as human errors. Future genomic studies on impassable natural migration barriers are necessary to confirm or dismiss the non-significant change in genetic diversity after barrier removal.

6. References

Aiello, B. R., Stewart, T. A., & Hale, M. E. (2016). Mechanosensation in an adipose fin. *The Royal Society, 283*(1825). doi:https://doi.org/10.1098/rspb.2015.2794

Arif, I. A., Bakir, M. A., Khan, H. A., Anis Ahamed, A., Al Farhan, A. H., Al Homaidan, A. A., . . . Shobrak, M. (2010). A Simple Method for DNA Extraction from Mature Date Palm Leaves: Impact of Sand Grinding and Composition of Lysis Buffer. *International Journal of Molecular Sciences, 11*(9), 3149-3157. doi:https://doi.org/10.3390/ijms11093149

Bagenal, T. B. (1973). Fish fecundity and its relations with stock and recruitment. *Rapp Pv Réun Cons Int Explor Mer 164*, 186-198.

Bayley, P. B., & Herendeen, R. A. (2000). The Efficiency of a Seine Net. *Transactions of the American Fisheries Society, 129*(4), 901-923. doi:https://doi.org/10.1577/1548-8659(2000)129<0901:TEOASN>2.3.CO;2

Bernaś, R., Wąs-Barcz, A., M. Árnyasi, M., Dębowski, P., Radtke, G., Poćwierz-Kotus, A., & Berrebi, P. (2021). Evidence of undirectional gene flow in a fragmented population of Salmon trutta L. *Scientific Reports 11*, 1. doi:https://doi.org/10.1038/s41598-021-02975-9

Bernatchez, L. (2016). On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. *Journal of Fish Biology, 89*(6). doi:https://doi.org/10.1111/jfb.13145

Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., . . . Guiguen, Y. (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications, 5*(3657). doi:https://doi.org/10.1038/ncomms4657

Black IV, W. C., Baer, C. F., Antolin, M. F., & DuTeau, N. M. (2001). Population genomics: genome-wide sampling of insect populations. *Annual review of entomology, 46*(1), 441-469. doi:https://doi.org/10.1146/annurev.ento.46.1.441

BGI Genomics https://www.bgi.com

Bohlin, T., Hamrin, S., Heggberget, T. G., Rasmussen, G., & Saltveit, S. J. (1989). Electrofishing — Theory and practice with special emphasis on salmonids. *Hydrobiologia, 173*, 9-43. doi:https://doi.org/10.1007/BF00008596

Catchen, J., Hohenlohe, P. A., Susan , Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology 22*(11), 3124-3140. doi:https://doi.org/10.1111/mec.12354

Charles, K., Roussel, J. M., & Cunjak, R. A. (2004). Estimating the contribution of sympatric anadromous and freshwater resident brown trout to juvenile production. *Marine and freshwater research, 55*(2), 185-191. doi:https://doi.org/10.1071/MF03173

Clair, T. A., & Hindar, A. (2005). Liming for the mitigation of acid rain effects in freshwaters: A review of recent results. *Environmental Reviews, 13*(3), 91-128. doi:https://doi.org/10.1139/a05-009

Clews, E., Durance, I., Vaguhan, I. P., & Ormerod, S. J. (2010). Juvenile salmonid populations in a temperate river system track synoptic trends in climate. *Global Change Biology, 16*(12), 3271-3283. doi:https://doi.org/10.1111/j.1365-2486.2010.02211.x

Creer, S., Deiner, K., Frey, S., Dorota Porazinska, D., Taberlet, P., Thomas, W. K., . . . Bik, H. M. (2016). The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution, 7*(9), 1008-1018. doi:https://doi.org/10.1111/2041-210X.12574

Crête-Lafrenière, A., Weir, L. K., & Bernatchez, L. (2012). Framing the Salmonidae Family Phylogenetic Portrait: A More Complete Picture from Increased Taxon Sampling. *PloS One, 7*(10). doi:https://doi.org/10.1371/journal.pone.0046662

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

Deiner, K., Garza, J. C., Coey, R., & Girman, D. (2007). Population structure and genetic diversity of trout (Oncorhynchus mykiss) above and below natural and man-made barriers in the Russian

River, California. *Conservation Genetics, 8*, 437-454. doi:https://doi.org/10.1007/s10592-006-9183-0
Dimitriew, C. M. (2011). The evolution of growth trajectories: what limits growth rate? *Biological Reviews, 86*(1), 97-116. doi:https://doi.org/10.1111/j.1469-185X.2010.00136.x
Duval, E., Skaala, Ø., Quintela, M., Dahle, G., Delaval, A., Wennevik, V., . . . Hansen, M. M. (2021). Long-term monitoring of a brown trout (Salmo trutta) population reveals kin-associated migration patterns and contributions by resident trout to the anadromous run. *BMC Ecology and Evolution 21*(143). doi:https://doi.org/10.1186/s12862-021-01876-9
Elliott, J. M. (1994). Quantitative ecology and the brown trout. *Oxford University Press*.
Elliott, J. M., & Hurley, M. A. (2000). Daily energy intake and growth of piscivorous brown trout, Salmo trutta. *Freshwater Biology, 44*(2), 237-445. doi:https://doi.org/10.1046/j.1365-2427.2000.00560.x
Ferguson, A., Reed, T. E., Cross, T. F., McGinnity, P., & Prodöhl, P. A. (2019). Anadromy, potamodromy and residency in brown trout Salmo trutta: the role of genes and the environment. *Journal of Fish Biology, 95*(3), 692-718. doi:https://doi.org/10.1111/jfb.14005
Fisher, R. A. (1999). The genetical theory of natural selection: a complete variorum edition. *Oxford University Press*.
Fleming, I. A., & Gross, M. R. (1991). Latitudinal elines: a trade-off between egg number and size in Pacificsalmon. . *Ecology and Evolution, 71*(1), 1-11. doi: https://doi.org/10.2307/1940241
Fraik, A. K., McMillan, J. R., Liermann, M., Bennett, T., McHenry, M. L., McKinney, G. J., . . . Nichols, K. M. (2022). The Impacts of Dam Construction and Removal on the Genetics of Recovering Steelhead (Oncorhynchus mykiss) Populations across the Elwha River Watershed. *Genes, 13*(9). doi:https://doi.org/10.3390/genes12010089
Frankham, R., Ballou, J. D., Ralls, K., Eldridge, M., Dudash, M. R., Fenster, C. B., & …Sunnucks, P. (2017). Genetic management of fragmented animal and plant populations. *Oxford University Press*.
Freeland, J. R. (2020). *Molecular Ecology* (3e ed.): John Wiley & Sons Ltd.
Goudet. J., Jombart, T., Kamvar, Z. N., Archer, E., & Hardy, O. (2022). hierfstat: Estimation and Tests of Hierarchical F-Statistics. *R package version 0.5-11*. Retrieved from <https://CRAN.R-project.org/package=hierfstat>.
Greeley, J. R. (2011). The Spawning Habits of Brook, Brown and Rainbow Trout, and the Problem of Egg Predators. *Transactions of the American Fisheries Society, 62*(1), 239-248. doi:https://doi.org/10.1577/1548-8659(1932)62[239:TSHOBB]2.0.CO;2
Hansen, M. M., & Mensberg K-L.D. (1998). Genetic differentiation and relationship between genetic and geographical distance in Danish sea trout (Salmo trutta L.) populations. *Heredity, 81*, 493-504. doi:https://doi.org/10.1046/j.1365-2540.1998.00408.x
Hartl, D. L., & Clark, A. G. (1997). *Principles of population genetics* (Vol. 116). Sunderland: Sinauer associates
Heggenes, J. (2016). *Fisken i fjellet i frotid og nåtid - Hva kan DNA-analyser fortelle?* Kristiansand: Portal forlag.
Hindar, K., Jonsson, B., Ryman, N., & Ståhl, G. (1991). Genetic relationships among landlocked, resident, and anadromous Brown Trout, Salmo trutta L. *Heredity 66*, 81-91. doi:https://doi.org/10.1038/hdy.1991.11
Ho, S. Y. W., & Larson, G. (2006). Molecular clocks: when timesare a-changin'. *Trends in Genetics, 22*(2), 79-83. doi:https://doi.org/10.1016/j.tig.2005.11.006
Huber, W., Carey, V., & Gentleman, R., et al. . (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods, 12*(2), 115-121. doi:https://doi.org/10.1038/nmeth.3252
Hughes, A. R., Inouye, B. D., Johnson, M. T. J., Underwood, N., & Vellend, M. (2008). Ecological consequences of genetic diversity. *Ecology Letters 11*(6), 609-623. doi: https://doi.org/10.1111/j.1461-0248.2008.01179.x
ICES. (2020). Baltic Salmon and Trout Assessment Working Group (WGBAST). *ICES scientific reports, 2*(22), 261. doi:https://doi.org/10.17895/ices.pub.5974

Jenkins Jr, T. M., Diehl, S., Kratz, K. W., & Cooper, S. D. (1999). Effects of population density on individual growth of brown trout in streams. *Ecology 80*(3), 941-956. doi:https://doi.org/10.1890/0012-9658(1999)080[0941:EOPDOI]2.0.CO;2

Jensen, K. W. (1977). On the dynamic and exploitation of the population of brown trout, Salmo trutta L., in Lake Oevre Heimdalsvatn, southern Norway. . *Report Institute of Freshwater Research, Drottningholm., 56*, 18-69.

Jolliffe, I. T., & J., C. (2016). Principal component analysis: a review and recent developments. *Phil. Trans. R. Soc. A., 374*. doi:https://doi.org/10.1098/rsta.2015.0202

Jones, J., Börger, L., Tummers, J., Jones, P., Lucas, M., Kerr, J., . . . Garcia de Leaniza, C. (2019). A comprehensive assessment of stream fragmentation in Great Britain. *Science of The Total Environment, 673*, 756-762. doi:https://doi.org/10.1016/j.scitotenv.2019.04.125

Jonkers, A. R. T., & Sharkely, K. J. (2016). The Differential Warming Response of Britain's Rivers (1982-2011). *PloS One, 11*(11). doi:https://doi.org/10.1371/journal.pone.0166247

Jonsson, B., & Jonsson, N. (1993). Partial migration: niche shift versus sexual maturation in fishes. *Reviews in Fish Biology and Fisheries, 3*, 348-365. doi:https://doi.org/10.1007/BF00043384

Jonsson, B., & Jonsson, N. (2021). Differences in growth between offspring of anadromous and freshwater brown trout Salmo trutta. *Journal of Fish Biology, 99*(1), 18-24. doi:https://doi.org/10.1111/jfb.14693

Kardos, M., Armstrong, E. E., Fitzpatrick, S. W., Hauser, S., Hedrick, P. W., Miller, J. M., . . . W.C., F. (2021). The crucial role of genome-wide genetic variation in conservation. *PNAS, 118*(48). doi:https://doi.org/10.1073/pnas.2104642118

Kartverket. *Geovekst, kommuner og OSM - Geodata AS* Retrieved from https://portal.fiskeridir.no/portal/apps/webappviewer/index.html?id=9aeb8c0425c3478ea021771a22d43476

King, R. A., Stockley, B., & Stevens, J. R. (2020). Small coastal streams—Critical reservoirs of genetic diversity for trout (Salmo trutta L.) in the face of increasing anthropogenic stressors. *Ecology and Evolution, 10*(12), 5651-5669. doi:https://doi.org/10.1002/ece3.6306

Klemetsen, A., Amundsen, P. A., Dempson, J. B., Jonsson, B., Jonsson, N., O'Connell, M. F., & Mortensen, E. (2003). Atlantic salmon Salmo salar L., brown trout Salmo trutta L. and Arctic charr Salvelinus alpinus (L.): a review of aspects of their life histories. *Ecology of freshwater fish, 12*(1), 1-59. doi:https://doi.org/10.1034/j.1600-0633.2003.00010.x

Knaus, B. J., & Grünwald, N. J. (2017). VCFR: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources, 17*(1), 44-53. doi:https://dx.doi.org/10.1111/1755-0998.12549.

Knaus, B. J., Tabima, J. F., & Grünwald, N. J. (2023). Quality control. Retrieved from https://grunwaldlab.github.io/Population_Genetics_in_R/qc.html

Koll, R., Ribera, J. M., Brunner, R. M., Rebl, A., & Goldammer, T. (2020). Gene Profiling in the Adipose Fin of Salmonid Fishes Supports Its Function as a Flow Sensor. *Genes, 11*(1:21). doi:https://doi.org/10.3390/genes11010021

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*(14), 1754-1760. doi: https://doi.org/10.1093/bioinformatics/btp324

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

Macqueen, D. J., & Johnston, I. A. (2014). A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Royal Society, 281*(1778). doi:https://doi.org/10.1098/rspb.2013.2881

Malcolm, I. A., Millidine, K. J., Jackson, F. L., Glover, R. S., & Fryer, R. J. (2019). The National Electrofishing Programme for Scotland (NEPS) 2019. *11*(9), DOI: 10.7489/12321-12321.

McCusker, M. R., & Bentzen, P. (2010). Positive relationships between genetic diversity and abundance in fishes. *Molecular Ecology, 19*(22), 4852-4862. doi:https://doi.org/10.1111/j.1365-294X.2010.04822.x

Menz, F. C., & Seip, H. M. (2004). Acid rain in Europe and the United States: an update. *Environmental Science Policy, 7*(4), 253-265. doi:https://doi.org/10.1016/j.envsci.2004.05.005

Miljødrektoratet. (2016). *Kalking i laksevassdrag skadet av sur nedbør Tiltaksovervåking i 2016*. Retrieved from https://www.miljodirektoratet.no/globalassets/publikasjoner/m821/m821.pdf

Morin, P. A., Luikart, G., & Wayne, R. K. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology and Evolution, 19*(4), 208-216. doi:https://doi.org/10.1016/j.tree.2004.01.009

Near, T. J., Eytan, R. I., Dornburg, A., Kuhn, K. L., Moore, J. A., Davis, M. P., . . . Smith, W. L. (2016). Resolution of ray-finned fish phylogeny and timing of diversification. *PNAS, 109*(34), 13698-13703. doi:https://doi.org/10.1073/pnas.1206625109

Nei, M., & Chesser, R. K. (1983). Estimation of fixation indices and gene diversities. *Annals of human genetics 37*(3), 253-259. doi:https://doi.org/10.1111/j.1469-1809.1983.tb00993.x

Ojanguren, A. F., Reyes-Gavilan, F. G., & Brana, F. (2001). Thermal sensitivity of growth, food intake and activity of juvenile brown trout. *Journal of Thermal Biology 26*(3), 165-170. doi:https://doi.org/10.1016/S0306-4565(00)00038-3

Pakkasmaa, S., & J., P. (2001). Morphological differentiation among local trout (Salmo trutta) populations *Biological Journal of the Linnean Society, 72*(2), 231-239. doi:https://doi.org/10.1111/j.1095-8312.2001.tb01313.x

Paradis, E., Blomberg, S., Bolker, B., & Brown, J. (2023). Analyses of Phylogenetics and Evolution. Retrieved from http://ape-package.ird.fr/, https://github.com/emmanuelparadis/ape

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PloS One 7*(5). doi:https://doi.org/10.1371/journal.pone.0037135

Portt, C. B., Coker, G. A., Ming, D. L., & Randall, R. G. (2006). A review of fish sampling methods commonly used in Canadian freshwater habitats.

Prodöhl, P. A., Ferguson, A., Bradley, C. R., Ade, R., Roberts, C., Keay, E. J., . . . Hynes, R. (2019). Impacts of acidification on brown trout Salmo trutta populations and the contribution of stocking to population recovery and genetic diversity. *Journal of Fish Biology, 95*(3), 719-742. doi:https://doi.org/10.1111/jfb.14054

Robinson , J. A., Brown. C., Kim, B. Y., Lohmuller, K. E., & Wayne, R. K. (2018). Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in island foxes. *Current Biology, 28*, 3487-3494.

Robinson , J. A., Räikkönen, J., Vucetich, L. M., Vucetich, J. A., Peterson, R. O., Lohmuller, K. E., & Wayne, R. K. (2019). Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. *Science Advances, 5*(5). doi:https://doi.org/10.1126/sciadv.aau0757

Rosseland, B. O., Skogheim, O. K., & Sevaldrud, I. H. (1986). Acid deposition and effects in Nordic Europe. Damage to fish populations in Scandinavia continue to apace *Water Air and Soil Pollution, 30*, 65-74. doi:https://doi.org/10.1007/BF00305176

RStudio. (2020). RStudio Team: Integrated Development for R. *RStudio, PBC, Boston, MA* Retrieved from http://www.rstudio.com/.

Ryman, N., Utter, F. M., & Laikre, L. (1995). Protection of intraspecific biodiversity of exploited fishes. *Reviews in Fish Biology and Fisheries, 5*, 417-446. doi:https://doi.org/10.1007/BF01103814

Sandlund, O. T., & Hesthagen, T. (2011). Fish diversity in Norwegian lakes: conserving species poor systems. *Water biodiversity assessment and protection*, 7-20.

Schiewe, H. (2013). Salmon. *Encyclopedia of Biodiversity (Second Edition)*, 522-531. doi:https://doi.org/10.1016/B978-0-12-384719-5.00293-8

Schill, D. J., & Beland, K. F. (1995). Electrofishing injury studies-a call for population perspective. *Fisheries, 20*(6), 28-29.

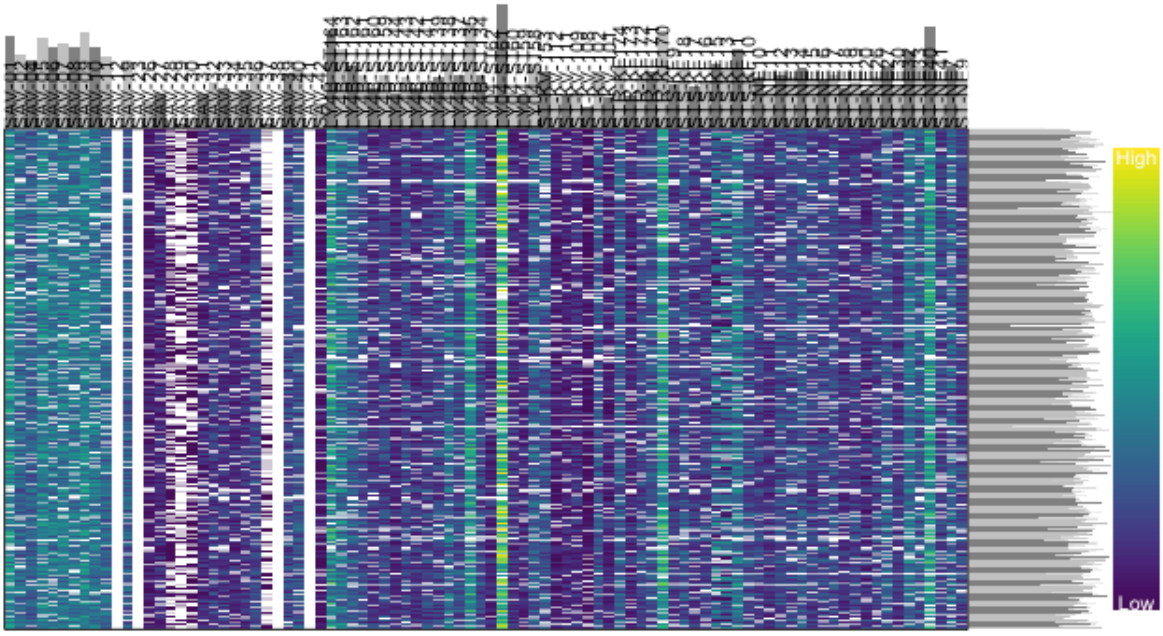
Schlaepfer, D. R., Braschler, B., Rusterholz, H.-P., & Baur, B. (2018). Genetic effects of anthropogenic habitat fragmentation on remnant animal and plant populations: a meta-analysis. *ECOSPHERE 9*(10). doi:https://doi.org/10.1002/ecs2.2488

Seeb, J. E., Carvalho, L., Hauser, L., Naish, K., Roberts, S., & Seeb, L. W. (2011). Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms.

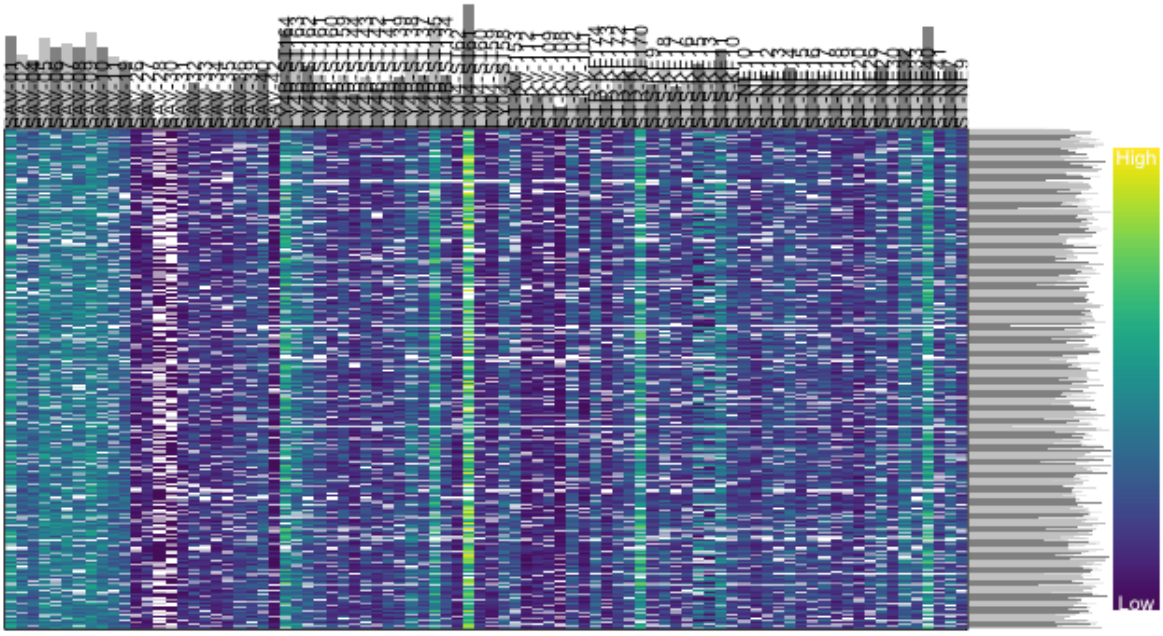
Molecular Ecology Resources, 11(1), 1-8. doi:https://doi.org/10.1111/j.1755-0998.2010.02979.x
Snyder, D. E. (1995). Impacts of electrofishing on fish. *Fisheries 20*(1), 26-27.
Snyder, D. E. (2003). Invited overview: conclusions from a review of electrofishing and its harmful effects on fish. *Reviews in Fish Biology and Fisheries, 13*, 445-453. doi:https://doi.org/10.1007/s11160-004-1095-9
Stanley, E. H., Catalano, M. J., Mercado-Silva, N., & Orr, C. H. (2007). Effects of dam removal on brook trout in a Wisconsin stream. *River Research and Applications 23*(7), 792-798. doi: https://doi.org/10.1002/rra.1021
Stapley, J., Reger, J., Feulner, P. G. D., Smadja, C., Galindo J., Ekblom, R., . . . Slate, J. (2010). Adaptation genomics: the next generation. *Trends in Ecology and Evolution, 25*(12), 705-712. doi:https://doi.org/10.1016/j.tree.2010.09.002
Taggart, J. B., Hynes, R. A., Prodöhl, P. A., & Ferguson, A. (1992). A simplified protocol for routine total DNA isolation from salmonid fishes. *Journal of Fish Biology, 40*(6), 963-965.
Teixeira, J. C., & Huber, C. D. (2020). The inflated significance of neutral genetic diversity in conservation genetics. *118*(10). doi:https://doi.org/10.1073/pnas.2015096118
Vandewoestijne, S., Schtickzelle, N., & Baguette, M. (2008). Positive correlation between genetic diversity and fitness in a large, well-connected metapopulation. *BMC Biology, 6*(46). doi:https://doi.org/10.1186/1741-7007-6-46
Vikøyr, B., Haraldstad, Ø., & Larsen, P. A. (1989). Kalkingsplan Lygna.
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. doi:https://doi.org/10.1093/jhered/89.5.438
Weir, B. S., & Cockerham, C. C. Estimating F-statistics for analysis of population structure *Evolution, 38*(6), 1358-1370. doi:https://doi.org/10.1111/j.1558-5646.1984.tb05657.x
Westgaard, J. I., Saha, A., Kent, M., Hansen, H. H., Knutsen, H., Hauser, L., . . . Torild., J. (2016). Genetic population structure in Greenland halibut (Reinhardtius hippoglossoides) and its relevance to fishery management. *Canadian Journal of Fisheries and Aquatic Sciences, 74*(4), 475-485. doi:https://doi.org/10.1139/cjfas-2015-0430
Whiteley, A. R., Hastings, K., Wenburg, J., Frissell, C. A., Martin, J. C., & F. W. Allendorf, F. W. (2010). Genetic variation and effective population size in isolated populations of coastal cutthroat trout. *Conservation Genetics, 11*(5), 1929-1943. doi:https://doi.org/10.1007/s10592-010-0083-y
Wickham., H. (2023). ggplot: Create Elegant Data Visualisations Using the Grammar of Graphics. Retrieved from https://github.com/tidyverse/ggplot2
Wilfinger, W. W., Mackey, K., & Chomczynski, P. (2018). Effect of pH and Ionic Strength on the Spectrophotometric Assessment of Nucleic Acid Purity. *BioTechniques 22*(3). doi:https://doi.org/10.2144/97223st01
Winans, G. A., Allen, M. B., Baker, J., Lesko, E., Shrier, F., Strobel, B., & J., M. (2018). Dam trout: Genetic variability in Oncorhynchus mykiss above and below barriers in three Columbia River systems prior to restoring migrational access. *13*(5). doi:https://doi.org/10.1371/journal.pone.0197571
Xue, Y. e. a. (2015). Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science, 348*, 242-245.
Zheng, X., Levine, D., Shen, J., Gogarten, S., Laurie, C., & Weir, B. (2012). A High-performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data. *Bioinformatics, 28*(24), 3326-3328. doi:https://doi.org/10.1093/bioinformatics/bts606

7. Appendix

Appendix 1. Heatmap showing low quality samples detected and removed in quality control. The white lines and the lines with little colour are missing or low-quality DNA.



(Before quality control)



(After quality control)

Appendix 2. R script

Script 1. - Quality control

```
### Quality control ###


library(SNPRelate)
library(ggplot2)

setwd("/Users/save-mariarornessveinson/Documents/Kvåsfossen ørret data ")
snpgdsVCF2GDS("ALL935i_jan23_Bt_rmDup_mac2BiHWE10_DP8mm8_42723s_90SaveI
nd_8999s.vcf", "Lygna_N2.gds", method=c("biallelic.only", "copy.num.of.ref"),
        snpfirstdim=FALSE, compress.annotation="ZIP_RA.max", compress.geno="",
        ref.allele=NULL, ignore.chr.prefix="bt", verbose=TRUE)

snpgdsClose(gen)

###
library(vcfR)
vcf=

vcf <-
read.vcfR("ALL935i_jan23_Bt_rmDup_mac2BiHWE10_DP8mm8_42723s_90SaveInd_899
9s.vcf")

head(vcf)
vcf

###
as.numeric(c(FALSE, TRUE))

sum(as.numeric(c(FALSE, TRUE)))

###
queryMETA(vcf, "DP")

dp <- extract.gt(vcf, element = "DP", as.numeric=TRUE)
sum(is.na(dp[,1]))

###
myMiss <- apply(dp, MARGIN = 2, function(x){ sum(is.na(x)) }
myMiss <- myMiss/nrow(vcf)

library(RColorBrewer)
palette(brewer.pal(n=12, name = 'Set3'))

par(mar = c(12,4,4,2))
barplot(myMiss, las = 2, col = 1:12)
```

```
title(ylab = "Missingness (%)")

###
par(mar = c(5,4,4,2))

###
myMiss <- apply(dp, MARGIN = 1, function(x){ sum(is.na(x)) })
myMiss <- myMiss/ncol(vcf@gt[,-1])

hist(myMiss, col = "#8DD3C7", xlab = "Missingness (%)", main = "")

###
library(vcfR)
vcf <-
read.vcfR('ALL935i_jan23_Bt_rmDup_mac2BiHWE10_DP8mm8_42723s_90SaveInd_8999
s.vcf')
head(vcf)

data()
dp <- extract.gt(vcf, element = "DP", as.numeric=TRUE)

###
par(mar=c(12,4,4,2))
boxplot(dp, col=2:8, las=3)
title(ylab = "Depth (DP)")

###
library(reshape2)
library(ggplot2)
library(cowplot)

# Melt our matrix into a long form data.frame.
dpf <- melt(dp, varnames=c('Index', 'Sample'), value.name = 'Depth', na.rm=TRUE)
dpf <- dpf[ dpf$Depth > 0,]

# Create a row designator.
#samps_per_row <- 20
samps_per_row <- 100
myRows <- ceiling(length(levels(dpf$Sample))/samps_per_row)
myList <- vector(mode = "list", length = myRows)

for(i in 1:myRows){
  myIndex <- c(i*samps_per_row - samps_per_row + 1):c(i*samps_per_row)
  myIndex <- myIndex[myIndex <= length(levels(dpf$Sample))]
  myLevels <- levels(dpf$Sample)[myIndex]
  myRegex <- paste(myLevels, collapse = "$|^")
  myRegex <- paste("^", myRegex, "$", sep = "")
  myList[[i]] <- dpf[grep(myRegex, dpf$Sample),]
  myList[[i]]$Sample <- factor(myList[[i]]$Sample)
}
```

```
 # Create the plot.
 myPlots <- vector(mode = "list", length = myRows)
 for(i in 1:myRows){
  myPlots[[i]] <- ggplot(myList[[i]], aes(x=Sample, y=Depth)) +
    geom_violin(fill="#8dd3c7", adjust=1.0, scale = "count", trim=TRUE)

  myPlots[[i]] <- myPlots[[i]] + theme_bw()
  myPlots[[i]] <- myPlots[[i]] + theme(axis.title.x = element_blank(),
                                       axis.text.x = element_text(angle = 60, hjust = 1))
  myPlots[[i]] <- myPlots[[i]] + scale_y_continuous(trans=scales::log2_trans(),
                                                    breaks=c(1, 10, 100, 800),
                                                    minor_breaks=c(1:10, 2:10*10, 2:8*100))
  myPlots[[i]] <- myPlots[[i]] + theme( panel.grid.major.y=element_line(color =
"#A9A9A9", size=0.6) )
  myPlots[[i]] <- myPlots[[i]] + theme( panel.grid.minor.y=element_line(color =
"#C0C0C0", size=0.2) )
 }

# Plot the plot.
plot_grid(plotlist = myPlots, nrow = myRows)

###
vcf

quants <- apply(dp, MARGIN=2, quantile, probs=c(0.1, 0.8), na.rm=TRUE)
quants[,1:6]

###
dp2 <- sweep(dp, MARGIN=2, FUN = "-", quants[1,])
dp[dp2 < 0] <- NA

dp2 <- sweep(dp, MARGIN=2, FUN = "-", quants[2,])
dp[dp2 > 0] <- NA

dp[dp < 4] <- NA

vcf@gt[,-1][ is.na(dp) == TRUE ] <- NA

vcf

###
library(vcfR)
library(pinfsc50)
vcf <- system.file("extdata", "pinf_sc50.vcf.gz", package = "pinfsc50")
vcf <-
vcfR::read.vcfR("ALL935i_jan23_Bt_rmDup_mac2BiHWE10_DP8mm8_42723s_90SaveIn
d_8999s.vcf")
dp <- extract.gt(vcf, element = "DP", as.numeric=TRUE)
```

```
vcf

quants <- apply(dp, MARGIN=2, quantile, probs=c(0.1, 0.9), na.rm=TRUE)
dp2 <- sweep(dp, MARGIN=2, FUN = "-", quants[1,])
dp[dp2 < 0] <- NA
dp2 <- sweep(dp, MARGIN=2, FUN = "-", quants[2,])
dp[dp2 > 0] <- NA
dp[dp < 4] <- NA
vcf@gt[,-1][ is.na(dp) == TRUE ] <- NA

vcf

###
heatmap.bp(dp[1:1000,], rlabels = FALSE)

###
myMiss <- apply(dp, MARGIN = 2, function(x){ sum( is.na(x) ) })
myMiss <- myMiss / nrow(dp)
barplot(myMiss, las = 3)

###
vcf@gt <- vcf@gt[, c(TRUE, myMiss < 0.6)]
vcf

###
dp <- extract.gt(vcf, element = "DP", as.numeric=TRUE)
heatmap.bp(dp[1:1000,], rlabels = FALSE)


write.vcf(vcf, file =
"ALL935i_jan23_Bt_rmDup_mac2BiHWE10_DP8mm8_42723s_90SaveInd_8999s.vcf")


snpgdsVCF2GDS("LygnaNesheim_DP8mm7mac2hwe10_106i_22405s.vcf copy.gz",
"Orret2.gds", method=c("biallelic.only", "copy.num.of.ref"),
        snpfirstdim=FALSE, compress.annotation="ZIP_RA.max", compress.geno="",
        ref.allele=NULL, ignore.chr.prefix="NC_", verbose=TRUE)


### 2. Create plot
library(SNPRelate)
library(ggplot2)

gen= snpgdsOpen("Orret2.gds")
pca=snpgdsPCA(gen)
snpgdsClose(gen)
qplot(pca$eigenvect[,1], pca$eigenvect[,2])

library(plyr)

popmap=read.table("popmap_LygnaNesheim_jan23 copy.txt", header=T)
```

```
dfPCA=data.frame(INDV=pca$sample.id, PCA1= pca$eigenvect[,1], PCA2=
pca$eigenvect[,2])
data=join(dfPCA,popmap,by="INDV", type="inner")

qplot( data$PCA1, data$PCA2, col=as.factor(data$POP))+ xlab("PCA1") + ylab("PCA2")+
theme(legend.title=element_blank())

Lygna_N_gen
```

Script 2. PCA plot

```
BiocManager::install("SNPRelate")

### plot1 ###

### R package: SNPrelate – Principle components ###

library(SNPRelate)
library(ggplot2)

setwd("Users/save-mariarornessveinson/Documents/Kvåsfossen ørret data")
snpgdsVCF2GDS("ALL935i_jan23_Bt_rmDup_mac2BiHWE10_DP8mm8_42723s_90SaveI
nd_8999s.vcf", "Lygna_N.gds", method=c("biallelic.only", "copy.num.of.ref"),
        snpfirstdim=FALSE, compress.annotation="ZIP_RA.max", compress.geno="",
        ref.allele=NULL, ignore.chr.prefix="Bt", verbose=TRUE)

library(SNPRelate)
library(ggplot2)

gen= snpgdsOpen("Lygna_N.gds")
pca=snpgdsPCA(gen)
snpgdsClose(gen)
qplot(pca$eigenvect[,1], pca$eigenvect[,2])

library(plyr)
popmap=read.table("popmap_LygnaNesheimR1_jan23 copy 4.txt", header=T)
dfPCA=data.frame(INDV=pca$sample.id, PCA1= pca$eigenvect[,1], PCA2=
pca$eigenvect[,2])
data=join(dfPCA,popmap,by="INDV", type="inner")

qplot( data$PCA1, data$PCA2, col=as.factor(data$POP))+ xlab("PCA1") + ylab("PCA2")+
theme(legend.title=element_blank())

pca$varprop

### change colour ###

pcol=c("red3","orangered","violetred1","goldenrod2","goldenrod3","green","aquamarine3","
aquamarine4","blue")
```

```
qplot( data$PCA1, data$PCA2, col=data$POP) + theme_classic() + xlab("PCA1") +
ylab("PCA2") + scale_colour_manual(values=pcol)
data2=data[data$PCA1>=(-0.1),]
dataOut1=data[data$PCA1<=(-0.1),]
qplot( data2$PCA1, data2$PCA2, col=data2$POP) + theme_classic() + xlab("PCA1") +
ylab("PCA2") + scale_colour_manual(values=pcol)
```

Script 3. – Phylogenetic trees

```
### plot 2 ###
### R package: SNPrelate + Ape: drawing trees ###

library(SNPRelate)
library(ggplot2)
library(ape)

setwd("Users/save-mariarornessveinson/Documents/Kvåsfossen ørret dat")
snpgdsVCF2GDS("ALL935i_jan23_Bt_rmDup_mac2BiHWE10_DP8mm8_42723s_90SaveI
nd_8999s.vcf", "Lygna_N.gds", method=c("biallelic.only", "copy.num.of.ref"),
        snpfirstdim=FALSE, compress.annotation="ZIP_RA.max", compress.geno="",
        ref.allele=NULL, ignore.chr.prefix="Bt", verbose=TRUE)

popmap=read.table("popmap_LygnaNesheimR1_jan23 copy 4.txt", header=T)

gen= snpgdsOpen("Lygna_N.gds")

g <- snpgdsGetGeno(gen)
d=as.matrix(dist(g, method = "euclidean"))
colnames(d)=ind
rownames(d)=ind
tr <- bionjs(d)
plot(tr, "u",no.margin=TRUE,show.tip.label=T)

temp=data.frame(INDV=read.gdsn(index.gdsn(gen, "sample.id")))
temp2=join(temp,popmap, type="left", by="INDV")
tiplabels(bg=temp2$COLR, pch=21)
snpgdsClose(gen)

snpgdsClose(gen)

g <- snpgdsGetGeno(gen)
d=as.matrix(dist(g, method = "euclidean"))
colnames(d)=ind
rownames(d)=ind
tr <- bionjs(d)
plot(tr, "p",no.margin=TRUE,show.tip.label=T)

temp=data.frame(INDV=read.gdsn(index.gdsn(gen, "sample.id")))
```

```
temp2=join(temp,popmap, type="left", by="INDV")
tiplabels(bg=temp2$COLR, pch=21)
snpgdsClose(gen)

snpgdsClose(gen)

g <- snpgdsGetGeno(gen)
d=as.matrix(dist(g, method = "euclidean"))
colnames(d)=ind
rownames(d)=ind
tr <- bionjs(d)
plot(tr, "r",no.margin=TRUE,show.tip.label=T)

temp=data.frame(INDV=read.gdsn(index.gdsn(gen, "sample.id")))
temp2=join(temp,popmap, type="left", by="INDV")
tiplabels(bg=temp2$COLR, pch=21)
snpgdsClose(gen)
```

Script 4. Summary statistics

```
### Import VCF file ###

library(vcfR)
library(adegenet)

trout =
read.vcfR("ALL935i_jan23_Bt_rmDup_mac2BiHWE10_DP8mm8_42723s_90SaveInd_899
9s.vcf", nrows = 8999, verbose = FALSE)
trout

trout = vcfR2genind(trout)
popinf=read.csv("popinfo_new.csv", header=F, sep=";")
#trout$pop = as.factor(substr(ind?Names(trout), 1, 3))
trout$pop = as.factor(popinf$V2)

### Summary statistics ###

library(adegenet)
library(poppr)
library(dplyr)
library(hierfstat)
library(reshape2)
library(ggplot2)
library(RColorBrewer)
library(scales)

str(trout)

trout
```

```
table(trout$loc.fac)

summary(trout$pop)

allelic.richness(genind2hierfstat(trout))$Ar %>%
  apply(MARGIN = 2, FUN = mean) %>%
  round(digits = 3)

basic_trout = basic.stats(trout, diploid = TRUE)

Ho_trout = apply(basic_trout$Ho, MARGIN = 2, FUN = mean, na.rm = TRUE) %>%
  round(digits = 2)
Ho_trout

He_trout = apply(basic_trout$Hs, MARGIN = 2, FUN = mean, na.rm = TRUE) %>%
  round(digits = 2)
He_trout

### Visualise heterozygosity per site ###

# Create a data.frame of site names, Ho and He and then convert to long format
Het_trout_df = data.frame(Site = names(Ho_trout), Ho = Ho_trout, He = He_trout) %>%
  melt(id.vars = "Site")

# Custom theme for ggplot2
custom_theme = theme(
  axis.text.x = element_text(size = 10, angle = 90, vjust = 0.5, face = "bold"),
  axis.text.y = element_text(size = 10),
  axis.title.y = element_text(size = 12),
  axis.title.x = element_blank(),
  axis.line.y = element_line(linewidth = 0.5),
  legend.title = element_blank(),
  legend.text = element_text(size = 12),
  panel.grid = element_blank(),
  panel.background = element_blank(),
  plot.title = element_text(hjust = 0.5, size = 15, face="bold")
)

  hetlab.o = expression(italic("H")[o])
  hetlab.e = expression(italic("H")[e])

  # Trout heterozygosity barplot

  ggplot(data = Het_trout_df, aes(x = Site, y = value, fill = variable))+
    geom_bar(stat = "identity", position = position_dodge(width = 0.6), colour = "black")+
    scale_y_continuous(expand = c(0,0), limits = c(0,0.50))+
    scale_fill_manual(values = c("royalblue", "#bdbdbd"), labels = c(hetlab.o, hetlab.e))+
    ylab("Heterozygosity")+
    ggtitle("Brown trout")+
```

```
    custom_theme

### Inbreeding coefficient (FIS) ###

# Calculate mean FIS per site.
  apply(basic_trout$Fis, MARGIN = 2, FUN = mean, na.rm = TRUE) %>%
    round(digits = 3)

###FST, PCA & DAPC ### Compute pairwise FST ###

 # Subset data sets to reduce computation time
trout_gen_sub = popsub(trout, sublist =
c("Nesheim","Sea_Nothølen2017","Res_Nothølen2017","Foss2016","Birkeland2016","Kvin
emonen2016","Kvåsfidjan2022","Kvinemonen2022","Lislåna2022"))

# Compute pairwise Fsts
trout_fst = genet.dist(trout_gen_sub, method = "WC84") %>% round(digits = 3)
trout_fst

### Visualise pairwise FST for lobster ###

# Desired order of labels
lab_order =
c("Nesheim","Sea_Nothølen2017","Res_Nothølen2017","Foss2016","Birkeland2016","Kvin
emonen2016","Kvåsfidjan2022","Kvinemonen2022","Lislåna2022")

# Change order of rows and cols
fst.mat = as.matrix(trout_fst)
fst.mat1 = fst.mat[lab_order, ]
fst.mat2 = fst.mat1[, lab_order]

# Create a data.frame
ind = which(upper.tri(fst.mat2), arr.ind = TRUE)
fst.df = data.frame(Site1 = dimnames(fst.mat2)[[2]][ind[,2]],
                    Site2 = dimnames(fst.mat2)[[1]][ind[,1]],
                    Fst = fst.mat2[ ind ])

# Keep the order of the levels in the data.frame for plotting
fst.df$Site1 = factor(fst.df$Site1, levels = unique(fst.df$Site1))
fst.df$Site2 = factor(fst.df$Site2, levels = unique(fst.df$Site2))

# Convert minus values to zero
fst.df$Fst[fst.df$Fst < 0] = 0

# Print data.frame summary
fst.df %>% str

# Fst italic label

fst.label = expression(italic("F")[ST])
```

```
# Extract middle Fst value for gradient argument
mid = max(fst.df$Fst) / 2

# Plot heatmap
ggplot(data = fst.df, aes(x = Site1, y = Site2, fill = Fst))+
  geom_tile(colour = "black")+
  geom_text(aes(label = Fst), color="black", size = 3)+
  scale_fill_gradient2(low = "blue", mid = "pink", high = "red", midpoint = mid, name =
fst.label, limits = c(0, max(fst.df$Fst)), breaks = c(0, 0.05, 0.10, 0.15))+
  scale_x_discrete(expand = c(0,0))+
  scale_y_discrete(expand = c(0,0), position = "right")+
  theme(axis.text = element_text(colour = "black", size = 10, face = "bold"),
        axis.title = element_blank(),
        panel.grid = element_blank(),
        panel.background = element_blank(),
        legend.position = "right",
        legend.title = element_text(size = 14, face = "bold"),
        legend.text = element_text(size = 10)
  )

### bootstrapped Fst
boot.ppfst(dat=trout,nboot=100,quant=c(0.025,0.975),diploid=TRUE)
```

Script.5 – Inbreeding coefficient F

```
### Inbreeding estimation Likelihood-based estimation of inbreeding F ###

library(SNPRelate)
library(ggplot2)

gen= snpgdsOpen("Lygna_N.gds")

popmap=read.table("popmap_LygnaNesheimR1_jan23 copy 4.txt", header=T)
snpgdsIndInb(gen,sample.id=NULL,snp.id=NULL,
autosome.only=TRUE,remove.monosnp=TRUE,maf=NaN,missing.rate=NaN,
method=c("mom.weir","mom.visscher","mle","gcta1","gcta2","gcta3"),
allele.freq=NULL,out.num.iter=TRUE,reltol=.Machine$double.eps^0.75, verbose=TRUE)

# Read csv
InbreedingF <- read.csv("InbreedingF.csv", header = TRUE, sep=";")

# Basic dot plot
p<-ggplot(InbreedingF, aes(x=Loc, y=F)) +
  geom_dotplot(binaxis='y', stackdir='center')
p


# Change dotsize and stack ratio
ggplot(InbreedingF, aes(x=Loc, y=F)) +
```

```
  geom_dotplot(binaxis='y', stackdir='center',
               stackratio=1.5, dotsize=1.2)
p

# Change dot plot colors by groups
p<-ggplot(InbreedingF, aes(x=Loc, y=F, fill=Loc)) +
  geom_dotplot(binaxis='y', stackdir='center')
p

p+scale_fill_manual(values=c("red3","orangered","violetred1","goldenrod2","goldenrod3","g
reen","aquamarine3","aquamarine4","blue"))
```

Script 6. – Length distribution

```
### Length distribution ###

# loading data set and storing it in ds variable

library(ggplot2)

ds <- read.csv("PopLength.csv", header = TRUE, sep=";")

# create a boxplot by using geom_boxplot() function
# of ggplot2 package

crop=ggplot(data=ds, mapping=aes(x=Location, y=Length))+geom_boxplot()
crop

t.test(x=ds$Length, conf.level=0.95)

t.test(x=ds$Location, y=ds$Length, paired=TRUE)

# Filtering
Length_Li22 <- ds %>%
  filter(Location=="1-Lislåna2022") %>%
  pull(Length)

Length_Kvi22 <- ds %>%
  filter(Location=="2-Kvinemonen2022") %>%
  pull(Length)

Length_Kvås22 <- ds %>%
  filter(Location=="3-Kvåsfidjan2022") %>%
  pull(Length)

Length_Kvi16 <- ds %>%
  filter(Location=="4-Kvinemonen2016") %>%
  pull(Length)
```

```
Length_Birk16 <- ds %>%
  filter(Location=="5-Birkeland") %>%
  pull(Length)

# Two Sample t-test

res0 <- t.test(Length_Kvi16, Length_Kvi22, var.equal = TRUE)
res0 #p-value = 0.05613
```