Bachelor's project

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biological Sciences Ålesund

Jan Ola Fodnes
Gina Kvaal

Variance in relative gene expression RT-qPCR analysis of Atlantic salmon - before, during and after delousing

Bachelor's project in Biotechnology
Supervisor: Ann-Kristin Tveten
May 2020

NTNU
Norwegian University of Science and Technology

Jan Ola Fodnes
Gina Kvaal

Variance in relative gene expression RT-qPCR analysis of Atlantic salmon - before, during and after delousing

Bachelor's project in Biotechnology
Supervisor: Ann-Kristin Tveten
May 2020

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biological Sciences Ålesund

NTNU
Norwegian University of
Science and Technology

Preface

This topic was chosen based on our interest in nucleic acid techniques, and the way acute stress can affect gene expression. Our work was fully conducted at the laboratories of NTNU Ålesund. We would like to extend our gratitude for the trust and freedom we were given, and for everyone who answered a passing question or opened a locked door. Without you, no pipetting would have been done after the first week, due to pipette tip shortage. We would like to thank our supervisor Ann-Kristin Tveten, for great and reliable support and guidance through this project, and through the last three years. Thank you for your patience and for answering our many, many questions these last months. Additionally, we would especially like to thank Sophie Pursti, for spending countless hours proofreading and lending an unbiased opinion throughout. A thank you is also extended to Ola Brandshaug, for consultation and insight into fish welfare.

Abstract

The expression of stress-related genes in Atlantic salmon (*Salmo salar*) is of major interest when it comes to welfare studies in farmed fish. Having a greater understanding of the stress response at a genetic level in fish can aid in improving fish welfare. Mechanical delousing is a known stressor of Atlantic salmon. Tissue samples were collected from test groups before (B), during (D), and after (AA) mechanical delousing. Gene expression analysis was performed by RT-qPCR with six stress-related genes of interest (*hsp70, p53, pcna, nrf2, lox5* and *tnfα*), which were normalised against a reference gene (*ef1aα*). *ef1aα* was chosen because of its stability and proven reliability as a reference gene in Atlantic salmon.

Analysis of variance was used to detected significant inter-group differences in relative gene expression from the RT-qPCR assays. Statistically significant variances were detected in groups B and D, but not in group AA. It is indicated that total sampling time influences the statistical trends in gene expression, where shorter sampling time per individual produces more stable results. Groups B and D had significantly larger sampling time relative to group AA. This may result in different gene expression between the first and last individual per group. Further research is recommended to standardise sampling.

Sammendrag

Ekspresjonen av stressrelaterte gen i laks (*Salmo salar)* er av stor interesse når det kommer til velferdsstudier hos oppdrettsfisk. Det å ha en bedre forståelse for stressrespons på et genetisk nivå, kan hjelpe til med å forbedre fiskevelferden. Mekanisk avlusning er en kjent stressfaktor for laksen. Vevsprøver ble samlet fra testgrupper før (B), under (D), og etter (AA) avlusing, og analysen av genekspresjon ble utført via RT-qPCR med seks stressrelaterte målgen (*hsp70, p53, pcna, nrf2, lox5* og *tnfα*), som så ble normalisert mot et referansegen (*ef1aα)*. *ef1aα* ble brukt som referansegen basert på dets stabile og tidligere viste pålitelighet som referansegen i laks.

Det ble brukt en variansanalyse for å avdekke signifikante forskjeller i genekspresjon mellom gruppene fra RT-qPCR forsøkene. Det ble påvist statistiske signifikante forskjeller i gruppe B og D, men ikke i gruppe AA. Det er indikert at tiden det tar å samle prøvene, påvirker de statistiske trendene i resultatene, der raskere prøvetaking per individ gir mer stabile resultater. Det ble brukt betydelig lengre tid for gruppe B og D, realtivt til gruppe AA. Dette kan resultere i ulik genekspresjon mellom første og siste individ per gruppe. Videre forskning er anbefalt for å standardisere prøvetakningen.

Contents

1 Introduction

Norway was, by 2018, the world's largest exporter of farmed salmon, with 1.3 tons salmon sold and 1100 farming facilities alongside the coast (1). It represents 93% of Norway's aquaculture production. Population density has grown in farming areas, and *Lepeophirus salmonis* (salmon louse) has become an increasing problem in aquaculture in the recent years (2). The salmon louse is a naturally occurring parasite in salmonids of marine environments in the northern hemisphere. It feeds on the skin, mucus, and blood of the fish. This may lead to physical damage, skin erosion, secondary infections, immunosuppression, and chronic stress in the fish (3). This results in impaired growth, and production loss (4). Larva from the louse may also transfer from farmed fish to wild fish. The salmon louse's great infection potential and the high number of potential hosts increases the potential for injury and fatality in both farmed and wild fish. This makes salmon lice one of the most serious issues in Norwegian aquaculture today (5).

Initially delousing relied on chemical treatments, but frequent treatments led to resistance to the chemicals used. New ways of delousing became necessary. This came in the form of cleaner fish and mechanical delousing. Atlantic salmon, and fish in general, are protected under the Norwegian Animal Welfare Act. § 3 states that animals have an intrinsic value regardless of the value they hold to humans, that they are to be treated well and to be protected against danger and unnecessary stressors and strain (6). Mechanical delousing is a rough treatment method that could be considered to be inhumane, considering the animals' intrinsic value.

While the louse itself rarely leads to host death, the delousing is both a direct and indirect cause of death. A common occurrence for non-medical delousing is crowding, which in itself is considered to be a stressor and detrimental to the fish. Delousing has therefore been identified to represent a significant challenge to fish welfare (5).

To study the effects delousing causes as a stressor on a genetic level, gill tissue was collected from farmed Atlantic salmon at Frøya from three groups. The groups represented individuals before delousing (group B), during delousing (group D), and after delousing (group AA). Gill tissue from a control group was also collected, from a fish farm facility at Furnes. RT-qPCR was performed to determine the genetic expression of the selected target genes. Six genetic markers were chosen, and their relative gene expression compared to one reference gene (5).

This thesis presents how the different genes of interest are expressed relative to the reference gene *ef1aα* and observing the interspecific and intraspecific differences for the gene expression in all sample groups. Samples were collected from 16 different individuals in all groups, and samples 1-12 were used in this analysis. All three groups were collected during different time intervals, group B during a sampling time of 144-168 minutes, group D of 72-96 minutes, and group AA of 24-36 minutes. If this influenced the gene expression, is also of interest and will be discussed.

1.2 Dictionary

ANOVA - Analysis of Variance
cDNA – complimentary/copied DNA
Ct – threshold cycle
Cq – quantitative cycle
ΔCt - delta, cycle threshold
ΔΔCt - delta delta, cycle threshold
DNA - deoxyribonucleic acid
dsDNA – double-stranded DNA
ef1aα - gene that codes for elongation factor 1Aα, all genes are written in cursive
GOI – gene of interest, target gene in RT-qPCR analysis
Housekeeping gene – often a used as a reference gene (REF)
hsp70 – gene that codes for the protein Heat shock protein 70
lox5 - gene that codes for arachidonate 5-lipoxygenase
mRNA – messenger RNA
nrf2 - gene that codes for nuclear factor erythroid-2-related factor
pcna - gene that codes for proliferating cell nuclear antigen
REF – reference gene, used for normalising data in RT-qPCR assay
RNA – ribonucleic acid
RT-qPCR – Reverse Transcription Quantitative Real-Time Polymerase Chain Reaction
SYBR green – fluorescent reporter, used in RT-qPCR
Target gene – also known as gene of interest (GOI)
T_m - melting curve temperature
tnfα - gene that codes for tumour necrosis factor α

1.3 Theory

Stress response in fish

Stress can be defined as any condition that causes a physical or physiological discomfort, which results in release of stress-related hormones or leads to specific physiological responses (7). Crowding, general handling, change in water quality, change in oxygen levels, and different delousing methods as thermal, mechanical, and freshwater treatments are examples of stressors in fish (5). Stress can be physical, physiological, or environmental, and may also be either short term or long term. The physiological responses to environmental stress in fish, has been categorised as primary, secondary, and tertiary responses. These also range from positive eustress, to the negative and potentially harmful distress. The primary response is categorised by adaptive processes, like mobilising energy for flight-or-fight responses. This happens in the form of a neuroendocrine release, like adrenaline and noradrenaline. Cortisol is also released as a primary response to an environmental stressor. It is a known stress hormone, it triggers several secondary and tertiary stress response mechanisms (8). This has made cortisol a common candidate as a biomarker of stress (4, 9).

The secondary response includes changes in tissue, plasma ion and metabolic levels, haematological features, and upregulation and activity of heat shock proteins. This has made transcriptional targets good candidates for measuring or identifying a secondary response at the cellular level. The effects may be an increase in respiratory frequency, increase in blood pressure and mounted immune responses. These are for the most part reversable. Persistent distress leads to a tertiary stress response. The tertiary stress response refers to the animal's behaviour, which can be swimming pattern, change in growth, condition, resistance to disease, metabolism, and reproductive activity. These maladapted processes is potentially irreversible, and may make the fish more vulnerable to infections from pathogens and parasites, and overall decreased welfare (7, 8).

Delousing as a stressor

The salmon louse has been a problem in the industry since its start in the seventies. The development of treatments has led to a more frequent use of non-chemical treatments such as cleaner fish and mechanical delousing. Such treatments may lead to stress and physical injuries in the treated fish (5). Mechanical delousing is done through mechanical force (friction) to remove the louse from the fish's surface. This eliminates the need for potentially harmful chemicals, which could influence the consumer or the environment (10). One form of mechanical delousing was developed by Marine Harvest and tested by SINTEF, dubbed HydroLicer, which delouses the fish through water turbulence. The fish is herded into a chamber where water turbulence is formed, and a the current "lifts" the louse off from the surface of the fish (11). The fish in this study were deloused via a pipe system, where the velocity the fish reached was high enough to force the louse from the surface.

Non-medical delousing often calls for crowding, which is a stressor alongside with change in temperature, and possibly physical injuries. While the full extent of the consequences of delousing are unknown, it is clear that frequent handling and mechanical delousing has a considerable negative effect on fish welfare (5). Welfare is connected to quality of life, and that includes good health and the fish's own experience of their surroundings, which include feelings such as fear and pain. By performing gene expression analyses, it is possible to detect key genes in an organism's response to environmental stress (12). It is therefore useful to study the gene expression of stress related genes, to analyse their expression under different conditions.

Gene expression analysis

RT-qPCR

Reverse transcription quantitative real-time PCR (RT-qPCR) has become a well-established method for quantifying levels of gene expression. This is achieved by isolating RNA from the tissue under study, then converting it to complementary DNA (cDNA) by a reverse transcriptase. The RNA used may be total RNA or transcripts (mRNA), but mRNA is used in gene expression analysis. The cDNA is then used as a template for the qPCR reaction. RT-qPCR may be performed in either a one-step or a two-step assay. In a one-step assay, the reverse transcription and qPCR is combined in a single tube with buffer, sequence-specific primers, reverse transcriptase and DNA polymerase. In a two-step assay the reverse transcription and qPCR occur as separate runs. Both have their advantages and disadvantages. The two-step assay allows for the creation of a cDNA pool to be stored and used at one's own discretion. It also allows for greater optimisation of the reverse transcription and the qPCR steps. qPCR quantitates by measuring the number of amplicons in real time, via a fluorescent signal. This fluorescence is achieved through reporters such as dyes or probes, that emit a signal when fluoresced by the qPCR instrument at the end of each cycle. A qPCR amplification plot will show the number of cycles required to reach a threshold level, known as the threshold cycle (Ct) or quantification cycle (Cq). It is at this threshold the fluorescent signal of the amplicons exceeds the background noise. This occurs during the geometric phase in PCR, which means that the efficiency is constant cycle-to-cycle (13).

SYBR Green

SYBR Green is a fluorescent dye, commonly used in many different types of molecular techniques. SYBR Green is intercalating, meaning it will bind between the strands of nucleic acids and then fluoresce. It binds preferentially to dsDNA, though it does bind to RNA with lesser sensitivity. As with any reagent or method, it has its advantages and disadvantages. Its non-specificity is both an advantage and disadvantage. One vial of SYBR Green may be used for a wide variety of assays regardless of genes of interest, or nucleic acids. This makes SYBR Green an economic choice. It does however make it necessary to perform quality control measures, such as a melting curve analysis to control primer dimers and unspecific products. This adds more work. Additionally, there have been reports of PCR inhibition when working with greater concentrations of SYBR Green, instability and poor sensitivity compared to newer intercalating dyes (13, 14).

Genes of interest (GOI):

The target genes used in this study were selected based on previous studies on Amoebic gill disease (AGD) by Marcos-López *et al.* (2018). This helped in selecting genes of interest related to cell proliferation and apoptosis, and oxidative/cellular stress. The genes selected include genes for heat shock protein 70 (*hsp70*), a marker of apoptosis (*p53*), a marker of cell proliferation *(pcna),* an oxidative stress marker (*nrf2*), a marker of inflammation (*lox5*), and a cytokine (*tnfα*) (15).

hsp70

Heat shock protein 70 (Hsp70, coded by the gene *hsp70*) is a central component in the cellular network of molecular chaperones and assists a large variety of protein folding processes in the cell. These processes include the folding and assembly of newly synthesised protein, refolding of misfolded and aggregated proteins, membrane translocation of organellar and secretory proteins, and control of the activity of regulatory proteins. This occurs by transient association of Hsp70 substrate binding domain with short hydrophobic peptide segments within the substrate protein. What drives the substrate binding and release cycle is the switching of Hsp70 between a low-affinity ATP bound state and a high-affinity ADP bound state. ATP-binding and hydrolysis is therefore essential for chaperone activity in Hsp70.

Hsp70 is also heavily relied on during *de novo* folding in eukaryotic proteins, and this reliance on Hsp70 chaperones increases under stress conditions. Mutated versions of proteins require more attention by Hsp70 chaperones, to preserve the function of the proteins. Hsp70 therefore functions as a capacitor, which buffers destabilizing mutations. Mutations like these are only made apparent when the need for Hsp70 chaperones exceeds their capacity, such as during stress conditions (16).

p53

p53 (coded by the gene *p53*) is a tumour-suppressor protein which uses a sequence-specific transcription factor that binds DNA as a tetramer, and by that activates or represses transcription in a large number of genes. It is critical in prevention of tumour development and is known to aid in causing cell cycle arrest and apoptosis in response to a wide range of cellular damage. It also has an important function of inhibiting growth of abnormal or stressed cells. p53 is activated by a number of different signals that may occur during tumour development in a cell, including carcinogen-induced DNA damage, telomere erosion, aberrant proliferative signals, hypoxia, and loss of adhesion or survival signals.

p53 can therefore interfere at several stages during the carcinogenic process. The activity of p53 in cells has a strong inhibitory effect on cell growth, so it is important to restrain the effect p53 has under non-stress conditions. Regulation of DNA binding activity is also controlled, and activity may be induced by stress (17).

pcna

Proliferating cell nuclear antigen (PCNA, coded by the gene *pcna*) is essential in nucleic acid metabolism. The protein is essential for DNA replication, is involved in DNA excision repair, and has been shown to be involved in transcription of DNA to RNA. It does this by encircling the DNA and slide bidirectionally along the structure. PCNA also acts as the processivity factor for DNA-polymerase epsilon, by tethering the polymerase catalytic unit to the DNA template for fast and processive DNA synthesis. In addition, PCNA has been shown to interact with other cellular proteins involved in cell cycle regulation and check point control (18).

nrf2

Nuclear factor erythroid-2-related factor (NRF2, coded by the gene *nrf2*) is a transcription factor which functions as a master regulator of the cellular redox homeostasis. The target genes of NRF2 consists of a high number of antioxidant enzymes and proteins that are involved in xenobiotic detoxification (19). They also repair and remove damaged or misfunctioning proteins, inhibits inflammation, and include other transcription factors (20). NRF2 also has a function in resistance to oxidative stress, by controlling the basal and induced expression for an array of genes responsible for the antioxidant response, to regulate the physiological response and outcome of oxidant exposure. The function NRF2 has in regulating an antioxidant defence has an evolutionary basis and has evolved to aid in resistance to ER stress, oxidative stress, and metal toxicity (21). Marcos-López *et al.* (2018) found *nrf2* to have such a consistent expression, that it showed a statistical significant difference, despite having a low fold change value (15).

lox5

Arachidonate 5-lipoxygenase (LOX5, coded by the gene *lox5*) has a central role in the synthesis of leukotrienes, inflammatory mediators that play an important role in allergies, and innate and adaptive immunity (15). It is a member of the lipoxygenase gene family, and the encoded protein catalyses the multi-step conversion from arachidonic acid to leukotrienes (22). L*ox5* is primarily expressed in bone-marrow-derived cells, including macrophages, monocytes, and neutrophils (23).

tnfα

Tumour necrosis factor α (TNFα, coded by the gene *tnfα*) is an inflammatory cytokine which is produced mainly by macrophages and monocytes during acute inflammation. It is also responsible for a wide range of signals leading to necrosis or apoptosis, internally in cells. It also acts as an important resistance to infections and cancers (24). TNFα sends signals through two transmembrane receptors, TNFR1 and TNFR2. It regulates many critical cell functions, like cell proliferation, survival, differentiation, and, as mentioned, apoptosis. TNFα has been given a role of master-regulator of inflammatory cytokine production, because of its pivotal role in regulating the cytokine cascade in many inflammatory diseases. Therefore, it has been proposed to be a therapeutic target for many of these diseases (25).

Reference gene (REF) *ef1aα* and *ef1aβ*

Elongation factor 1A (eF1A) is one of three elongation factors – eF1A, eF1B and eF2 - which are involved in the elongation cycle in protein biosynthesis. eF1A is activated through GTP binding and forms a ternary complex with aminoacylated elongator tRNAs (aa-tRNAs). This ternary complex of eF1A decodes the genetic information, meaning that the bases only pairs with their complimentary bases. This takes place on the ribosome, and the mRNA codon meets the anticodon of a cognate tRNA (26). The process of synthesising eukaryotic proteins is mainly defined in three phases: initiation, elongation, and termination. eF1A splits into two subunits, eF1Aα and eF1Aβ, and has a main function in the elongation phase of translation. The delivery step of aa-tRNA is catalysed by the eukaryotic Elongation Factor 1 (eEF1) (27).

By using a reference gene, one can compensate and minimise errors in the relative quantification of the mRNA transcripts in the qPCR, because reference genes are genes that show a stable expression regardless of the state of the cell. It serves as an internal reference, that other mRNA values can be normalised against. Gene expression of *ef1a* has consistently shown to be one of the most stable and suitable reference genes for Atlantic salmon, when being tested *in vitro* (28). Olsvik *et al.* (2005) tested the stability of six potential reference genes for Atlantic salmon with the goal of determining the most suitable genes for RT-qPCR analyses. *ef1a* expressed stable relative transcription levels in gills, liver, head, kidney, spleen, thymus, brain, muscle, and posterior intestine in six adult Atlantic salmon. It was therefore concluded to be one of the most suitable reference genes in RT-qPCR assays when testing for gene expression in Atlantic salmon (29).

Quality control

The concentration of both the RNA and cDNA (dsDNA) was measured and controlled with the Qubit™ Fluorometer (Invitrogen, Carlsbad, USA). It utilises target selective fluorescent dyes that emits fluorescence signals when selectively bound to DNA, RNA, or proteins. The detection range for the high sensitivity (HS) assay for dsDNA goes from 10pg/µL to 100ng/µL, and the detection range for HS assay for RNA goes from 250pg/µL to 100ng/µL. The sample concentration is measured against two provided standard solutions, after the instrument is calibrated.

1.4 Statistics

2 -ΔΔCt method for normalising RT-qPCR data

Livak and Schmittgen published the 2-ΔΔCt method for analysing the relative gene expression using RT-qPCR data in 2001 (30). The 2-ΔΔCt method has become the most common approach for calculating gene expression. The mathematical proof of the 2-ΔΔCt method will not be deduced in this paper.

To utilise the 2-ΔΔCt method there are some requirements for the assay design. Two groups of test samples, one "test" group, and the other a control group. A Calibrator will be derived from the control group, either one sample or the mean of the control group. One needs the Ct values from a gene of interest (GOI) and reference gene (REF).

Equation 1.

$$\Delta Ct = Ct(GOI) - Ct(REF)$$

Equation 2.

$$\Delta\Delta Ct = \Delta Ct(Test\ sample) - Ct(Calibrator)$$

Equation 3.

$$2^{-\Delta\Delta Ct}$$

The *ΔCt* is the normalisation of the samples and is the difference between GOI and REF (*eqn. 1*). This takes into account possible differences in RNA concentrations between samples. The *ΔΔCt* is calculated as the difference of *ΔCt* value (from *eqn. 1*) and the chosen Calibrator, which the samples are relative to. The relative gene expression value is the log transformation of the negative *ΔΔCt* value (from *eqn 2.*) with the base of 2 (*eqn. 3*). The number 2 in 2 to the power of negative *ΔΔCt,* is the Amplification factor (*E*), which is a 100% primer efficiency.

At the same time as Livak *et al*. published their method, Pfaffl published his own method for analysing the relative gene expression (31). The mathematical proof of the Pfaffl method will not be deduced in this paper.

Equation 4.

$$\frac{E_{GOI}^{\Delta Ct\ GOI}}{E_{REF}^{\Delta Ct\ REF}}$$

The Pfaffl method calculates the ratio between GOI and REF (*eqn. 4*). The *E* is the amplification factor and must be calculated before the given qPCR assay. This will give an amplification factor for each GOI and for the REF. The *ΔCt* is the difference between Ct value of the Calibrator and the Ct value of the test sample. The numerator is the relative quantities for the gene of interest, which is the amplification factor *E* (of GOI) to the power of the *ΔCt.* The same for the denominator, but with the values for the REF.

The $2^{-\Delta\Delta Ct}$ method and Pfaffl gives a normalised gene expression relative to control samples.

Later Vandesompele *et al.* (2002) published an alternative method to Pfaffl's method (32). The only difference between the methods are that the alternative method utilises the geometric mean of multiple reference genes (*eqn. 5*).

Equation 5.

$$\frac{E_{GOI}^{\Delta Ct\ GOI}}{GeoMean\ [E_{REF}^{\Delta Ct\ REF}]}$$

The alternative method presented by Vandesompele *et al.* gives a multi-normalised gene expression relative to control samples.

Statistical significance testing

The main statistical analysis used for this experiment was an Analysis of Variance (ANOVA). The ANOVA test allows a comparison of more than two groups at the same time, and to determine whether a relationship exists between the groups. With the ANOVA test one can determine the variability between and within samples (33).

It is assumed that the data samples from this experiment follows a normal distribution and the population variance is equal between samples.

The statistical significance is deduced from a calculated *F*-statistic. If there are no differences between the groups, $F = 1$. If there are differences, $F > 1$. The *F*-statistic follows an *F* one-tailed distribution.

Hypothesis testing for an ANOVA test; null hypothesis (H_o) wherein all population means are the same, or alternative hypothesis (H_1) wherein at least one population mean is different:

Equation 6.

$$H_0;\ \mu_1 = \mu_2 = \mu_3$$
$$H_1; not\ all\ means\ are\ equal$$

The ANOVA analysis consists of a series of calculations, which is collected in tabular from.

Calculate the mean within group (*eqn. 7*), where *i* is the group, and *j* is the sample, which gives the j_{th} sample in group *i* with the symbol x_{ij}. n_i is the number of samples in i_{th} group.

Equation 7.

$$\bar{x}_i = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i}$$

Equation 8.

$$\bar{x} = \frac{\sum_{i=1}^{a} \sum_{j=1}^{n_i} x_{ij}}{N}$$

Calculate the total mean, Grand Mean, which is the mean of the sample means (*eqn. 8*). *a* is the number of groups, and *N* is the sum of the total number of samples.

Equation 9.

$$SS_G = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (\bar{x}_i - \bar{x})^2 = \sum_{i=1}^{a} n_i * (\bar{x}_i - \bar{x})^2$$

Equation 10.

$$S_G^2 = \frac{SS_G}{a-1}$$

Calculate the between-group sum of squared differences SS_G, sum of squares; groups (*eqn. 9*). Then calculate the variance between groups, which is the SS_G divided by the degrees of freedom, *a – 1,* where *a* is the number of groups (*eqn. 10*).

Equation 11.

$$SS_E = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

Equation 12.

$$S_E^2 = \frac{SS_E}{N - a}$$

Calculate the within-group sum of squares SS_E, sum of squares; error (*eqn. 11*). Then calculate the variance within-groups, which is the SS_E divided by the degrees of freedom, $N - a$, where N is the sum of the total number of samples (*eqn. 12*).

Equation 13.

$$F_{obs} = \frac{S_G^2}{S_E^2}$$

The *F*-value (F_{obs}) is the comparison of the between group variance and the within-group variance (*eqn. 13*).

For this experiment a 5% significance level ($\alpha = 0.05$) was selected. The *F*-critical value (*F*-statistic) can be calculated as:

Equation 14.

$$F_{crit} = f_{\alpha\,((a-1),\ (N-a))}$$

The decision rule is to reject the null hypothesis (H_o) if the observed value F_{obs} is greater than the *F*-critical value, and accept the alternative hypothesis (H_1), or fail to reject the null hypothesis otherwise.

By using a computer software (such as Microsoft® Excel), a *p*-value can be calculated. The decision rule is to reject the H_o and accept the H_1, if and only if the *p*-value is less or equal to the significance level = α (33).

Statistical correlation

The other statistical analysis used for this experiment was a linear correlation measurement. The linear correlation measures the linear association between two sample variables *x* and *y*. The measure of linear relationship generates the correlation coefficient *r* (33).

Equation 15.

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2)\,(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2)}}$$

The numerator, s_{xy}, is the covariance which is the sum of matches and mismatches among the pairs *x* and *y*. The denominator is the product of multiplying the standard deviation (SD) of variable *x* with the SD of variable *y*. The correlation coefficient *r* is calculated by dividing the covariance with the product of multiplying the SD of variable *x* and *y* (*eqn. 15*).

This calculated value *r* is a value between *+1* and *-1*, where *+1* is a total positive linear correlation, *-1* is a total negative linear correlation and *0* is no linear correlation (33).

2. Method and materials

2.1 Biopsy samples from Atlantic salmon

Atlantic salmon was collected by netting, then sedated before euthanasia. Average size of the fish was 2.2kg. Gill samples were collected from newly euthanized fish. 1 – 4 gill filaments were removed by tweezer and scissors and placed into individual tubes containing 500µL RNAlater® (Qiagen, Hilden, Germany) for conservation. Stored at 4◦C for 20 hours followed by storing at -80◦C.

Individuals from group B, D and AA (16 individuals per group) came from a fish farm at Frøya. Sampling time per individual was 12-14min for group B, 6-8min for group D, and 2-3min for group AA. Group B samples were collected before delousing, group D during delousing. Group AA were collected 18 hours after, when the fish were eating as normal. Mechanical delousing by well boat were the main stressor.

The control group was collected and handled as the test groups B, D and AA, without any stressor. The origin of the four individuals came from a fish farm at Furnes.

This concession has the license number: M VS0016.

2.2 Lysis and RNA extraction

TissueLyser II (Qiagen, Hilden, Germany), RNeasy® mini kit (ID: 74104, Qiagen, Hilden, Germany) and RNase-Free DNase I Set (ID: E1091, OMEGA bio-tek, Norcross, USA) were applicated for lysis and RNA extraction.

Gill tissue samples were weighed and transferred to TissueLyser II tubes with 5mm stainless steel beads on ice, with RLT Buffer (Qiagen, Hilden, Germany) using tweezers. Tweezers were washed in nuclease-free water three times, followed by 70% Ethanol three times between each sample. Initiated lysis by incubation on ice for 5min. Lysis and homogenisation were performed by Tissuelyser II for 2min at 25Hz, followed by incubation on ice for 1min. This procedure, lysing for 2min at 25Hz, followed by incubation on ice for 1min, was performed for a total of four times. Rearranged the tubes between each incubation according to manufacturer's handbook (*Appendix 5*).

RNA extraction proceeded according to manufacturer's protocol for RNeasy® mini kit (ID: 74104, Qiagen, Hilden, Germany) (*Appendix 3*). Performed on-column DNase digestion with RNase-Free DNase I Set (ID: E1091, OMEGA bio-tek, Norcross, USA) (*Appendix 4*) for samples 9B – 16B, 1D – 8D, 1AA – 16AA. Quality control was performed by quantification of RNA concentration with Qubit™ Fluorometer (Invitrogen, Carlsbad, USA) using the Qubit™ RNA HS Assay Kit (ID: Q32852, Invitrogen, Carlsbad, USA) (*Appendix 8*). Refer to *Figure 1.* and *Figure 2.* for tissue and RNA yield, respectively.

2.3 cDNA synthesis

First strand cDNA synthesis was performed immediately after RNA extraction. Utilised the qScript® cDNA Synthesis Kit (ID: 95047, QuantaBio, Beverly, USA) according to protocol (*Appendix 6*). Maximised RNA input (15µL) and cDNA synthesis was performed on a 2720 Thermo Cycler (Applied Biosystems, Waltham, USA). Quality control was performed by quantification of DNA concentration with Qubit™ Fluorometer (Invitrogen, Carlsbad, USA) using the Qubit™ 1X dsDNA HS Assay Kit (ID: Q33231, Invitrogen, Carlsbad, USA) (*Appendix 9*). Referring to *Figure 2.* for cDNA yield. RNA and cDNA pools were stored immediately at -80◦C. Working solutions of cDNA template for each sample was diluted from cDNA pools, to construct a cDNA concentration of 200pg/µL. Working solution of cDNA was stored at -20◦C.

2.4 qPCR primer design

Primer sequences for the genes *p53*, *pcna*, *lox5*, *nrf2* and *tnfα* were obtained from Marcos-López *et al.* (2018). These were again obtained from searches in GenBank®, BLAST queries of other fish species and published data of Atlantic salmon (15).
Primers for *ef1aα* and *ef1aβ* were obtained from Olsvik *et al.* (29), and *hsp70* were designed by Ann-Kristin Tveten using the NCBI primer design tool.

As a quality control measure the primer sequences were analysed by the NCBI Primer-BLAST tool (34). Entered primer pair sequences of each respective gene and selected primer specificity for Atlantic salmon (taxid:8030). There were predicted to be no significant cross specificity to non-relevant genes for *hsp70*, *pcna*, *lox5*, *tnfα*, *ef1aα* and *ef1aβ* primer pairs. For the gene *p53* and *nrf2* there was predicted cross specificity to numerous non-relevant genes. The majority of these generated PCR products were greater than 1000bp.

Table 1. Oligonucleotid primers used in the qPCR assay. The sequence and supplier for the forward (FWD) and reverse (RWD) primer for the given gene. The sequence is given in the 5'-3' direction for both FWD and RWD primers.

Target gene	Sequence (5' – 3')	Supplier
***hsp70* FWD**	TTC CGA CCT CTT CAG GGG AA	Eurogentec
***hsp70* RWD**	TTG GGC CTT GTC CAT CTT GG	Eurogentec
***p53* FWD**	CAT CAT CAC CCT GGA GAC A	Invitrogen
***p53* RWD**	CAC ACA CGC ACC TCA AAG	Invitrogen
***pcna* FWD**	GCC GTG ACC TGT CTC AGA TTG	Invitrogen
***pcna* RWD**	CCG AGA ACT TAA CGC CAT CCT T	Invitrogen
***lox5* FWD**	ATC CAC CAG ACA GTC ACA CAC CTT C	Invitrogen
***lox5* RWD**	GCC ACT CCA AAC ACC TCC GAG AC	Invitrogen
***nrf2* FWD**	GAG GGA CGA GGA TGG GAA G	Invitrogen
***nrf2* RWD**	ATC GGT GGT CTG CTG GAG	Invitrogen
***tnfα* FWD**	GTG TAT GTG GGA GCA GTG TT	Invitrogen
***tnfα* RWD**	GAA GCC TGT TCT CTG TGA CT	Invitrogen
***ef1aα* FWD**	CCC CTC CAG GAC GTT TAC AAA	Invitrogen
***ef1aα* RWD**	CAC ACG GCC CAC AGG TAC A	Invitrogen
***ef1aβ* FWD**	TGC CCC TCC AGG ATG TCT AC	Invitrogen
***ef1aβ* RWD**	CAC GGC CCA CAG GTA CTG	Invitrogen

2.5 qPCR assay

Performed qPCR assays with the PerfeCTa® SYBR® Green SuperMix, Low ROX™ kit (ID: 95056, QuantaBio, Beverly, USA) with 96 well-plates (VWR, Radnor, USA) on the AriaMx Real-time PCR System (Agilent Technologies, Santa Clara, USA). The qPCR assays were performed according to manufacturer's protocol (*Appendix 7*).

The plate setup for the experiment included four test samples, one control, and one no template control (NTC), all done with duplicate technical replicates (*Appendix 11*). The test groups, B, D and AA, respectively, were represented by 12 biological individuals. The same control sample (3C-a) was used throughout the entire experiment.

Applied all components (*Table 2.*) to light safe micro-centrifuge tubes for constructing Mastermixes for each respective gene. A cDNA template input of 400pg was obtained by applying 2µL of 200pg/µL working solution from each respective sample. A total of 13 qPCR runs were performed, four of which were excluded from this experiment due to invalid results. These invalid results were caused by primer contamination, inadequate Ct-values of reference genes and pipetting errors. The thermal profile was; 2min of hot start at 95°C, 3-step cycling for a total of 35 cycles of 15sec at 95°C, 30sec at 58°C, 30sec at 72°C, and a melt curve of 30sec at 95°C, 30sec at 65°C, 30sec at 95°C.

Table 2. One reaction of one gene with a total volume of 15µL, are constructed by 7.5µL SYBR Green SuperMix, 0.6µL for each primer, 4.3µL nuclease-free water and 2 µL of cDNA template. Every qPCR run had one NTC, four test samples and one control, in duplicates, which gives a total of 12 reaction for one gene. A Mastermix for one gene is constructed by scaling the components proportionally and applicating SYBR Green SuperMix, primers for the given gene and nuclease-free water to a micro-centrifuge tube. To make up for pipetting errors, 2 reactions were added. Applicated 13µL of Mastermix to one well, and then 2µL of cDNA template. For the NTC's 2µL of nuclease-free water were added.

Component	Volume 1rxn. (µL)	Volume 12+2rxn. (µL)	Final Concentration	Mastermix Volume 1rxn. (µL)	Template Volume 1rxn. (µL)
SYBR Green SuperMix (2X)	7.5	105	1x		
Forward primers (10 µM)	0.6	8.4	400 nM		
Reverse primers (10 µM)	0.6	8.4	400 nM		
Nuclease-free water	4.3	60.2	-		
Template (200pg/µL)	2	-	400 pg		
Final Volume (µL)	**15**	**182**	**-**	**13**	**2**

2.6 Calculating relative gene expression

Detection and collection were done by the AriaMx Real-time PCR System, which generated Ct-values and melting curve temperature (T_m). Exported data to Microsoft® Excel, excluded invalid results according to the Tukey's fence outlier detection method (35) and calculated relative gene expression according to the $2^{-\Delta\Delta Ct}$ method (*Appendix 1*).

2.7 ANOVA analysis

After calculating relative gene expression, the ΔΔCt-values of groups B, D and AA were used for statistical testing as according to Vandesompele *et al.* The ΔCt-values were used for the control group. Microsoft® Excel for Mac (v. 16.35) were used with the add-in program Analysis ToolPak for data analysis. *F*-statistic and *P*-values were calculated with the Analysis tool ANOVA: Single Factor (*Appendix 1*).

2.8 Calculating correlation coefficient

The linear correlation between weighted gill tissue and extracted RNA, the linear correlation between weighted gill tissue and synthesised cDNA and the linear correlation between extracted RNA and synthesised cDNA were calculated using the Microsoft® Excel function *CORREL(R_1; R_2)* (*Appendix 1*).

3 Results

Results are presented after data analysis. According to the outlier detection method, some Ct-values were excluded from further data analysis based on deviating melting temperature (*Appendix 1*). The Ct-values for control sample 3C-a (calibrator sample) of *ef1aβ* were absent and/or excluded. This applies to every major group, B, D and AA. The ΔΔCt-values are therefore only calculated using *ef1aα* as reference gene. All calculations were done as previously described (*Appendix 1*).

3.1 Gill tissue samples

The gill tissue samples were weighted to a mean of 7.8mg, 5.1mg, 8.5mg and 15.1mg for groups B, D, AA and control, respectively. Standard deviation (SD) for the groups were 3.1mg, 2.2mg, 5.2mg, 1.9mg (*Figure 1)* (*Appendix 1*).

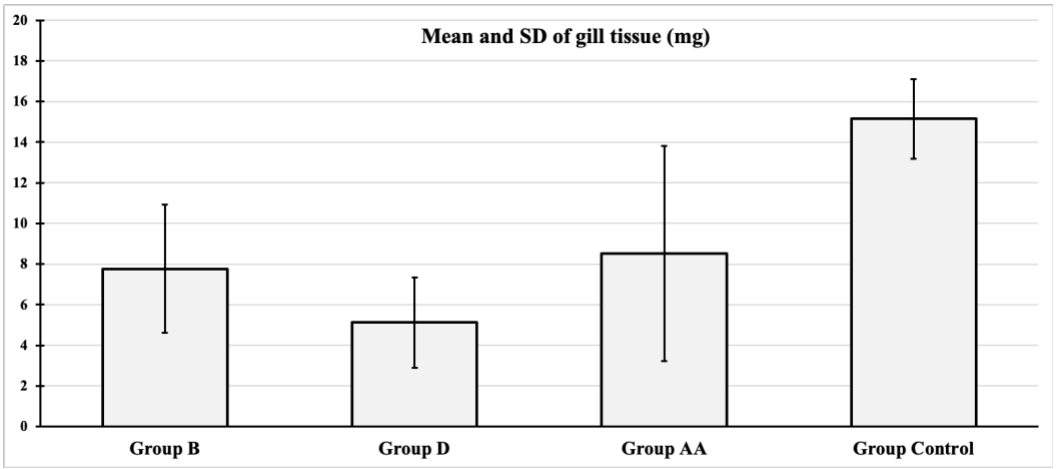


Figure 1. Weighted gill tissue (mg) mean and standard deviation for the respective groups. X-axis represents groups and Y-axis represent weight in milligrams.

Measured RNA concentration yield mean after extraction was 161µg/mL (SD=49) for group B, 181µg/mL (SD=40) for group D, 129µg/mL (SD=44) for group AA and 119µg/mL (SD=26) for control group. After cDNA synthesis this were measured to a mean of 38µg/mL, 48µg/mL, 26µg/mL and 49µg/mL for the groups B, D, AA and control, with a SD of 13µg/mL, 19µg/mL, 17µg/mL and 5µg/mL.

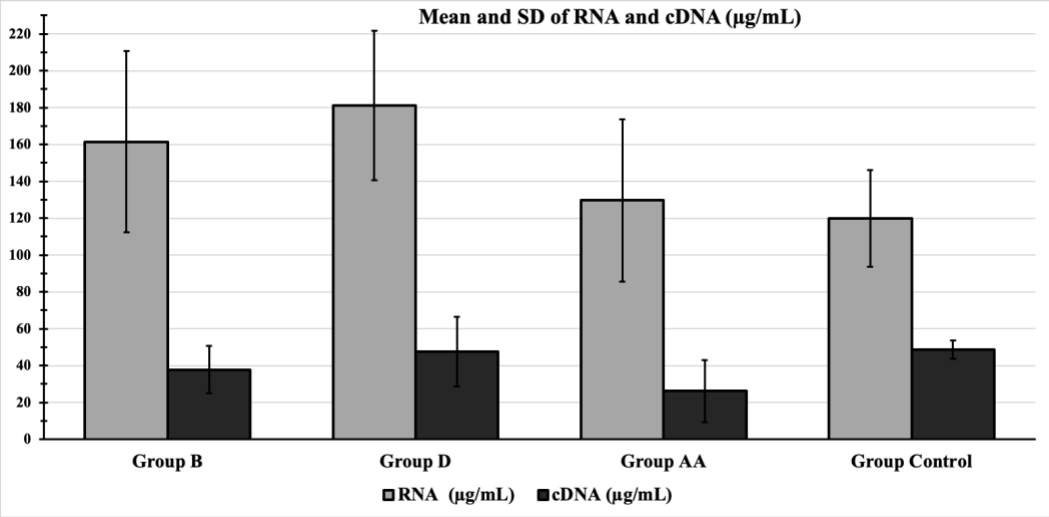


Figure 2. Mean and standard deviation of measured concentrations of RNA (grey) and cDNA (black) for the respective groups. X-axis represents groups and Y-axis represent concentration in µg/mL.

The correlation coefficient between weighted gill tissue and extracted RNA was -0.48, and -0.03 between weighted gill tissue and synthesised cDNA. The correlation coefficient between extracted RNA and synthesised cDNA was 0.41 (*Appendix 1*).

Table 3. Weighted gill tissue (mg), measured RNA concentration after RNA extraction (µg/mL) and measured cDNA concentration after cDNA synthesis for all samples for the experiment.

Sample	Gill tissue (mg)	RNA (µg/mL)	cDNA (µg/mL)
1B	6.73	180	47.1
2B	6.29	200	48.5
3B	11.86	65.8	23.2
4B	6.09	170	45.9
5B	7.45	160	45.9
6B	3.39	38.4	15.9
7B	14.65	160	52
8B	9.83	190	56
9B	5.05	190	25.2
10B	4.67	180	27.2
11B	8.75	190	42.3
12B	5.74	190	34.6
1D	3.5	200	34.8
2D	5.68	170	39
3D	6.92	200	35.6
4D	4.93	190	19.8

5D	2.67	190	38.7
6D	4.31	200	39.4
7D	5.45	200	50
8D	11.59	200	31.8
9D	5.2	200	68.7
10D	4.14	200	58
11D	5.06	200	54
12D	7.36	200	56
1AA	22.2	66	4.48
2AA	12.02	100	9.73
3AA	11.53	180	2.27
4AA	6.32	98	34.1
5AA	2.8	75	8.14
6AA	14.82	79	17.1
7AA	7.51	180	51
8AA	11.99	150	11.8
9AA	2.72	140	30.7
10AA	6.08	170	42.7
11AA	5.91	180	10.3
12AA	7.42	170	44
1C-a	16.54	140	47
1C-b	16.18	120	42.9
2C-a	14.98	79	41.6
2C-b	12.84	130	49.6
3C-a	11.71	130	53
3C-b	17.55	160	56
4C-a	15.1	110	50.7
4C-b	16.25	90	49.2

3.2 Analysis of variance

Descriptive statistics and ANOVA analysis for the experiment will be presented.

Group B – before delousing

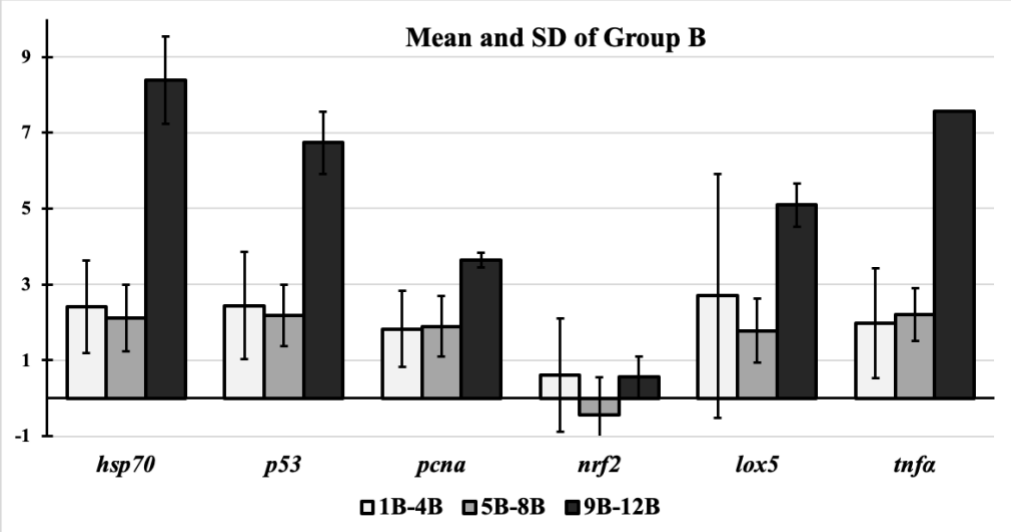


Figure 3. Mean and standard deviation of ΔΔCt values of sub-groups within group B. Sub-groups 1B-4B, 5B-8B and 9B-12B are represented by white, gray and black, respectively. X-axis represents the different genes of interest and Y-axis represents ΔΔCt values.

In group B *hsp70* ΔΔCt-values had a mean of 2.4, 2.1 and 8.3, and SD of 1.2, 0.9 and 1.2 for the sub-groups *1B-4B*, *5B-8B* and *9B-12B*. *p53* had a mean of 2.4, 2.2 and 6.7, and SD of 1.4, 0.8 and 0.8. *pcna* had a mean of 1.8, 1.9 and 3.6, and SD of 1.0, 0.8 and 0.2. *nrf2* had a mean of 0.6, -0.4 and 0.7, and SD of 2.2, 1.0 and 0.5. *lox5* had a mean of 2.7, 1.8 and 5.1, and SD of 3.2, 0.9 and 0.6. *tnfα* had a mean of 2.0, 2.2 and 7.6, and SD of 1.4 and 0.7 for the sub-groups *1B-4B* and *5B-8B*. There is not generated any SD of group *9B-12B*, based on excluded and/or lacking Ct-values of the samples *9B*, *11B* and *12B* (*Figure 3*).

The ANOVA analysis of group B indicates that there is statistically significant inter-group variance of certain genes. Both *pcna* and *tnfα* have *p*-values of 0.011 and are statistically significant at a significance level at $\alpha = 0.05$. *hsp70* and *p53* have a *p*-value < 0.01 and are statistically significant at a significance level of $\alpha = 0.01$. *nrf2* and *lox5* do not have any significant variance between their sub-groups (*Figure 4*).

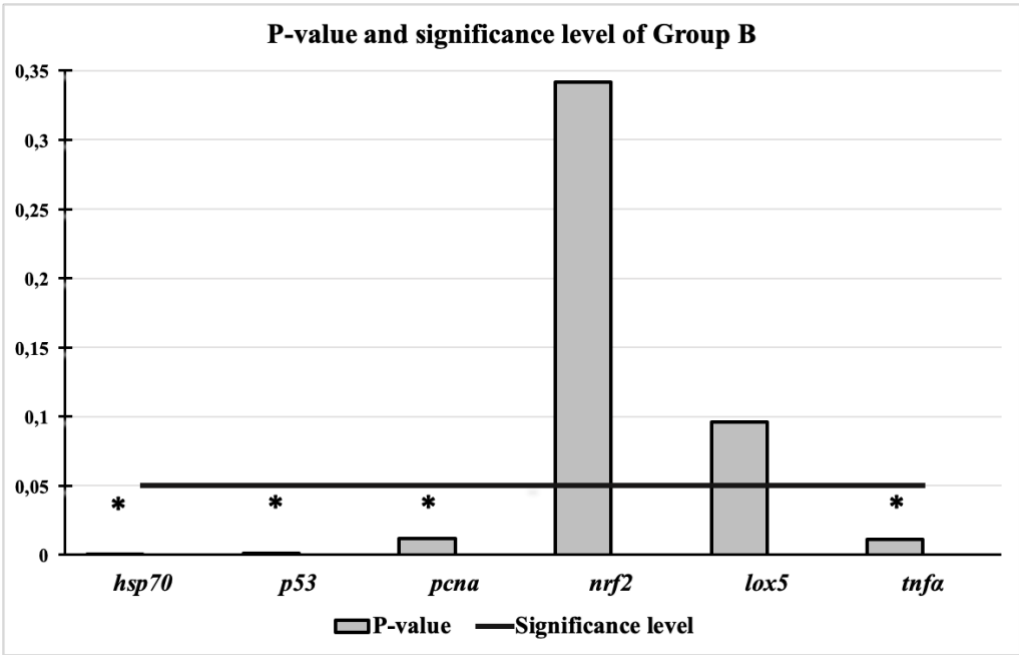


Figure 4. P-value generated from the ANOVA statistical analysis. Asterix indicates statistical significance at a significance level of α = 0,05. hsp70, p53, pcna and tnfα have a p < 0.05, and are therefore statistically significant. X-axis represents gene of interest and Y-axis represents p-value.

Group D – during delousing

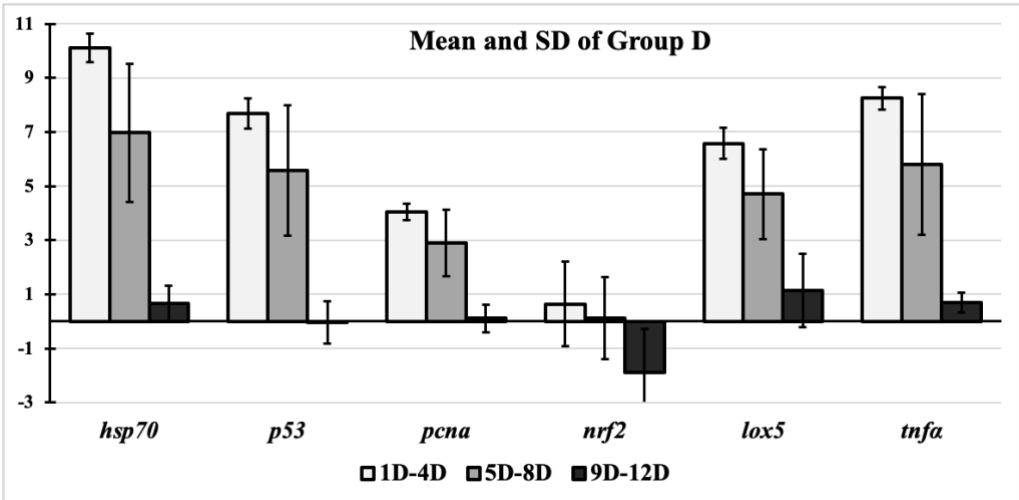


Figure 5. Mean and standard deviation of ΔΔCt values of sub-groups within group D. Sub-groups 1D-4D, 5D-8D and 9D-12D are represented by white, gray and black, respectively. X-axis represents the different genes of interest and Y-axis represents ΔΔCt values.

In group D *hsp70* ΔΔCt-values had a mean of 10.1, 7.0 and 0.6, and SD of 0.5, 2.5 and 0.6 for the sub-groups *1D-4D*, *5D-8D* and *9D-12D*. *p53* had a mean of 7.7, 5.6 and -0.03, and SD of 0.6, 2.4 and 0.8. *pcna* had a mean of 4.0, 2.9 and 0.1, and SD of 0.3, 1.2 and 0.5. *nrf2* had a mean of 0.6, -0.1 and -1.88, and SD of 1.6, 1.5 and 1.6. *lox5* had a mean of 6.6, 4.7 and 1.2, and SD of 0.7, 1.7 and 1.3. *tnfα* had a mean of 8.4, 5.8 and 0.7, and SD of 0.4, 2.6 and 0.4 for the three sub-groups (*Figure 5*).

The ANOVA analysis of group D indicates that there is statistically significant inter-group variance of certain genes. The genes *hsp70*, *p53*, *pcna*, *lox5* and *tnfα* have a *p*-value < 0.01 and are therefore statistically significant at a significance level at $\alpha = 0.01$. *nrf2* do not have any significant variance between sub-groups (*Figure 6*).

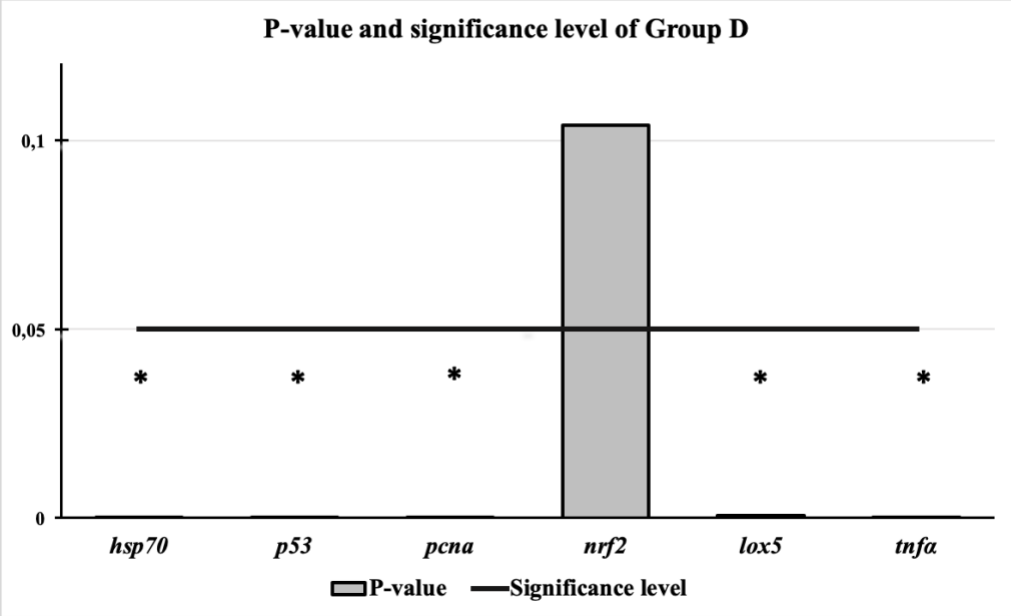


Figure 6. P-value generated from the ANOVA statistical analysis. Asterix indicates statistical significance at a significance level of α = 0.05. hsp70, p53, pcna, lox5 and tnfα have a p < 0.05, and are therefore statistically significant. X-axis represents gene of interest and Y-axis represents p-value.

Group AA – after delousing

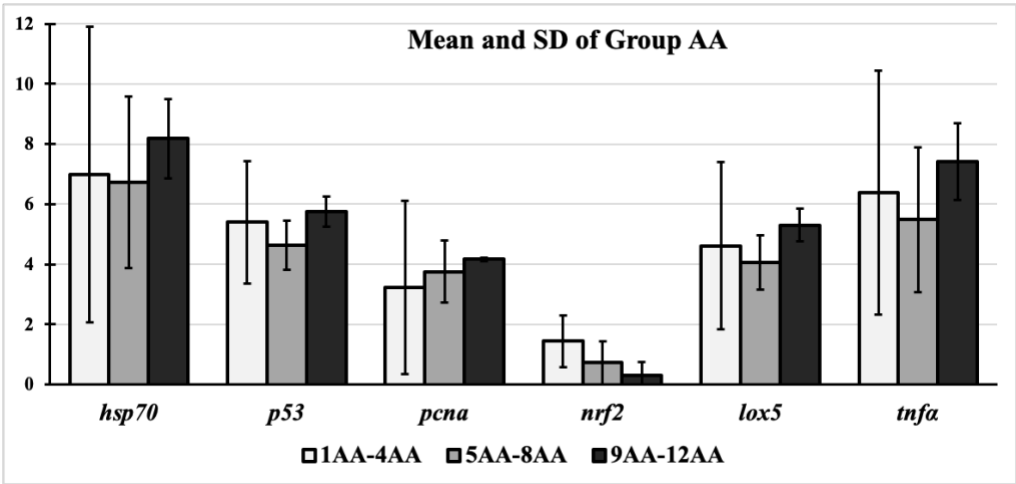


Figure 7. Mean and standard deviation of ΔΔCt values of sub-groups within group AA. Sub-groups 1AA-4AA, 5AA-8AA and 9AA-12AA are represented by white, gray and black, respectively. X-axis represents the different genes of interest and Y-axis represents ΔΔCt values.

In group AA *hsp70* ΔΔCt-values had a mean of 7.0, 6.7 and 8.2, and SD of 4.9, 2.8 and 1.3 for the sub-groups *1AA-4AA*, *5AA-8AA* and *9AA-12AA*. *p53* had a mean of 5.4, 4.6 and 5.8, and SD of 2.0, 0.8 and 0.5. *pcna* had a mean of 3.2, 3.8 and 4.2, and SD of 2.9, 1.0 and 0.1. *nrf2* had a mean of 1.4, 0.7 and 0.3, and SD of 0.9, 0.7 and 0.4. *lox5* had a mean of 4.6, 4.0 and 5.3, and SD of 2.8, 0.9 and 0.5. *tnfα* had a mean of 6.4, 5.5 and 7.4, and SD of 4.1, 2.4 and 1.3 for the three sub-groups (*Figure 7*).

The ANOVA analysis of group AA indicates that there is no statistically significant inter-group variance between sub-groups. All genes without exception have a *p*-value > 0.05. With a significance level of $\alpha = 0.05$ this is not statistically significant (*Figure 8*).

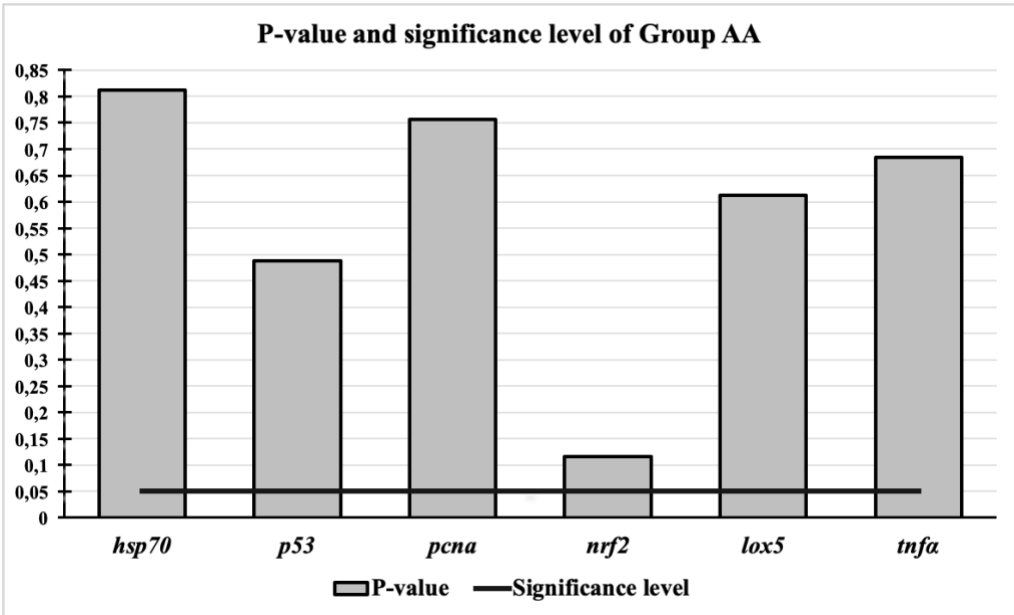


Figure 8. P-value generated from the ANOVA statistical analysis. Significance level of α = 0.05, indicated by the horizontal bar. None of the genes have a p < 0.05 and are therefore not statistically significant. X-axis represents gene of interest and Y-axis represents p-value.

Control group through major groups

The ΔCt-values were used for control group, being the ΔΔCt-values are the difference between ΔCt-values and calibrator, equals 0.

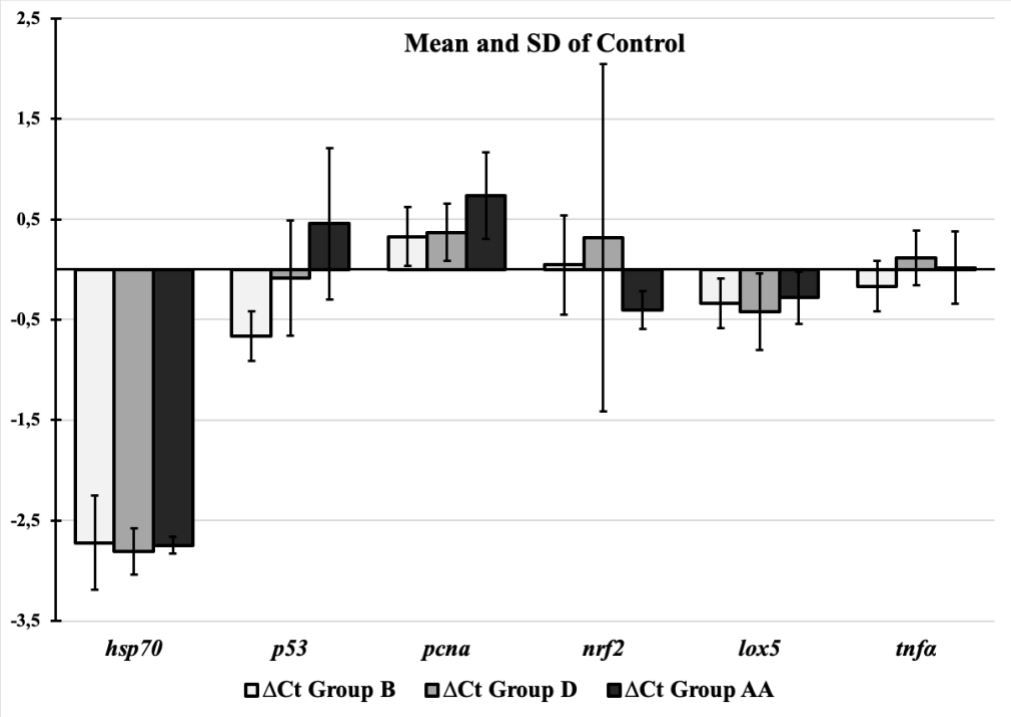


Figure 9. Mean and standard deviation of ΔCt values of control group between the major groups. The major groups B, D and AA are represented by white, gray and black, respectively. X-axis represents the different genes of interest and Y-axis represents ΔCt values.

In the control group, *hsp70* ΔCt-values had a mean of -2.7, -2.8 and -2.7, and SD of 0.5, 0.2 and 0.1 for the groups B, D and AA. *p53* had a mean of -0.7, -0.1 and 0.5, and SD of 0.2, 0.6 and 0.8. *pcna* had a mean of 0.3, 0.4 and 0.7, and SD of 0.3, 0.3 and 0.4. *nrf2* had a mean of 0.05, 0.3 and -0.3, and SD of 0.5, 1.7 and 0.2. *lox5* had a mean of -0.3, -0.4 and -0.3, and SD of 0.2, 0.4 and 0.3. *tnfα* had a mean of -0.2, 0.1 and 0.02, and SD of 0.3, 0.3 and 0.4 for the three sub-groups (*Figure 9*).

The ANOVA analysis of the control group indicates that there is no statistically significant variance between the major groups. All genes without exception have a *p*-value > 0.05. With a significance level of α = 0.05 this is not statistically significant (*Figure 10*).

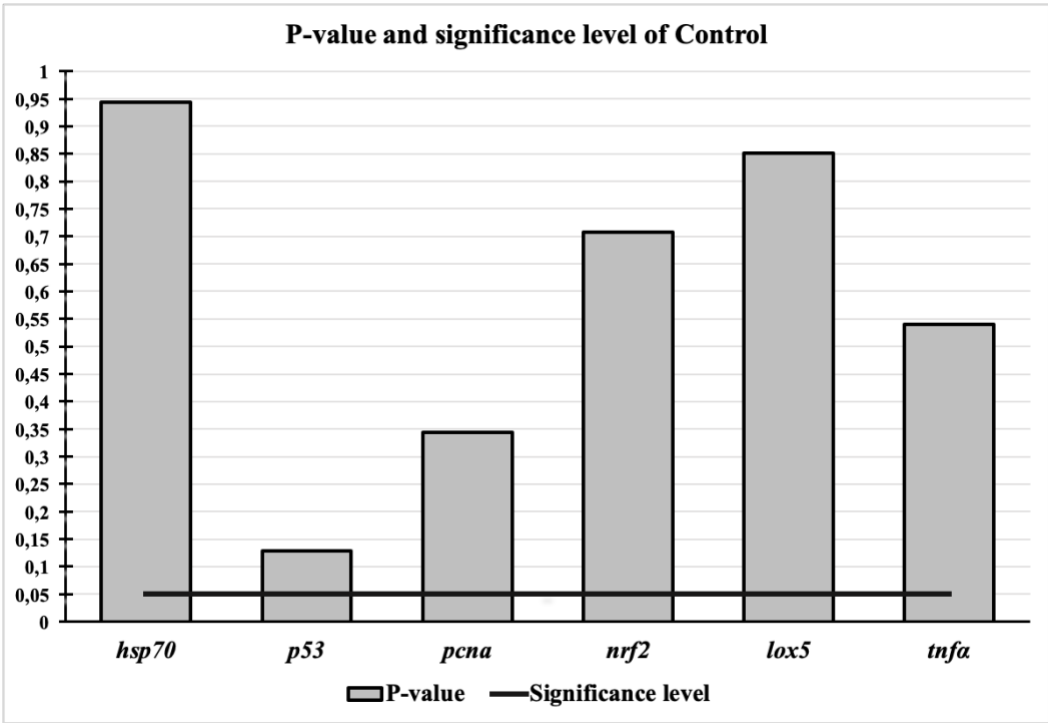


Figure 10. P-value generated from the ANOVA statistical analysis. Significance level of α = 0.05, indicated by the horizontal bar. None of the genes have a p < 0.05 and are therefore not statistically significant. X-axis represents gene of interest and Y-axis represents p-value.

4 Discussion

This experiment was an *in vivo* study of variance within genetic stress markers in three different groups of Atlantic salmon. It was not the intent of this study to determine any fold expression change between the major groups. The experiment included six stress related genes of interest and two reference genes, and its applicability is to shed light upon possible factors that contribute to inter-group (between sub-groups) variance and intra-group (within sub-group) variance of gene expression.

There was found a significant variance between sub-groups of various genes in groups B (*Figure 4*) and D (*Figure 6*) which are the before and during delousing groups, respectively. A significance level of $\alpha = 0.05$ indicated no significant difference between the group AA, after delousing (*Figure 8*) and in the control group (*Figure 10*) (*Appendix 1*).

The following topics will discuss trends in the data, possible factors of variance, sources of bias, and further research.

4.1 Statistical trends and the influence of sampling differences

The ANOVA analysis tests for difference between population means, and if there are significant differences between means. The test does not indicate which means differ. Visualisation of the data, as presented by figures, aids in the interpretation of the results.

There was a statistically significant variance in the gene expression of *hsp70*, *p53*, *pcna* and *tnfα* in group B (*Figure 4*). *Figure 3* implies an overall larger ΔΔCt-value of sub-group *9B-12B*. This indicates a trend where sub-group *9B-12B* is the source of significant inter-group variance within all of the genes. This is supported by the numerical results (*Appendix 1*). The SD is also consequently larger in the sub-group *1B-4B* compared to the other sub-groups. The large intra-group variance of *1B-4B* in *lox5* originates from sample *3B*, with a conspicuously large Ct-value, corresponding with a disproportionately low Ct-value of the reference gene. When compared to the other samples in *1B-4B,* this data should have been excluded from the final results.

There was a statistically significant variance in the gene expression of *hsp70*, *p53*, *pcna*, *lox5* and *tnfα* in group D (*Figure 6*). Observations of *Figure 5* implies an overall lower ΔΔCt-value of sub-group *9D-12D*. This indicates a trend where sub-group *9D-12D* is the source of significant inter-group variance within the genes. This is supported by the numerical results (*Appendix 1*). The intra-group variance is also consequently larger in the sub-group *5D-8D* compared to the other sub-groups.

There was no statistically significant variance in gene expression within the sub-groups in group AA. The SD is larger in the sub-group *1AA-4AA* compared to the other sub-groups (*Figure 7*). Besides this conspicuously large intra-group variance, there are no indications of systematic variance within *1AA-4AA* and are therefore treated as random error.

As for the control group, there was no statistically significant variance in gene expression of control sample *3C-a* between the major B, D and AA groups. This corresponds with the hypothesis, where the control sample should not differ significantly compared to the groups within the same gene. The SD's are < 1 in every group besides group D in *nrf2*. It originates from sample *3C-a (10D-12D+1D)* which has a conspicuously large Ct-value of *nrf2*, where the Ct-value of reference gene corresponds with the other samples. When compared to the other samples in *nrf2* of group D, this data should have been excluded from the final results.

A possible contributing factor for the observed trends in the two first groups, B and D, might be differences in sampling time. mRNA concentration and composition in a cell is under constant regulation and therefore not static. Most eukaryotic mRNA has a half-life of 30 minutes. The mRNA's with the shortest life span are those that encode for proteins with rapidly changing concentrations to accommodate the cell's current needs, such as during a stress response (35). Acute stress responses are quickly down regulated as the stress stimuli ceases, and the gene expression will reflect that. Delousing is an acute form of stress and the stress responses are down regulated fairly quickly after exposure. This rapid down regulation compounded by the short life span of mRNA, may be why the stress related genes of interest are expressed different in sub-group *9D-12D* than in sub-groups *1D-4D* and *5D-8D*.

Gill tissue for this experiment were all collected at different time spans for all three major groups. Group B were collected during a time period of 12-14min per sample, totalling in 144-168min for all 12 individuals. Tissue collection from group D took 6-8min per individual and 72-96min total.

Lastly, sampling from group AA took 2-3min per individual, or 24-36min total. This is a significantly shorter time span compared to groups B and D. Group B had the longest sampling time span, from start to finish. Too long a time span between the first and last sampling leaves later individuals open for other environmental stressors that could cause significant differences in gene expression. Stressors like general handling, crowding, water quality, and time passed since last feeding are all stressors in a fish farm environment. These factors combined with short mRNA half-life, are potential sources of variance for group B. The environmental factors affecting individual *1B* and individual *12B* could be drastically different, which could be why the genes in sub-group *9B-12B* are showed to differentiate significantly compared to sub-groups *1B-4B* and *9B-12B*.

nrf2 stands out compared to the others, since it did not show significant variation in either of any groups. This is similar to the result found by Marcos-Lopez *et al.* (2018), which found *nrf2* to have a significant stable expression (14).

Another thing that may affect variance, is differing efficiencies of disruption and homogenisation of the gill tissue. The tissue samples in group C were markedly larger than the tissue samples from groups B, D, and AA. While there is nothing presented in the results which indicates correlation between only sample size and RNA output, a correlation may exist between sample size, lysis-method, and RNA output. Samples *1AA-8AA* was disrupted for 1.5min, put on ice for 1min, and disrupted again for 1.5min. The rest of the tissue samples from group AA (sub-group *9AA-12AA*), as well as all samples from groups B, D, and C were disrupted for 2min and put on ice for 1min, followed by the 2min of disruption and incubation for 1min for a total of four rounds. The later method of disrupting the gill tissue proved to be the most optimal for greater RNA output. The average RNA output was 116µg/mL for *1AA-8AA*, 165µg/mL for *9AA-12AA*, 160µg/mL for *1B-12B*, 196µg/mL for *1D-12D*, and 120µg/mL for group C (8 samples) *(Table 3*). Comparing the weight and lysis-method, *9AA-12AA*, *1B-12B*, *1D-12D* all had a greater RNA output than *1AA-8AA*. Group C underwent the same lysis-method as *9AA-12AA*, *1B-12B*, and *1D-12D*, but had a notably lesser RNA output than these groups. This could be because these tissue samples weighed considerably more than samples from group B, D, and AA, which were all similar in weight (relatively, when compared to group C). Since group C was of a larger size, the lysis method used for group C could have been changed to achieve a more similar RNA output to group B, D, and *9AA-12AA*.

For all cDNA synthesis runs, 15µL isolated RNA was added, regardless of concentration. There is no correlation between RNA input and cDNA output in cDNA synthesis (*Appendix 1*). There are also no indications of correlation between QC data and the Ct-values.

4.2 Sources of bias

Plate setup

For the groups B, D and AA the qPCR plate setup was one NTC, four samples, followed by 3C-a (no treatment control). A total of 12 individual samples per major group, divided by four samples per qPCR run, equals three qPCR runs per major group. In the qPCR plate setup for groups B and AA, the four samples were divided into *1X-4X*, *5X-8X* and *9X-12X*, where "*X*" is either groups B or AA. The qPCR plate setup for group D was *2D-5D*, *6D-9D* and *10D-12D + 1D*. Under the ANOVA analysis the samples were arranged as qPCR runs of groups B and AA; *1X-4X*, *5X-8X* and *9X-12X*, where "*X*" is now either group B, D or AA. This could lead to unfortunate plate-to-plate inter-assay variance between sub-groups in group D when doing the statistical testing. In the statistical sub-group *1D-4D,* sample *1D* was not run on the same qPCR plate as the three other samples. Therefore, sample *1D* and samples *2D-4D* were not exposed to the same within-plate intra-group variance. Sources of inter-group variance are especially connected to manual execution as pipetting errors and handling. Automated systems and data sampling software do have some variance, but this is almost negligible compared to human interference.

For validation purposes regarding inter-assay variance all samples in one group should be run on the same qPCR plate and compared to see if there is acceptable variance between the inter-assay variances. That was not done in this experiment.

Calculating relative gene expression

A possible major downside with the Livak and Schmittgen $2^{-\Delta\Delta Ct}$ method is that it assumes a primer efficiency equal to 100% (amplification equal to 2) for both primer pairs of gene of interest and reference gene (30). This is an assumption that is almost always not the case (31, 32, 36). The Pfaffl method does take in to account these differences in primer efficiencies (31).

Both the Livak *et al*. and Pfaffl methods are based on mathematical principals that only accept a single reference gene. Some experiments use multiple reference genes and should take this into account during data analysis.

Vandesompele *et al.* published an alternative mathematical approach which accepts multiple reference genes (32). As the Pfaffl method, this alternative method also takes in consideration possible differences in primer efficiencies.

Post Hoc tests

The ANOVA-analysis is an omnibus test, which only tests for an overall difference between groups. If there are significant differences between group means, the test does not indicate which means differ. This is a drawback of the ANOVA test if the experimental design wishes to explore which group(s) differ. To test which group(s) differ one can utilise post-hoc tests. These *after* ANOVA tests are only necessary if the ANOVA test was significant. Post-hoc tests like pairwise t-tests, Bonferroni, Tukey HSD and others can be performed with SPSS software (IBM SPSS Statistics) (37).

Correlation analysis

There are some downsides to a correlation analysis. The correlation coefficient is neither robust for asymmetric distribution nor resistant to outliers. A scatterplot may give some indications of skewness and outliers in a dataset (33, 38). A statistical significance test like *t*-test for linear correlation can be performed with small *r* (33).

This experiment used the correlation coefficient function in Microsoft® Excel for calculating *r*. This function generates the population correlation coefficient and not an estimation of the population correlation for samples. With a sample size of n=44 it was concluded not to correct this calculation, since the large sample size gives a good indication of the population correlation coefficient without degrees of freedom. Microsoft® Excel does not have a sample correlation coefficient calculation function. None of the above proposed tools aiding interpretation of the linear correlation was used in this experiment.

4.3 Further research

RT-qPCR is an extremely powerful technique with high levels of sensitivity and specificity. This makes it extra vulnerable to significant variance within datasets. With an approximately two-fold amplification of target sequences, small differences may make a great impact. This is crucial in fold expression change studies, which typically analyse differences in fold change between a control group and a treated group. This can lead to erroneous assumptions and conclusions. A hypothetical study comparing a control and treated group is dependent of a small intra-group variance to state that a treatment has a true effect or not, and not a significant variance that implies a false reality.

A significant fold change does not necessarily result in a true difference in protein expression. Correlation between gene and protein expression may be insufficient (39). This is due to regulation in post-transcription, translation and protein degradation. The protein abundance is regulated according to the current cell state (40). It is demonstrated that under dynamic transitions between cell states, such as during stress response, the mRNA concentration does not reflect the protein concentration (41). It is therefore advised to do proteomic analysis studies for better understanding of actual protein expression.

Errors will occur in every experiment which may lead to significant variation within the dataset. A well-designed experiment will try to minimise systematic errors, but random errors are difficult (if not impossible) to exclude from any experiment. Further research regarding stress related experiments on Atlantic salmon should investigate possible optimisation of sampling. If there are physiological and/or logistical factors limiting further optimisation, a standardisation of the sampling procedure is a minimum requirement.

The initial number of samples were 16 samples per group for this experiment. An early evaluation of the sample quantity concluded to reduce the sample quantity per major group to 12 samples due to suspicion of significant variation between the first and last sample. Even this reduction of sample quantity was not enough to avoid significant variance between samples in the same major group. Optimisation and standardisation of sampling with emphasis of reducing total sampling time for all groups will open for larger sample quantity, without risking significant variance.

Further experiments should also verify RNA and cDNA quality as well as quantity. Contaminants such as proteins and salts may inhibit downstream applications. Assessing RNA purity and concentration is therefore an important step in quality control. The Nanodrop is a system that offers purity and concentration assessments (42). Assessing RNA quality should also be done as a quality control step. RNA integrity has a significant impact on RT-qPCR performance (43). The Bioanalyzer 2100 is a recommended system for assessing RNA quality (42). An alternative method is to assess RNA quality with agarose gel electrophoresis (44). Another advised quality assurance step is validation of the amplicon from the target sequence by endpoint PCR followed by agarose gel electrophoresis.

This study is a novel and small piece in the big fish welfare puzzle puddle. Further research may compare common biochemical biomarkers of stress to stress related genetic markers.

5 Conclusions

Analysis of variance have been demonstrated to detect significant inter-group difference in relative gene expression RT-qPCR assays of Atlantic salmon. The data analysis detected statistically significant variance in groups B and D, but not in group AA. It is proposed to be a result of poor standardisation of sampling, with emphasis on total sampling time. The purpose of this study is to shed light on factors which may lead to profound variance in an *in vivo* relative gene expression analysis. Further research is recommended to optimise sampling of gill tissue and utilise proposed quality control steps for valid results.

6 Appendix

1. Total calculations
2. Protocols
3. RNeasy Mini Kit Handbook (Qiagen)
4. RNase-free DNase Set (Omega bio-tek)
5. TissueLyser Handbook (Qiagen)
6. qScript cDNA Synthesis Kit Quick guide (Quantabio)
7. PerfeCTa SYBR Green Supermix Low ROX (Quantabio)
8. Qubit RNA HS Assay Usermanual (Invitrogen)
9. Qubit 1X dsDNA HS Assay Kit UG (Invitrogen)
10. RNAlater handbook (Qiagen)
11. qPCR assay plate setup

7 Sources

1. Lønnum Andreassen B. Oppdrett: - Fiskevelferd bør bli et kriterium for å få nye konsesjoner forskning.no: forskning.no; 2019 [updated 29.08.2019] [Available from: https://forskning.no/fisk-fiskehelse-hav-og-fiske/oppdrett--fiskevelferd-bor-bli-et-kriterium-for-a-fa-nye-konsesjoner/1550310]

2. Aquaculture G. Aquaculture in Norway aquaculture.ggn.org: Certified aquaculture - GGN; [Available from: https://aquaculture.ggn.org/en/aquaculture-in-norway.html]

3. Overton K, Dempster T, Oppedal F, Kristiansen T, Gismervik S, Stien L. Salmon lice treatments and salmon mortality in Norwegian aquaculture: a review. Reviews in Aquaculture. 2018.

4. Van Weerd JH, Komen J. The effects of chronic stress on growth in fish: a critical appraisal. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology. 1998;120(1):107-12.

5. Hjeltnes B, Bang-Jensen B, Bornø G, Haukaas A. The Health Situation in Norwegian Aquaculture 20182019 [Available from: https://www.vetinst.no/rapporter-og-publikasjoner/rapporter/2019/fish-health-report-2018]

6. Landbruks- og matdepartementet. Lov om dyrevelferd lovdata.no: LD; 2009 [updated 20.07.2018] [Available from: https://lovdata.no/dokument/NL/lov/2009-06-19-97]

7. Gabriel U, Akinrotimi O. Management of stress in fish for sustainable aquaculture development. Researcher. 2011;3:28-38.

8. Hoem KS, Tveten A-K. Current approaches in decoding the molecular mechanisms of long-term stress in adult farmed Atlantic salmon (Salmo salar). Reviews in Aquaculture.

9. Cao Y, Tveten AK, Stene A. Establishment of a non-invasive method for stress evaluation in farmed salmon based on direct fecal corticoid metabolites measurement. Fish Shellfish Immunol. 2017;66:317-24.

10. Lunden Nilsen K. Forsker på bedre fiskevelferd med sensorfisk sintef.no: Gemini.no; 2019 [updated 10.01.2019] [Available from: https://www.sintef.no/siste-nytt/forsker-pa-bedre-fiskevelferd-med-sensorfisk/]

11. Erikson UG, Solvang T, Schei M, Strand A, Aalberg K. Utredning av fiskevelferd ved bruk av HydroLicer sintef.no: Sintef; 2018 [Available from: https://www.sintef.no/prosjekter/utredning-av-fiskevelferd-ved-bruk-av-hydrolicer/]

12. Feder ME, Walser J-C. The biological limitations of transcriptomics in elucidating stress and stress responses. Journal of Evolutionary Biology. 2005;18(4):901-10.

13. Navarro E, Serrano-Heras G, Castaño MJ, Solera J. Real-time PCR detection chemistry. Clinica Chimica Acta. 2015;439:231-50.

14. Buh Gašparič M, Tengs T, La Paz JL, Holst-Jensen A, Pla M, Esteve T, et al. Comparison of nine different real-time PCR chemistries for qualitative and quantitative applications in GMO detection. Analytical and Bioanalytical Chemistry. 2010;396(6):2023-9.

15. Marcos-López M, Calduch-Giner JA, Mirimin L, MacCarthy E, Rodger HD, O'Connor I, et al. Gene expression analysis of Atlantic salmon gills reveals mucin 5 and interleukin 4/13 as key molecules during amoebic gill disease. Scientific Reports. 2018;8(1):13689.

16. Mayer MP, Bukau B. Hsp70 chaperones: cellular functions and molecular mechanism. Cell Mol Life Sci. 2005;62(6):670-84.

17. Woods DB, Vousden KH. Regulation of p53 Function. Experimental Cell Research. 2001;264(1):56-66.

18. Kelman Z. PCNA: structure, functions and interactions. Oncogene. 1997;14(6):629-40.

19. Merriam-Webster. Xenobiotic merriam-webster.com: Merriam-Webster dictionary 2020 [Available from: https://www.merriam-webster.com/dictionary/xenobiotic]

20. Holmström KM, Kostov RV, Dinkova-Kostova AT. The multifaceted role of Nrf2 in mitochondrial function. Curr Opin Toxicol. 2016;1:80-91.

21. Ma Q. Role of nrf2 in oxidative stress and toxicity. Annu Rev Pharmacol Toxicol. 2013;53:401-26.

22. NCBI. ALOX5 arachidonate 5-lipoxygenase [Homo sapiens (human)] ncbi.gov: NCBI; 2020 [updated 26.04.2020] [Available from: https://www.ncbi.nlm.nih.gov/gene/240#gene-expression]

23. Schentrup AM, Lima JJ, Johnson JA. Very Important Pharmacogene: ALOX5 pharmgkb.org: PGKB; 2006 [Available from: https://www.pharmgkb.org/vip/PA166170368/overview]

24. Idriss HT, Naismith JH. TNF alpha and the TNF receptor superfamily: structure-function relationship(s). Microsc Res Tech. 2000;50(3):184-95.

25. Parameswaran N, Patial S. Tumor necrosis factor-α signaling in macrophages. Crit Rev Eukaryot Gene Expr. 2010;20(2):87-103.

26. Andersen GR, Nissen P, Nyborg J. Elongation factors in protein biosynthesis. Trends in Biochemical Sciences. 2003;28(8):434-41.

27. Sasikumar AN, Perez WB, Kinzy TG. The many roles of the eukaryotic elongation factor 1 complex. Wiley Interdiscip Rev RNA. 2012;3(4):543-55.

28. Jorgensen SM, Kleveland EJ, Grimholt U, Gjoen T. Validation of Reference Genes for Real-Time Polymerase Chain Reaction Studies in Atlantic Salmon. Marine Biotechnology. 2006;8(4):398-408.

29. Olsvik PA, Lie KK, Jordal A-EO, Nilsen TO, Hordvik I. Evaluation of potential reference genes in real-time RT-PCR studies of Atlantic salmon. BMC Mol Biol. 2005;6:21-.

30. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402-8.

31. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 2001;29(9):e45.

32. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology. 2002;3(7):research0034.1.

33. Helbæk M. Statistikk for kjemikere: Fagbokforlaget; 2016. 296 p.

34. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 2012;13:134.

35. van der Spoel E, Choi J, Roelfsema F, Cessie Sl, van Heemst D, Dekkers OM. Comparing Methods for Measurement Error Detection in Serial 24-h Hormonal Data. Journal of Biological Rhythms. 2019;34(4):347-63.

36. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biology. 2007;8(2):R19.

37. Field A. Discovering Statistics Using SPSS: SAGE Publications; 2009.

38. Ang A, Chen J. Asymmetric correlations of equity portfolios. Journal of Financial Economics. 2002;63(3):443-94.

39. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol. 1999;19(3):1720-30.

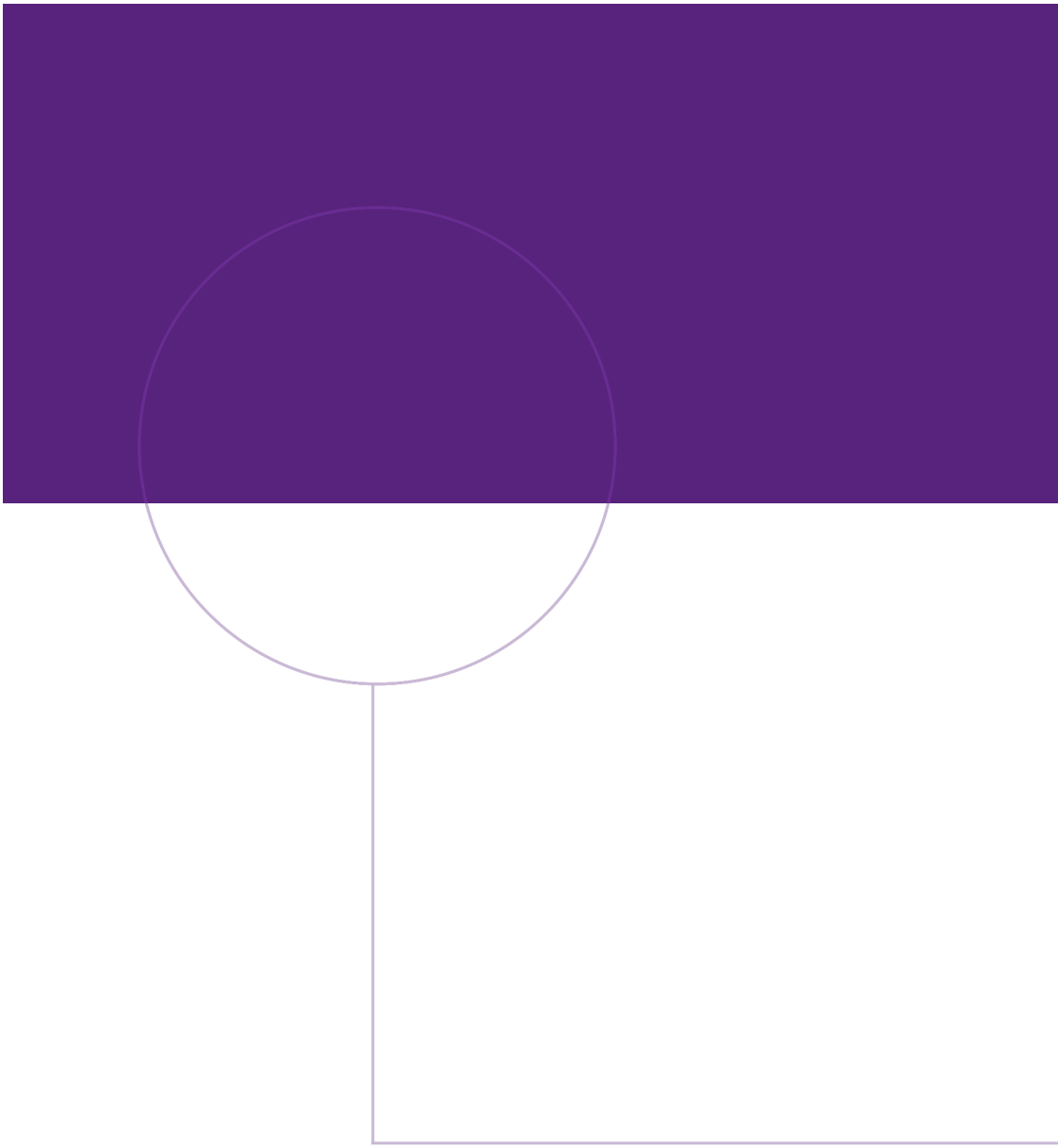
40. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet. 2012;13(4):227-32.

41. Liu Y, Beyer A, Aebersold R. On the Dependency of Cellular Protein Levels on mRNA Abundance. Cell. 2016;165(3):535-50.

42. Fleige S, Pfaffl MW. RNA integrity and the effect on the real-time qRT-PCR performance. Molecular Aspects of Medicine. 2006;27(2):126-39.

43. Fleige S, Walf V, Huch S, Prgomet C, Sehm J, Pfaffl MW. Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative real-time RT-PCR. Biotechnology Letters. 2006;28(19):1601-13.

44. Aranda PS, LaJoie DM, Jorcyk CL. Bleach gel: a simple agarose gel for analyzing RNA quality. Electrophoresis. 2012;33(2):366-9.



NTNU

Norwegian University of
Science and Technology