

Norwegian University
of Life Sciences

Master's Thesis 2020
Faculty of Biosciences

60 ECTS

CRISPR based functional characterization of the *abcg2b* gene contribution to muscle pigmentation in Atlantic salmon

Johanna Henny Wagnerberger
M.Sc. Biotechnology and Chemistry (Molecular Biology)

Preface

This master thesis was developed under the ongoing GENEinnovate project, a research project funded by BIONÆR (Research Council of Norway). The project aims to investigate genetic variants with significant impact on production traits in plants, fish and animals by utilizing and improving gene-editing tools within these species. GENEinnovate is a collaboration between the Norwegian University of Life Sciences (NMBU) and several industry partners, of which the aquaculture breeding company AquaGen was most essential for this thesis. The laboratory work was executed at NMBU between February and September 2020, with a 3 month interlude due to restrictions during the Covid-19 pandemic.

I would first and foremost like to thank my main supervisor Simen Rød Sandve (associate professor, NMBU) and assisting supervisor Jacob Seilø Torgersen (Senior scientist, AquaGen) for their guidance, which reached far beyond what was mandated. You have made me feel welcome and trusted me to be independent, but still always been available if I needed help. I would also like to thank the rest of the Sandve-lab group for including me from the beginning. A special thanks to Yang Jin (Postdoctoral fellow, NMBU) for helping me create a nice figure and supplying me with extra background material and to Krasimir Slanchev (Staff Scientist, Max Planck Institute of Neurobiology) for including me in the CRISPR genome editing process prior to the beginning of my thesis.

I would also like to thank my family, my partner and my friends who have shown me unwavering support and great patience over the last months. You have made writing a master thesis during an isolating pandemic much more bearable, and I am grateful to all of you.

Johanna H. Wagnerberger

Sammendrag – norsk

Den røde kjøttfargen til laks er et viktig varemerke og ofte forbundet med kvaliteten av fileten. Fargeintensiteten styres i stor grad av kostholdssammensetninger og skyldes hovedsakelig det røde pigmentet astaxanthin, som absorberes fra fôret i midt-tarmen og transporteres via blodet til muskelen. Pigmentet er lipidløselig og tett knyttet til lipidtransport. Kjøttfarging er imidlertid også under genetisk kontroll og representerer en viktig faktor i lakseavlprogrammer. Som en del av det pågående GENEinnovate-prosjektet ble CRISPR-Cas9-formidlede knockouts av tre forskjellige gener i Atlantisk laks utført. Ett av disse genene, *abcg2b*, er hovedfokuset i denne oppgaven. Aktiviteten til *abcg2b* er kjent for å påvirke kjøttfargen i Atlantisk laks negativt. Den nøyaktige funksjonelle rollen *abcg2b* har i forbindelse med laksekjøttfarging er imidlertid ukjent. Ettersom *abcg2b* produserer en membrantransportør er forventningen at proteinet eksporterer astaxanthin fra enterocytter i midt-tarmen tilbake til tarmlumen. I denne oppgaven sammenligner vi lipidinnholdet i midt-tarm enterocytter fra *abcg2b* knockout-laks og villtype-laks ved hjelp av fluorescens-mikroskopi. Bildene viser en klar økning i lipidinnhold i *abcg2b* knockout tarmtotter sammenlignet med tarmtotter fra villtype-laks i samme alder. Gjennomsnittlig lipiddekning og normalisert antall lipid-dråper var mer enn dobbelt så høyt i knockout tarmtotter enn i villtypen. Dette styrker hypotesen om at *abcg2b* eksporterer lipider fra enterocytten tilbake til tarmlumen. Astaxanthin transporteres sannsynligvis sammen med lipider av *abcg2b*, noe som fører til lavere konsentrasjon av astaxanthin i blodet og en lysere kjøttfarging hos laks med høy *abcg2b*-aktivitet.

Denne oppgaven testet også getPCR-metoden for å estimere genredigerings-effektivitet (KO-effektivitet) i mosaikk F0 laks. Metoden bruker en modifisert versjon av qPCR for å beregne KO-effektiviteten relativt til en intern og ekstern kontroll. Åtte laks som var genredigert i enten *abcg2b* eller *bco1-like* ble testet ved hjelp av samme getPCR-reaksjon. Den resulterende KO-effektiviteten ble deretter sammenlignet med en KO-effektivitet estimert ved hjelp av "gull-standarden" dyp sekvensering. Sammenligningen viser at det er korrelasjon mellom det getPCR-baserte estimatet og det sekvenseringsbaserte estimatet, men også at getPCR er sårbar for PCR-hemmende forurensninger. getPCR-metoden fungerer best på individer med høy KO-effektivitet. Den kan brukes som en rask og kostnadseffektiv måte for å undersøke et stort antall fisk og velge ut individer med høyest KO-effektivitet for videre evaluering. Metoden er foreløpig ikke nøyaktig nok til å erstatte sekvensbasert KO-effektivitetsestimering. getPCR kan derimot være et mindre risikofyllt, raskt og kostnadseffektivt alternativ til KO-effektivitetsestimering via et visuelt markør-gen.

Summary - english

Red flesh coloration is an important trademark of salmon meat and commonly associated with its quality. The colour intensity is largely controlled by diet composition and mainly caused by the red pigment astaxanthin, which is absorbed from the feed in the mid-gut and transported via the blood to the muscles. The pigment is lipid soluble and tightly linked to lipid transport. However, flesh coloration is under genetic control as well, and represents an important trait in salmon breeding programs. As part of the ongoing GENEinnovate project, Atlantic salmon CRISPR-Cas9 mediated knockouts (KO) of three different genes were performed. One of these genes, *abcg2b*, is the main subject of this thesis. The activity of *abcg2b* is known to impact Atlantic salmon flesh coloration negatively. However, the exact functional role of *abcg2b* in connection to salmon flesh coloration is unknown. As *abcg2b* produces a membrane transporter, it is hypothesized to export astaxanthin from the enterocytes of mid-gut villi back to the intestinal lumen. In this thesis we compare the lipid content of mid-intestine villi from *abcg2b* knockout salmon and wild type salmon using fluorescent microscopy. The images show a clear increase in lipid content in *abcg2b* knockout salmon villi compared to wild type salmon villi of the same age. The average lipid coverage, as well as the normalized number of droplets, was more than twice as high in knockout salmon villi as in control salmon villi. This supports the hypothesis that *abcg2b* exports lipids from the enterocyte back to the intestinal lumen. Astaxanthin is likely co-transported with lipids by *abcg2b*, leading to lower astaxanthin blood levels and a paler flesh coloration in salmon with high *abcg2b* activity.

This thesis also tested the getPCR method to estimate editing (KO) efficiency in mosaic F0 salmon. The method uses a modified version of qPCR to calculate KO efficiency relatively to an internal and external control. Eight salmon carrying gene edits of either *abcg2b* or *bco1-like* were tested using the same getPCR reaction. The resulting KO efficiency was then compared to a KO efficiency estimated with the “gold standard” method of deep sequencing. The comparison shows that getPCR estimations correlate with the estimations produced by the sequencing-based method, but also that getPCR is vulnerable to PCR inhibiting contaminants. The getPCR method works best on high-knockout individuals and could be used as a fast and cost-efficient way to screen large numbers of fish and select the individuals with highest KO efficiency for further evaluation. It is currently not accurate enough to replace sequencing-based KO efficiency estimation. However, it could be a less invasive, fast and cost-effective alternative to visual marker-gene assisted KO efficiency estimation.

Table of contents

Introduction & literature review.....	1
Salmon Farming	1
The red muscle colour.....	2
The variations in salmon flesh coloration.....	4
Genetic variation associated with redness in salmon	5
Genome editing	7
Thesis aim	9
Materials and Methods.....	10
Gene editing using CRISPR Cas9.....	10
Genome editing test PCR – using qPCR for knockout estimation.....	11
Tissue sampling.....	14
Genomic DNA isolation	15
Lipid staining and microscopy.....	15
Converting microscope image to binary segmentation.....	16
Results.....	17
getPCR.....	17
Quantification of lipid droplets in mid-intestine salmon villi	19
Discussion.....	22
getPCR estimates high KO efficiencies well	22
<i>abcg2b</i> knockout salmon show a clear increase in villus lipid content	24
Conclusion.....	27
REFERENCES.....	28
Appendix	a-f

Introduction & literature review

Salmon Farming

Historical notes

The early days of Norwegian aquaculture can be traced back to the first hatching of brown trout in 1850, but large-scale seafood production did not gain momentum until 1970. Around that time the first cage for off-shore cultivation of fish was constructed, which gave Norway a huge opportunity for upscaling the production of saltwater fish (Venvik, 2005). Norway's coastline was perfect for these cages, as its fjords and islands created many calm ocean areas, and the gulf stream ensured a stable water temperature. In 1982 the United Nations Convention on the Law of the Sea established that countries would have an exclusive economic zone stretching 200 nautical miles from the coastline (UNCLOS, 1982). Norway's long coastline now secured large ocean areas for the Norwegian aquaculture to bloom.

Economics and importance

The seafood industry has been on steady increase since its beginnings, showing a significant growth in GDP contribution over the past 10 years (Richardsen, Myhre & Tyholt, 2019). According to a report from the Norwegian Seafood Council (2020) this industry has more than doubled its turnover since 2009, passing 100 billion Norwegian kroner in export value for the first time in 2019. Of all exported seafood in Norway, the Atlantic salmon (*Salmo salar*) towers above the rest. Norwegian salmon farms are the biggest producers of Atlantic salmon in the world. It is also the only salmon species produced commercially in Norway (Moe, 2017) and will hereby only be referred to as "salmon". About 95% of the salmon produced in Norway is exported. This salmon export generates 67.6% of the revenue of all Norwegian seafood export (Norwegian Seafood Council, 2020)(Figure 1). Beyond the sheer production volume, the value of the Norwegian salmon production is tightly linked to several quality properties of the salmon meat. This includes the fillet protein content, lipid content, fillet texture, odour, and importantly, the characteristic fillet colour (Espe, 2008). Consumers value colour a great deal and are indeed willing to pay higher prices for a darker red fillet colour (Alfnes, Guttormsen, Steine & Kolstad, 2006; Bjorn Bjerkeng, 2008).

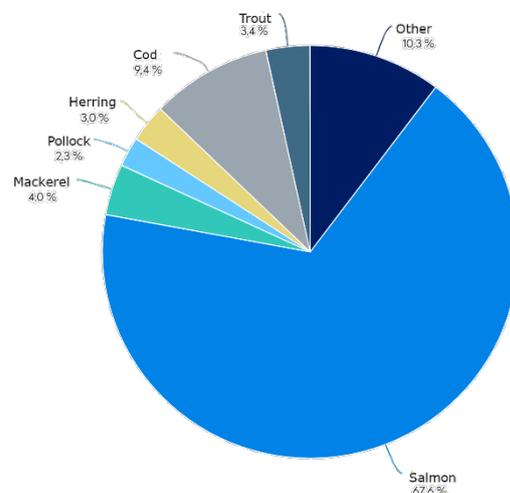


Figure 1: The overall 2019 Norwegian seafood export divided into species of fish exported. Image retrieved and translated with permission from the Norwegian Seafood Council, <https://nokkeltall.seafood.no/> (2020)

Beyond the sheer production volume, the value of the Norwegian salmon production is tightly linked to several quality properties of the salmon meat. This includes the fillet protein content, lipid content, fillet texture, odour, and importantly, the characteristic fillet colour (Espe, 2008). Consumers value colour a great deal and are indeed willing to pay higher prices for a darker red fillet colour (Alfnes, Guttormsen, Steine & Kolstad, 2006; Bjorn Bjerkeng, 2008).

The red muscle colour

Astaxanthin accumulates in salmonid muscles

The red muscle coloration arises when salmon feed on crustaceans that contain carotenoid pigments in their shell and flesh. These pigments cannot be synthesized by crustaceans, but have been obtained through the food chain, with the primary producers being marine phytoplankton and microalgae like *Haematococcus pluvialis* (Lorenz & Cysewski, 2000). The dietary pigments are metabolized and deposited in the flesh of the salmon (Schiedt, Leuenberger, Vecchi & Glinz, 1985), with the most important pigment for salmon muscle coloration being a group of carotenoids called xanthophylls. The presence of oxygen containing side-groups makes xanthophylls more polar than the very similar, pure hydrocarbon carotenes. In salmon, canthaxanthin and astaxanthin are the dominating xanthophylls associated with muscle pigmentation (Storebakken & No, 1992; Torrissen, 1989). This thesis will focus on astaxanthin (Figure 2) as this is most relevant for muscle pigmentation in farmed salmon.

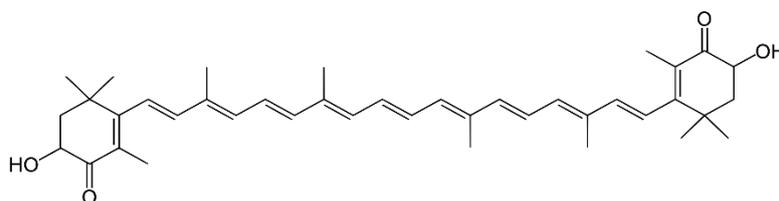


Figure 2: Non-stereochemical depiction of the astaxanthin molecule.
Adapted from "[Astaxanthin](#)" by Yikrazuul, licensed under public domain

Astaxanthin is despite its polar side-groups still lipid-soluble and sometimes found integrated in cell walls, orienting itself parallel to the phospholipids in the lipid bilayer. This positioning enables it to expose the more polar ends of the molecule to the hydrophilic outside of the bilayer, while keeping the hydrophobic middle buried (Kidd, 2011). The astaxanthin molecule has strong free radical scavenging and anti-oxidizing abilities, and is hypothesized to aid human health in numerous ways when consumed (Dose et al., 2016; Higuera-Ciapara, Felix-Valenzuela L Fau - Goycoolea & Goycoolea, 2006). In fact, it has been shown that astaxanthin has up to ten times stronger anti-oxidizing ability than other carotenoids like canthaxanthin and β -carotene, showing an activity similar to vitamin E (Miki, 1991). For salmonids, astaxanthin serves a function as provitamin A as well (Christiansen, Lie & Torrissen, 1994; Schiedt et al., 1985). This function is however not present in mammals (Bohn, 2008).

Astaxanthin Metabolism

In salmon astaxanthin is mainly absorbed through the mid intestine (White, Ørnsrud & Davies, 2003). The uptake from the intestinal lumen is hypothesized to be closely linked to fatty acid uptake. Its hydrophobicity makes the pigment soluble with fats, and it has been shown that salmon given a diet with higher fat content show an increased uptake of astaxanthin through the intestine, as well as increased retention in the muscle (B. Bjerkeng et al., 1997; Furr & Clark, 1997). This is partially because astaxanthin is transported through the hydrophilic lumen on the surface of mixed lipid micelles made of bile salts, phospholipids and fatty acids (Borel et al., 1996). Higher fat content aids the formation of such micelles by stimulating the release of emulsifying bile acids, thereby aiding astaxanthin to reach the surface of enterocytes in the intestinal wall (Zaripheh & Erdman, 2002). The transport into the enterocytes was previously thought to only happen through passive diffusion, but the exact mechanisms that facilitate this uptake are unknown (Parker, 1996). However, if the salmonid pathway for carotenoid uptake is similar to the mammalian pathway, some active transport both in and out of the enterocyte is expected to be present (Figure 3) (Emmanuelle Reboul, 2013). For example, cell wall transport of carotenoids in human enterocytes have been shown to be facilitated by cholesterol transporter proteins like NPC1L1 and ABCA1 (E. Reboul & Borel, 2011).

After uptake in the enterocyte, some of the astaxanthin will be converted to important metabolites like apocarotenoids, idoxanthin or retinol by enzymes like BCO1 and BCO2 (Lobo et al., 2012). Factors like temperature, stress and the concentration of retinol and long-chain polyunsaturated fatty acids play a role in the rate of astaxanthin metabolization (Trine Ytrestøyl et al., 2019). The remaining astaxanthin is first transported out of the enterocyte to the blood. High-density lipoproteins (HDL), chylomicrons and other serum proteins like serum albumin aid in further transportation through the bloodstream (Fielding & Fielding, 2008; G. H. Aas, Bjerkeng, Storebakken & Ruyter, 1999). The astaxanthin carried by lipoproteins is transported to the liver, where it is metabolized for other purposes and not used further for muscle pigmentation. However, some of the astaxanthin in the liver re-enters the bloodstream carried by very low-density lipoproteins (Page & Davies, 2003; Rajasingh, Øyehaug, Våge & Omholt, 2006). A lot of the astaxanthin carried by serum albumin travels directly to the muscle, circumventing a second metabolization step (Trine Ytrestøyl et al., 2019). Exactly how the carotenoid enters the muscle cell from the blood is unknown, but this mechanism is hypothesized to be a salmonid-specific pathway, since white-fleshed fish will remain white even if astaxanthin levels in the blood are high (T. Ytrestøyl & Bjerkeng, 2007). After entering the muscle, the storage of the pigment in the muscle cell is enabled by F-actin filaments, where one of the cyclic end groups of astaxanthin is proposed to bind to a hydrophobic pocket between the G-actin subunits (von der Ecken et al., 2015; Young et al., 2017).

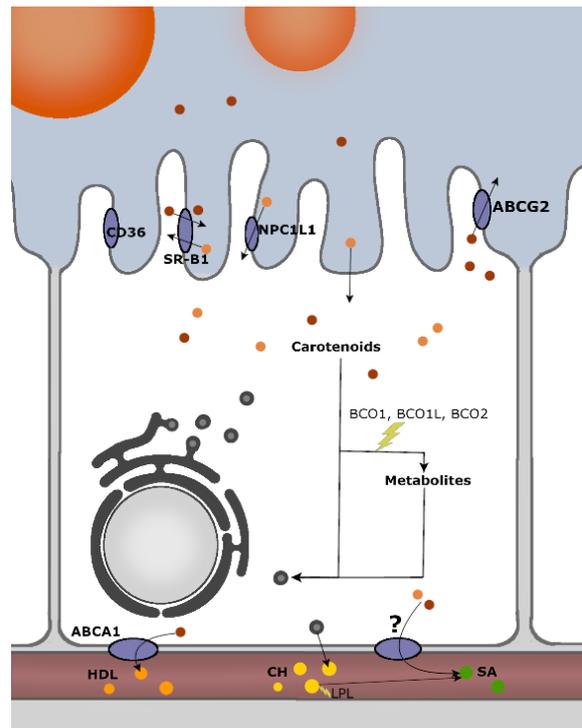


Figure 3: Proposed uptake and transportation of misc. carotenoids (orange) and astaxanthin (red) through the salmonid enterocyte. The depicted mechanism is based on knowledge from other vertebrates and the salmonid-specific mechanisms proposed by Rajasingh et al. (2006) and Zoric (2017). Enzymes are depicted with a yellow lightning bolt. The transportation in the blood is facilitated by high-density lipoproteins (HDL), chylomicrons (CH) and serum albumin (SA).

The importance of red muscle

The exact function of the red pigmented muscle tissue in many salmonid species is debated. Some salmon species, like Sockeye salmon, display high amounts of pigmentation displacement to the skin during spawning phase, suggesting a role for pigmentation in sexual selection. Males have also been shown to prefer more red coloured females (Foote, Brown & Hawryshyn, 2004). Higher astaxanthin levels have also been linked to better growth and survival of first-feeding fry (Christiansen & Torrissen, 1996). During this early developmental stage astaxanthin likely functions as a provitamin A (Christiansen et al., 1994). Other functions like the reduction of oxidative stress and strengthening of the immune system have been suggested as well (Christiansen et al., 1995; Skarstein & Folstad, 1996). Salmon undergoing a lot of stress have also been shown to have paler muscle coloration (Trine Ytrestøyl et al., 2019), which can indicate that the muscle cells serve as a temporary storage space for this strong antioxidant so that it may be used when needed. For wild salmon this is particularly important during stressful and strenuous migration periods when the salmon is not eating. For farmed salmon the most stressing situations would include crowded pens, frequent delousing or illness.

The variations in salmon flesh coloration

Diet and digestion – a bottleneck for flesh coloration

Farmed salmon has had less access to natural astaxanthin, since the commercial fish feed used today only contains low amounts of carotenoid-rich fish meal. In 1990 about 90% of the fish feed was marine based. A large increase in salmon production led to increasing supplementation of feeds with more sustainable plant-based material, resulting in feed being only 25% marine based by 2016 (Trine Ytrestøyl, Aas & Åsgård, 2015; T. S. Aas, Ytrestøyl & Åsgård, 2019). The feed does not contain enough astaxanthin naturally, which leads to a demand in artificial pigment supplementation to keep the farmed salmon from turning too pale. It is also easier to add an exact amount of pigment in each batch of feed, thereby securing a more consistent colour profile of farmed salmon meat.

However, despite adding an increasing amount of astaxanthin to the fish feed, flesh redness is declining. In 2009 a concentration of 30-40 mg of astaxanthin per kg of feed was enough to reach about 7 mg/kg in salmon muscle, while in 2019 50-70 mg/kg feed was needed for 6 mg/kg salmon (Trine Ytrestøyl et al., 2019). The big difference in astaxanthin concentration between feed and salmon muscle is largely because astaxanthin is not easily taken up by the intestine. A study conducted by T. Ytrestøyl and Bjerkgeng (2007) showed that muscle astaxanthin concentration increased linearly with weight specific doses of astaxanthin if the need for uptake through the gut was circumvented. The limitations thereby lie in the salmon's ability to digest and retain astaxanthin, which creates a bottleneck in its pathway to muscle deposition (B. Bjerkgeng & Berge, 2000). While this explains the high difference in astaxanthin concentration between feed and flesh, it does not explain why less astaxanthin has been stored in the muscle over the past 10 years. One explanation for this decline could be the xanthophyll's close connection to lipid uptake and transportation. When fish meal started to be replaced with plant-based alternatives this also led to lower amounts of phospholipids, cholesterol, vitamin A and omega-3 fats, and a higher amount of dietary fibre in the feed. All these variations can impact how much astaxanthin gets taken up from the feed and deposited in the muscle (Arnesen, Krogdahl & Sundby, 1995; Borel et al., 1996; Desmarchelier & Borel, 2017; Trine Ytrestøyl et al., 2019). Correct diet formulation thereby plays a significant role for achieving a redder salmon.

Genetics of muscle colour

A more intense red fillet colour can also be achieved through selective breeding. Salmon breeders select for parents with over-average muscle pigmentation when fed on the same feed and kept in the same environment as all other farmed salmon. This shifts the overall population phenotype towards a naturally redder flesh colour over time. Selective breeding for measurable traits like size, weight or parasite resistance has been done through family-based selection since 1975 (Gjedrem, 2010). This strategy uses trait quality information that is known about relatives (full siblings and half siblings mostly) to estimate an expected trait quality of the breeding candidate (Thodesen & Gjedrem, 2006). Family-based selection has proven effective for salmon, since they can produce large families and display high variation for many traits (Gjedrem, 2010). More recently, family selection has been combined with new strategies based on genotyping (Correa et al., 2017; Sae-Lim, Kause, Lillehammer & Mulder, 2017). Genotype-based selective breeding became more attractive in the past years following the rise of high-throughput genotyping technology (Leng, Lübberstedt & Xu, 2017; Varshney, 2016). Using this new resource led to the formation of breeding indexes that combined genetic and familial quality information for individual fish (Elliott & Kube, 2009; Houston et al., 2014; Moen & Ødegård, 2014; Tsai et al., 2017). This new approach is more accurate, requires less broodfish to be sacrificed for phenotypical analysis and reduces inbreeding since well-performing fish in poorly performing families now can be selected as well.

Improving breeding indexes through scientific methods

The genotype-based approach of broodfish selection requires a known connection between a genotype and the phenotype one selects for. Initially these connections are investigated through genome-wide association studies (GWAS) (reviewed in Korte and Farlow (2013)). These studies group their specimens by similar trait quality and compare them genotypically both within and between groups. The aim is to identify single-nucleotide polymorphisms (SNPs) that carry the same allele within groups, but not across groups. Most times linkage disequilibrium between proximal SNPs leads to a larger genomic area seemingly being connected to the trait (The International HapMap Consortium, 2005). This area is described as a quantitative trait locus (QTL) (Flister et al., 2013). Knowledge within protein chemistry, gene regulation and further testing needs to be employed to identify the true causative SNP or SNPs within a QTL, as GWAS only identifies association and not causation. The more genotype – phenotype connections are unravelled in the salmon genome, the more informative and profitable the genotyping of possible broodfish will become. In this thesis, one gene that is hypothesized to be connected to salmon flesh pigmentation is investigated further.

Genetic variation associated with redness in salmon

Several genomic regions associated with variation in redness have been identified in the salmon genome. This trait shows signs of being controlled by few loci of large effect (Rajasingh, Gjuvsland, Våge & Omholt, 2008). One of these regions is located on chromosome 26 and contains two beta-carotene oxygenase genes, *bco1* and *bco1-like (bco1l)* (Baranski, Moen & Våge, 2010). The *bco1l* variant is a paralogue to *bco1* and is known to cleave various carotenoids into vitamin A. The catalytic activity of Bco1 in salmon is unknown, but a structural comparison of the Bco1 and Bco1like protein models suggests a similar function with differing substrate specificity. Bco1 production in the intestine of pale fleshed salmon is also twice as high as in red salmon (Zoric, Torgersen, et al., 2017). Bco1like on the other hand is expressed more in red salmon (Helgeland et al., 2019). This implies a feedback regulation between these genes, strengthening the hypothesis of similar function, but differing use of dietary substrates. It could suggest that *bco1* might have evolved to target astaxanthin as a substrate for vitamin A production. Further testing is needed to confirm this. The paralogues are both believed to be connected to astaxanthin turnover and through this impact flesh coloration negatively.

Another genomic region associated with redness is situated on chromosome 2 and contains what is currently annotated as an *abcg2-like* pseudogene in the NCBI salmon genome assembly (ICSASG_v2, LOC106595425). This pseudogene classification is highly questionable, as RNAseq data and western blotting supports active transcription and translation of this gene (Zoric, Moen, et al., 2017). Assuming that this *abcg2* variant is functional, it represents a very interesting candidate gene for variation in flesh redness.

Known functions of *abcg2*

The acronym *abcg2* stands for ATP-binding cassette sub-family G member 2. The proteins that belong to this family are all membrane bound transporters. They are known to export a diverse range of hydrophobic to amphipathic compounds from the cytoplasm of enterocytes and other tissues. These compounds vary greatly in their chemical attributes (Robey et al., 2009). In humans overexpression of this transporter protein has been linked to multidrug resistance for certain types of cancer and neurodegenerative diseases (Iorio et al., 2016). In salmon and other vertebrates it aids in the export of xenobiotics like heavy metals and a selection of insecticides (Ferreira, Costa & Reis-Henriques, 2014). The broad spectrum of molecules this transporter protein can export, as well as the connection to flesh coloration via GWAS, makes the *abcg2* variant found on chromosome 2 in salmon a possible candidate for astaxanthin transport. A comparison of red and pale salmon also reported a 2.5 fold higher abundance of the Abcg2 protein in the enterocytes of pale salmon (Zoric, Moen, et al., 2017). This may indicate that this Abcg2 transporter aids astaxanthin efflux from the enterocytes back into the intestinal lumen and thus contributes to a paler flesh coloration (Figure 3).

The Atlantic salmon *abcg2* gene family

In the current salmon genome assembly (published by Lien et al. (2016)) 8 copies of *abcg2* were found (Zoric, Moen, et al., 2017). In this thesis the naming system is based on which *abcg2* gene in zebrafish they are most similar to (personal comm., Yang Jin (CIGENE) (2020)) (Figure 4). Since zebrafish only contain 4 *abcg2* genes, a suffix has been added to the salmon gene copies. Of the *abcg2* genes in the salmon genome, only *abcg2a-c*, *abcg2b-a* and *abcg2d-c* encode for the full Abcg2 protein with all its subdomains. The other genes encode truncated versions.

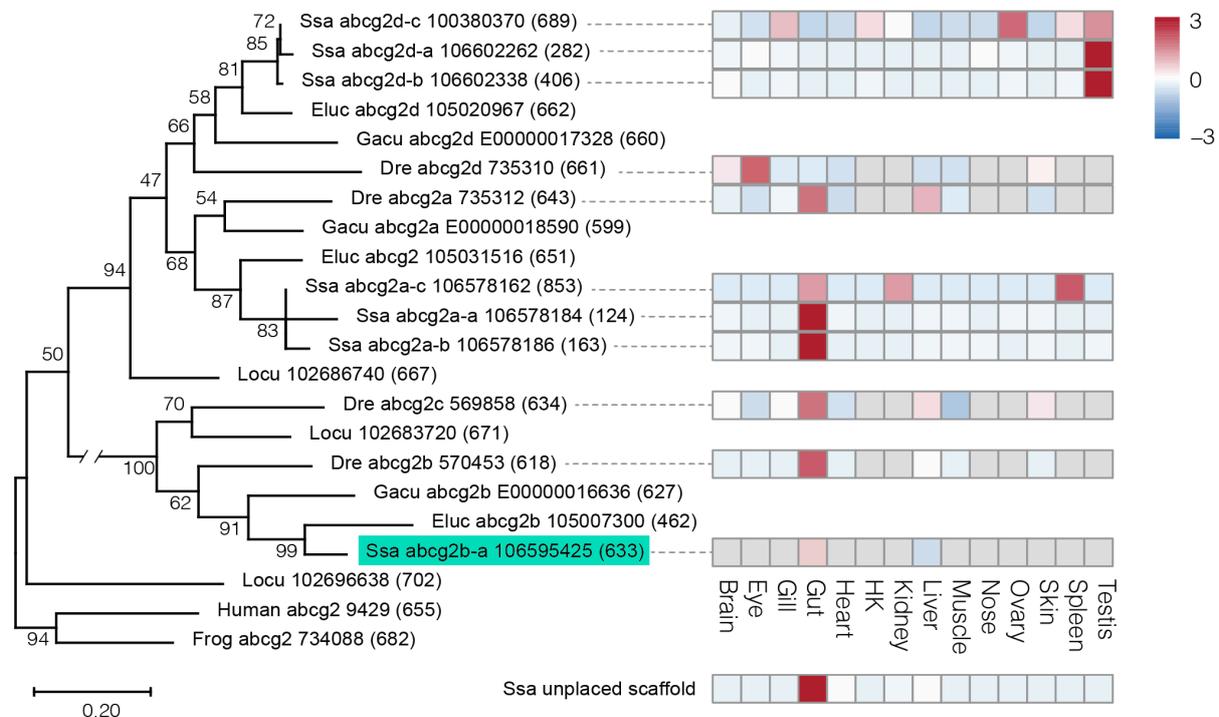


Figure 4: Phylogenetic relationships between the salmon (*Ssa*) *abcg2* genes and their homologs in other species. The species used for comparison are human, frog, spotted gar (*Locu*), three-spined stickleback (*Gacu*), northern pike (*Eluc*) and zebrafish (*Dre*). The gene name (if available) is given after the species abbreviation, followed by either their NCBI gene ID or, in cases where no NCBI ID is available, their ENSEMBL protein ID (#). The amino acid chain length is given in brackets at the end. Zebrafish and salmon genes are connected to their respective tissue expression profile. The colour scale refers to the number of standard deviations the observed tissue expression is distanced from the row-scaled mean expression across all tissues. Blue represents relative downregulation, while red represents relative upregulation. The gene containing the QTL associated to flesh coloration, *Ssa abcg2b-a*, is highlighted green.

The flesh colour associated QTL on chromosome 2 was identified by Zoric, Moen, et al. (2017) and studied further to identify single nucleotide polymorphisms (SNP) that could have causative effect on muscle pigmentation. One SNP that located within *abcg2b-a* was found “more strongly associated to fillet color than any other DNA polymorphism located on *ssa02*” (see p.17 of Zoric, Moen, et al. (2017)). This SNP causes a missense mutation, leading *Abcg2b-a* of pale fish to carry the amino acid Serine at the same location where red fish carry Asparagine. The two amino acids are highly similar, which means substituting one with the other is unlikely to cause a drastic change in protein behaviour. It is also not clear how this amino acid substitution could cause a 2.5-fold difference in protein abundance. Further investigation is therefore needed to confidently state this missense mutation as a mutation that causes change in flesh pigmentation. However, due to the high association to fillet colour, this SNP is currently the mutation within *abcg2b-a* that is most probable to cause change in flesh pigmentation.

Functional genomics to understand the Ax metabolism

Although we have convincing evidence that allelic variation in *abcg2b-a* (hereby only referred to as *abcg2b*) in salmon correlates with variation in flesh redness, we lack functional proof that the *abcg2b* gene is responsible for this phenotype. Until very recently, unravelling such causal genotype-phenotype links were next to impossible in salmon. However, recent advances in genome editing now allow for targeted manipulation of virtually any genome, including salmon. In the following section I will briefly describe the basic principles of genome editing, with particular regard to the CRISPR method.

Genome editing

Genome editing is the targeted altering of genes by altering their base code with biotechnological tools. Most methods used currently induce a double-stranded break (DSB) in the target region. The resulting space can then be used to alter the bases flanking the break, which is repaired again after the alteration has taken place. As of to date there is no widely used mechanism that both breaks, alters and repairs the DNA strand, although site-specific DNA recombinases have shown promising results (Bessen et al., 2019). Other technologies, like base editors, are being developed to circumvent the need for a DSB entirely, but these are still low in efficiency and need further optimization (Zhu, Li & Gao, 2020). The currently most widely used technologies create the targeted DSB, and then rely on the cell's natural repair mechanisms to do the rest. In eukaryotes a DSB is most likely to be repaired via non-homologous end joining (NHEJ)(Burma, Chen & Chen, 2006). This process is error-prone, and often leads to small indels (Heidenreich et al., 2003). If the DSB is in a gene, the indel will often lead to a frameshift. Every triplet downstream of the break is thereby altered, usually leading to a "knockout", which means the gene will not be able to produce a functional protein. If the indel alters whole triplets however, the reading frame stays intact and only a few amino acids are gained/lost in the protein. In such cases the altered gene might not be knocked out, even though the gene code has been edited. Knocking out a gene is the most usual goal in current gene editing, as this is the simplest to achieve. Sometimes however, it might be desirable to alter or add a new gene so new functions may occur. In such cases, the cell's ability to perform homology-directed repair (HDR) can be taken advantage of (Sander & Joung, 2014). HDR requires both a DSB and a template containing the desired alteration. The template should have sequence homologous to the sequence flanking the DSB on each end, and the desired alteration in the middle. The flanking bases are used to match the template to the region containing the DSB, and the middle is inserted into the break via DNA synthesis. Ligase then closes the single stranded breaks on each side of the new insert (San Filippo, Sung & Klein, 2008). This results in the insertion of a known sequence in the target region, rather than a random indel as in NHEJ.

The first successful attempts at gene modification were produced in the 1970s (Cohen & Chang, 1973; Jaenisch & Mintz, 1974), but the approach first gained momentum in the 1990s, when the use of zinc finger nucleases (ZFN) was discovered. ZFNs consist of a Fok I nuclease domain fused to zinc-finger proteins (ZFP) (Y. G. Kim, Cha & Chandrasegaran, 1996). Different zinc-fingers domains bind specifically to different base-pair triplets in the DNA (Figure 5). By choosing the right domains for the ZFP one could thereby direct where the nuclease would bind and induce a double-stranded break in the DNA (Urnov et al., 2010). However, the initial use of ZFN often turned cytotoxic, most likely because they created too many off-target cleavages (Porteus & Baltimore, 2003; Szczepek et al., 2007). The need for more specific targeting led to the discovery of transcription activator-like effector nucleases (TALEN), which used the same nuclease domain as ZFNs, but used a different targeting protein for sequence recognition (Miller et al., 2011). The way of inducing a double-stranded break remained the same as in ZFN, requiring dimerization of two Fok I nuclease domains to create a successful break (Figure 5). Interestingly, even though both methods create a DSB with the same nuclease dimer, TALENs produce mainly deletions, while ZFNs produce insertions and deletions at similar rates (Y. Kim, Kweon & Kim, 2013).

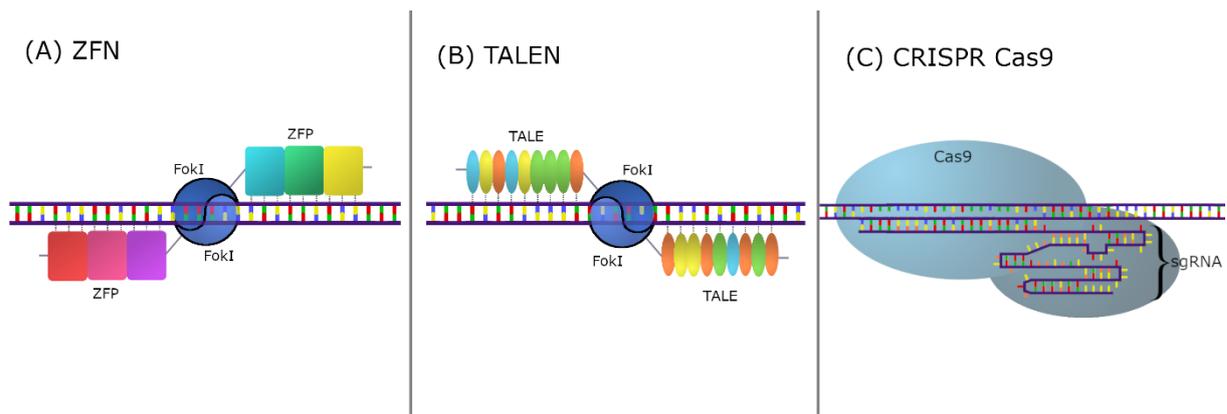


Figure 5: Gene modification tools bound to DNA.

(A) Zinc finger nuclease (ZFN) dimer consisting of zinc-finger proteins (ZFP) and a nuclease (FokI) dimer

(B) transcription activator-like effector nuclease (TALEN) dimer containing the TALE and a nuclease (FokI) dimer

(C) the Cas9 protein with an integrated single guide RNA (sgRNA)

CRISPR

A new form of creating a targeted DSB was found shortly after the discovery of TALENs, through the discovery of a bacterial immune system (Barrangou et al., 2007). The system uses clustered regularly interspaced palindromic repeats (CRISPR) in the DNA of the bacteria to integrate and store genomic sequence spacers from previously encountered viral infections. This genomic region works as a reference library to the cell when there is suspicion of a new viral infection. CRISPR associated proteins (Cas) aid the immune system by continuously scanning potential viral DNA sequences, and cutting sequences that match viral sequences in the CRISPR (Makarova et al., 2011). This prevents translation of previously encountered viral DNA, and thus halts infection. It was later found that a particular type of CRISPR immune system, the CRISPR-Cas9 system found in *Streptococcus pyogenes*, could be utilized for a cheaper, simpler and more precise form of genome editing (Deltcheva et al., 2011; Jinek et al., 2012). In *S. pyogenes* the Cas9 endonuclease consists naturally of a large protein bound to a duplex of a mature CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA). The crRNA is transcribed from the CRISPR region and contains a 20 nucleotide sequence from a previously encountered viral DNA on the 5' end (Gasiunas, Barrangou, Horvath & Siksnys, 2012). The region downstream of those 20 nucleotides hybridizes with the tracrRNA before they are co-processed to a mature guide-RNA (gRNA) and integrated in the Cas9 protein (Deltcheva et al., 2011). The fully assembled Cas9 protein then interrogates available DNA that could possibly be from viral sources. For the Cas9 protein to unzip a DNA strand, a protospacer adjacent motif (PAM) must be present in the respective DNA. In the case of Cas9 this PAM has the base sequence 5'-NGG-3'. If Cas9 encounters the PAM on one strand, the region upstream of it is unzipped, and the opposing strand is matched to the single-stranded part of the crRNA. If hybridization between the crRNA and the strand opposite of the PAM sequence occurs, a DSB is induced. The PAM is in this case vital for both recognition of non-self sequences and initiation of cleavage (Sternberg et al., 2014). When new viral infections are met, the region right upstream of a PAM is integrated into CRISPR, but the PAM is not. Since the CRISPR library lacks the PAM it is thus protected from cleavage by its own endonuclease.

The version of the Cas9 endonuclease used for gene modification is often altered from its naturally occurring state. Instead of the original crRNA:tracrRNA duplex it contains a chimeric single guide RNA (sgRNA) that combines the two into a single strand (Figure 5)(Jinek et al., 2012). The sgRNA can at times be shorter than the duplex, but sgRNAs that are not shorter are now used more widely, as they seem to have increased expression and cleaving efficiency *in vivo* (Hsu et al., 2013). Many more modifications exist, all with the intent to tailor the method to specific uses, increase stability and/or reduce off-target effects (Dang et al., 2015; Hendel et al., 2015).

A FO gene-edited organism, e.g. individuals that were edited at the single cell stage, most often become “mosaic”. This means that only a proportion of the somatic cells carries edited DNA. In this thesis, the fraction of gene-edited cells is referred to as knockout (KO) efficiency, as the gene-edit in question is supposed to create a gene knockout. It is however not guaranteed that a CRISPR induced DSB within the reading frame of a gene leads to a knockout, as an indel of 3 (or a multiple of 3) bases would not result in a frameshift. It is therefore important to keep in mind that the word “knockout” or “KO” in this thesis refers to a successful gene-editing event (with the intent of creating a knockout), and not a proven successful knockout of a functional protein.

Thesis aim

This master project is part of the ongoing GENEinnovate project, which makes Atlantic salmon CRISPR-Cas9 mediated knockouts of *abcg2b*, *bco1*, and *bco1-like* genes. There are two main aims of this master project:

First, a new method for estimation of KO efficiency using qPCR is tested as a cheaper alternative to sequencing-based KO efficiency estimation.

Second, a deeper analysis of *abcg2b* knockout phenotypes is performed to better understand its role as a membrane transporter in the enterocyte of the mid-gut. Since *abcg2b* produces an intestinal exporter of lipophilic compounds, in addition to carrying a SNP highly correlated to flesh colouration in Atlantic salmon, we hypothesise that *Abcg2b* plays a role in astaxanthin export from the enterocyte to the intestinal lumen. By analysing the CRISPR-Cas9 knockout salmon we expect to find that an *abcg2b* knocked-out enterocyte will have increased amount of lipids, and through this also a higher concentration of astaxanthin in the villus.

Materials and Methods

Gene editing using CRISPR Cas9

The CRISPR RNA (crRNA) sequences were designed to guide the Cas9 endonuclease to cut in the coding sequence of the target gene. Candidate sequences were found by choosing the sequence upstream of a PAM. Within *abcg2b* (LOC106595425) the sequence ‘AGACGGTTGGACCAGGCACC’ was chosen. A second crRNA sequence within the gene *slc45a2* (LOC106563596) was also designed (GGGGAACAGGCCGATAAGAC). A *slc45a2* knockout results in albinism and works as a visual marker for successful CRISPR injections.

The salmon in this thesis were edited via different types of Cas9 injection cocktails. One type contained mRNA that coded for the Cas9 protein, while another contained mature Cas9 protein. The products used for the preparation of these Cas9 injection cocktails are given in Table 1. The Cas9 mRNA was produced in-house using the mMMESSAGE mMACHINE™ SP6 Transcription Kit (Thermo Fisher Scientific, Waltham, USA) and a vector containing Cas9-2NLS coding sequence (provided by the Max-Planck Institute of Neurobiology in Planegg, Germany).

Table 1: Products used in the CRISPR injection cocktails. Only one of the products marked with ‘*’ was used in each cocktail.

Product	Stock concentration	Provider
Alt-R™ crRNA	100 μM	Integrated DNA Technologies (Coralville, USA)
Alt-R™ CRISPR-Cas9 tracrRNA	100 μM	
nuclease-free duplex buffer	30x	
Cas9-3NLS protein *	10 μg/μL	
Cas9-2NLS mRNA *	1 μg/μL	Produced in-house

The Cas9 injection cocktails were prepared by first adding crRNA and tracrRNA to a 1x nuclease-free duplex buffer (final concentration of 6μM for each RNA). This solution was heated up to 95°C for 5 minutes, before letting it cool to room temperature to allow formation of a mature gRNA. Second, three parts of the gRNA solution was then combined with one part 5x injection-buffer (30x nuclease-free duplex buffer + saturated phenol red solution) and one part of either the Cas9-2NLS mRNA or the Cas9-3NLS protein.

The use of different injection cocktails led to the formation of two different groups of salmon that both were gene-edited in the *abcg2b* gene. For one group we used the Cas9 mRNA, while for the other we used the mature Cas9 protein in the injection cocktail. In the protein-mediated gene-edit group we also targeted *slc45a2* to generate a visual knockout marker phenotype. In addition to the two *abcg2b* groups, a second gene, *bco1-like*, was edited in a separate group of salmon. This was done via a CRISPR cocktail containing the *bco1-like* gRNA and Cas9 mRNA. For the exact combination of gRNA and Cas9 delivery for each group, see Table 2.

Table 2: Setup of CRISPR injection cocktails. The components marked with a dark blue cell were present in the cocktail of each respective group.

Group name	gRNA			Type of Cas9 delivery	
	<i>abcg2b</i>	<i>slc45a2</i>	<i>bco1-like</i>	mRNA	protein
<i>abcg2b</i> _KO					
<i>abcg2b</i> / <i>slc45a2</i> _KO					
<i>bco1l</i> _KO					

Before the Cas9 cocktails were injected, salmon eggs (1 dL) were gently mixed with 200 μ L salmon milt at 4°C (Eggs and milt from AquaGen, Kyrkseterøra, Norway). 100mL filtered fresh water containing 0.5 mM glutathione (gH₂O) was added to initiate fertilization and prevent polysaccharide crosslinking in the chorion. Two minutes after adding the milt, the eggs were rinsed with gH₂O and left at 4°C for 3 hours to let embryogenesis initiate. The eggs were kept in gH₂O throughout the whole following genome-editing process, always at temperatures between 4-8°C and pH 8 to ensure optimal egg survival.

The Cas9 cocktail was injected with glass micro-needles. These needles were produced by heating and pulling apart glass capillaries with a PC-10 glass micropipette puller (Narishige, Tokyo, Japan). First, the pointed tip of the pulled glass capillary was inspected under the microscope (Stereo SZX10 by Olympus, Tokyo, Japan) for quality control. Next, the injection solution was inserted through the back of the glass capillary with a Microloader™ pipette tip (Eppendorf, Hamburg, Germany). The filled capillary was then inserted into an MN-153 Micromanipulator (Narishige) connected to a FemtoJet® Microinjector (Eppendorf). Lastly, the closed tip of the micropipette was carefully broken off with the tip of a tweezer to create a hollow open point. The needle was inserted into the germinal disc of a fertilized salmon egg, and enough solution to cover about 20% of the disc (1-3nL) was injected.

After injection, the eggs were transferred from gH₂O to fresh water and kept in a 6°C environment. Start-feeding of salmon fry was initiated 10 weeks after egg-injection. They were kept on a “Nutra XP” (Skretting, Stavanger, Norway) diet containing 4 mg astaxanthin per kg feed for the first 20 weeks.

Genome editing test PCR – using qPCR for knockout estimation

The proportion of gene-edited cells in mosaic organisms can vary substantially between individuals and have an impact on the organismal phenotype. When working with F0 mosaic knockout organisms it is therefore important to have methods to estimate the KO efficiency. In this thesis we optimized and tested the genome editing test PCR (getPCR) method by Li et al. (2019) as an alternative to a more elaborate and expensive ‘sequencing-based strategy’ to measure KO efficiency in Atlantic salmon.

getPCR is a modified qPCR method that utilises the high sensitivity between primers and their target site to estimate the fraction of target sites in the DNA that have been edited. For this, two primer pairs are designed. One primer pair (i.e. the “test” pair) is designed to amplify the wt sequence of the CRISPR target region, while the other “control” pair amplifies a separate un-edited control region within the genome (Figure 6A). The test primer pair is designed so that either one or both primers anneal to the expected Cas9 cut site, with an overhang of 2-4 bases matching the sequence downstream of the cut site. If the DNA has been successfully edited, the primer(s) covering the Cas9 cut site will in theory not anneal and thus not result in a PCR product. In mosaic gene edited organisms the control region will be amplified at optimal efficiency, as this region is always un-edited. However, for the test primers only some amplification will happen, with the amplification efficiency being directly linked to the editing efficiency in the CRISPR region. The differences in amplification efficiencies leads to a quantifiable difference in editing efficiency when compared to a qPCR performed with the same primers on a wt salmon (see Figure 6).

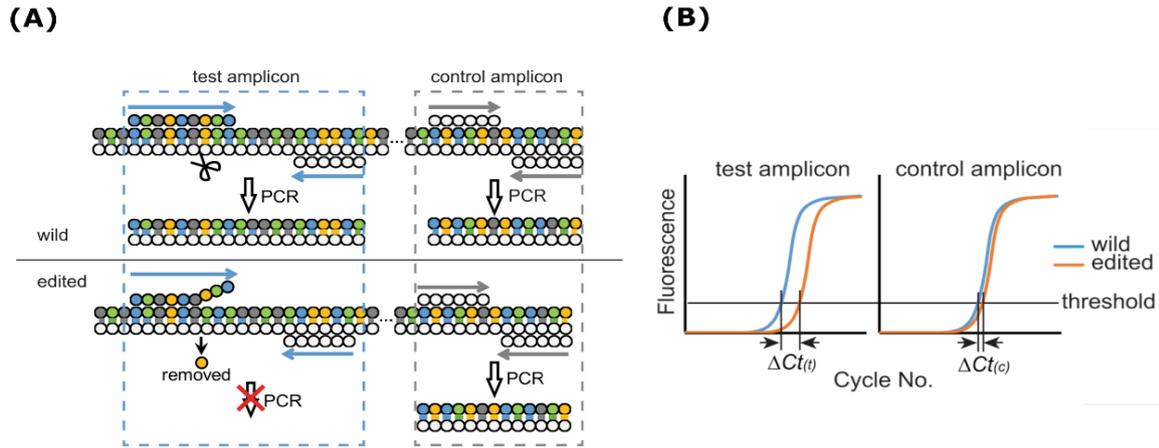


Figure 6: Derivative of "Principle and flowchart of getPCR"

(A) Difference in PCR reaction process between wild type and genome-edited cells. This figure shows only one primer covering the Cas9 cut site (implied by scissors).

(B) Amplification plots for both amplicons produced during a getPCR. Curves for a wild type fish (blue) and a mosaic gene-edited fish (orange) are plotted together for comparison.

Figure is taken from Li et al. (2019), licensed under [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/).

The knockout (KO) efficiency in the gene-edited salmon can then be calculated with the following equation:

$$KO \text{ efficiency} = 1 - \frac{(1 + E_t)^{-\Delta Ct_t}}{(1 + E_c)^{-\Delta Ct_c}} \quad \text{Eq. (1)}$$

Equation 1 is an adaptation of the $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen, 2001). Unlike the $2^{-\Delta\Delta Ct}$ method, it does not assume the same efficiency (E) for both primer pairs. ΔCt is also not first calculated between two different amplicons on the same sample, but between the same amplicon on two different samples ($Ct_{\text{edited}} - Ct_{\text{wild}}$). The values connected to the control primers are subscripted with "c", while the values connected to the test primers are subscripted with "t".

The getPCR method for KO efficiency estimation is not widely used and was therefore first tested relative to a more established sequencing-based KO efficiency estimation method. Two different gene knockouts were tested by using salmon from group *abcg2b_KO* and *bco1l_KO* (four salmon from each group). The PowerUp™ SYBR™ Green Master Mix (Thermo Fisher Scientific) along with its user guide was used as a guideline for getPCR protocol design. Essential modifications were made to fit the getPCR protocol described by Li et al. (2019). This resulted in three qPCR reaction mixes, all containing the components described in Table 3. Each reaction mix contained only one of the three primer pairs needed to test the getPCR method (control, *abcg2b*-test or *bco1-like*-test). The genomic DNA (gDNA) to be tested on the respective primers was added last, after the reaction mix had been divided into wells of 96-well qPCR optical plates (Bio-Rad, Hercules, USA). The final reaction volume in each well was 10 μ L. All reactions were run in triplicate on a qPCR cycling process as described in Table 4, using the CFX96 Touch™ Real-Time PCR detection system (Bio-Rad). The annealing temperature was set to 4°C above the optimal T_m of the test primer covering the Cas9 cut site, to ensure that this primer would only anneal to perfectly matching sequences within the gDNA. Each

reaction mix was also used to run triplicate non-template controls (NTC), which instead of gDNA contained more Milli-Q® purified water (dH₂O) (Merck, Darmstadt, Germany).

Table 3: Setup of getPCR reagent solution. Both master mix and primers ordered from Thermo Fisher Scientific. Stock concentrations are given in brackets after the reagent name.

Reagent	Volume used per 10µL reaction
PowerUp™ SYBR™ Green Master Mix (2x)	5µL
Forward primer (10µM)	0.5 µL
Reverse primer (10µM)	0.5 µL
gDNA (variable)	enough to contain minimum 5ng gDNA
dH ₂ O	fill to 10µL

Table 4: getPCR cycling mode. Also used for selection of primers and standard-curve production.

Step	Temperature (°C)	Duration
UDG activation	50	2 minutes
Dual-Lock™ DNA polymerase	95	2 minutes
Denaturation	95	15 seconds
Annealing	69	15 seconds
Extension	72	30 seconds
Melt curve – initial denaturation	95	15 seconds
Melt curve	60 → 95	Increasing 0.2°C every 15s

The same internal control primer pair, designed to amplify a region within LOC106577560, was used for both knockout groups. However, test primer pairs had to be designed and optimized for each of the two knockout-genes individually. Several combinations of forward and reverse test primers were first tested on wild type gDNA and non-template controls to ensure the absence of any additional amplicons or primer dimers. The final primer pairs that were chosen for getPCR are given in Table 5 (see Supplemental table 1 for a full list of all tested primers). The chosen primer pairs were first used in a qPCR on a dilution series of wild type gDNA (5 gDNA concentrations ranging from 0.5 to 100 ng/µL). This created a standard curve from which their amplification efficiency could be extrapolated (Supplemental figure 1). The dilution series qPCR used the same reagents and cycling mode as described in Table 3 and Table 4.

Table 5: Primer pairs that were used for the knockout salmon getPCR. Bases reaching past the Cas9 cut site in the respective gene are coloured red. LOC106577560 did not contain a Cas9 cut site as it served as the internal control amplicon. The primer melting temperature (T_m) and amplification efficiency of each pair is given as well.

Target gene	Direction	Primer sequence	T_m (°C)	Efficiency
<i>abcg2b</i>	Forward	CAGAGGTGATGTTCCAGGTGCC	65.1	94.6%
<i>abcg2b</i>	Reverse	GTGTAGTCTGGAGAAGGAGACGGTT	64.5	
<i>bco1-like</i>	Forward	TGATGCCATGTCTCAGTCCCTCATC	64.9	98.0%
<i>bco1-like</i>	Reverse	CTGCTTTGACAGGTTCCGGGTGA	64.6	
LOC106577560	Forward	TCAGCGTTCATTGGCTTTGAGACCA	65.1	103.7%
LOC106577560	Reverse	CAGGCTGTTCTTGATCTCGTACTGGTT	64.4	

The final getPCR experiment was run in triplicate for both control and target regions in each knockout salmon. A wild type salmon was run in triplicate for all three regions, acting as the “wild” external control (see Figure 6B). The PCR cycling process is described in Table 4.

The getPCR KO efficiency estimations were compared to KO efficiencies estimated from MiSeq sequencing data. The sequencing data was collected before the beginning of this thesis, so it will only be briefly discussed. First, DNA was isolated from embryos in group *abcg2b_KO* and *bco1l_KO*. The respective Cas9 target region was then amplified via PCR, and sequenced using MiSeq technology (Illumina Inc., San Diego, USA). KO efficiency was estimated by investigating the fraction of reads that mismatched to the salmon genome in bases close to the Cas9 cut site. The base showing the highest fraction of mismatches around the cut site was used to estimate the KO efficiency in the whole salmon (Figure 7).

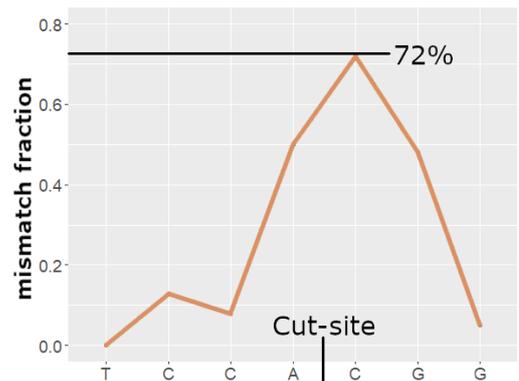


Figure 7: Hypothetical example of how ko-efficiency (in this case 72%) was estimated from MiSeq data.

The getPCR method was, after initial testing, used as the main method to estimate KO efficiency in the *abcg2b/slc45a2_KO* group. gDNA from intestine tissue of fourteen salmon in the *abcg2b/slc45a2_KO* group was used in a getPCR as previously described. Five of these salmon were also used in a second getPCR that tested the knockout-efficiency in diluted gDNA (5ng/μL) extracted from both intestine and liver tissue.

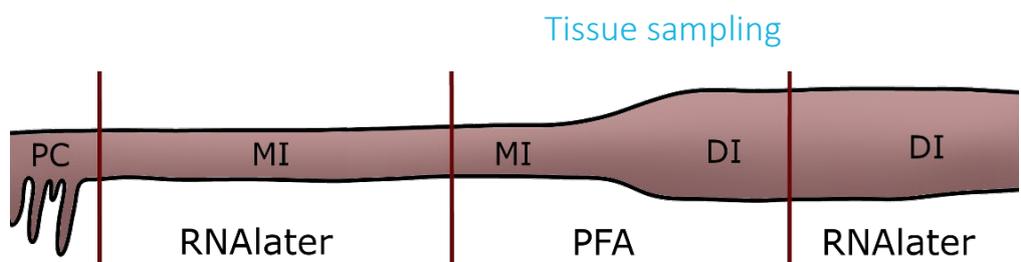


Figure 8: Division of the salmon gut (red lines indicate scalpel cut sites). The solution each piece was transferred to is written below the respective piece. Part of the Pyloric caeca (PC), the whole mid-intestine (MI) and whole distal intestine (DI) of a salmon are shown.

Intestine and liver tissues were sampled 4 months after start feeding (no fasting period). The fish were sedated in water containing Tricaine mesylate and killed with a sharp blow to the head. An incision was made from the gill, below the lateral line, to the anus. Organs were then removed from the body cavity, before tissue samples were collected. For the liver, two slices of approximately 10mm³ in size were dissected out and stored in RNAlater. Before the intestine tissue was sampled the intestine was emptied by carefully pushing out its contents. Next, the intestine was divided into three pieces with a scalpel (Figure 8). The tissues placed in RNAlater were stored on -20°C. The tissue piece containing both mid- and distal intestine was placed in 4% paraformaldehyde (PFA) and kept in the fridge overnight to allow tissue fixation to occur. The following day the intestine piece was washed with 1xTBS and slowly dehydrated with increasing ethanol concentrations (25%, 50%, and finally 70% ethanol for 30 minutes each). Subsequently, the intestine piece was stored on -20°C in fresh 70% ethanol.

Genomic DNA isolation

All gDNA was isolated by using the DNAdvance™ genomic DNA Isolation Kit (Beckman Coulter, Brea, USA) and following the recommended instructions. After isolation, the gDNA was tested for contamination and degradation via agarose gel electrophoresis. 8µL gDNA was mixed with 2µL 5x loading buffer (Thermo Fisher Scientific) and inserted into wells on a 1% Agrose gel containing RedSafe™ (Merck). The gel was placed in a Tris/Acetate/EDTA buffer bath for 45 minutes over 60 Volts and subsequently imaged under UV light using “Filter 1” and “gelRed” settings on the Gel Doc™ (Bio-Rad). The image was inspected to ensure only one clear band was visible. The gDNA was then stored at -20°C.

Lipid staining and microscopy

To investigate the lipid content of *abcg2b* knockout enterocytes, fixated mid intestine pieces of *abcg2b/slc45a2_KO* salmon were stained with florescent dyes, and compared to control fish intestine that had undergone the same procedure. This was done to visualize the histology of the nucleus, lipid droplets, and cholesterol in the cell under fluorescent microscopy.

PFA-fixated mid intestine on 70% ethanol was first rehydrated by slowly reducing the ethanol-concentration (30 minutes on 70%, 50% and 25% ethanol respectively), and finally placing the intestine piece in tris-buffered saline (TBS) for 30 minutes. After rehydration the intestine was cast in 3% agarose and cut into 300-400µm thick cross-sections using a Compressstome® VF-300 Microtome (Precisionary Instruments Inc., Natick, USA). The cross-sections were removed from the agarose cast and the cells were permeabilized by placing the cross-section in a solution of 0.5% saponin in 1x TBS (TBSS) for one hour. The TBSS was then replaced by the staining solution described in Table 6, in which the intestine slices remained submerged overnight. The three stains (all by Thermo Fisher Scientific) were used to visualize the nucleus, lipid droplets and cholesterol.

Table 6: Staining solution used for fluorescence microscopy.

Product	Excitation/Emission	Staining target	Working solution
SYTOX™ Green	504/523 nm	DNA (nucleus)	0.5 µM
HCS LipidTOX™ Red neutral	577/609 nm	Lipids	150x stock dilute
Cholera Toxin Subunit B, Alexa Fluor™ 647 conjugate	650/668 nm	Cholesterol	2.5 µg/mL
TBSS (0.5% saponin)	--	--	Fill to achieve desired stain concentrations

The slices were washed in TBSS for 2x20min the following day, before being placed in a petri-dish filled with TBS for imaging under the Axio Imager.Z2 Microscope (Carl Zeiss AB, Oberkochen, Germany). The water-submersible W “Plan-Apochromat” 40x objective (Carl Zeiss AB) was used in combination with the laser specifications given in Table 7 to take images of stained villi in the cross-sections. For every intestine, three slices were stained and visually inspected, before capturing a picture of minimum three different villi in each slice. The image was taken in the centre of each villus by making sure that cellular orientation along the lacteal was visible. This resulted in at least 9 pictures of different villi for each salmon.

Table 7: laser specifications used for imaging of mid intestine cross-sections.

Target stain	Laser wavelength	Laser strength	Detector gain
SYTOX™ Green	488 nm	0.15 - 0.35%	650V
HCS LipidTOX™ Red neutral	561 nm	0.15%	715V
Cholera Toxin Subunit B, Alexa Fluor™ 647 conjugate	640 nm	0.20%	653V

Converting microscope image to binary segmentation

To quantify differences in intestinal lipid content between knockout and wild type fish we combined fluorescence microscopy with binary segmentation of the captured images. The microscope images contained three greyscale image channels, one for each staining target (see Table 6). The channel containing the lipid droplets was isolated and edited using the image analysis tool Fiji (Schindelin et al., 2012). Brightness and contrast were auto optimized, before processing the image using the “Subtract background” function (rolling ball radius = 80px, sliding paraboloid). The images were exported to HDF5-format and used for “Pixel classification” in the Ilastik image segmentation toolkit (Berg et al., 2019). The Ilastik pixel classifier was trained on 16 of the total 157 images. The trained algorithm was used to create binary images, segmenting the lipid droplets from the background. The region of interest (ROI) within the picture was set to only contain the villus depicted in the picture. This was achieved by opening the original greyscale image in Fiji, overexposing the image enough to create a clear villus outline, and tracing this outline by hand using the Intuos M pen tablet (Wacom, Kazo, Japan). This ROI was then transferred onto the binary image in Fiji. Within the ROI, the “Analyze Particles” tool was used to extract number and area of individual lipid droplets, as well as villus area, median droplet size, and fraction of villus covered by lipid droplets. All particles with an area smaller than 4 pixels ($\sim 0.037\mu\text{m}^2$) were excluded from the particle analysis to reduce background noise.

Results

getPCR

Establishing a getPCR protocol for detecting CRISPR mutations in salmon

Before evaluating getPCR as an alternative to the sequencing-based method of KO efficiency estimation, test primers were designed for the *abcg2b* and *bco1-like* gene knockouts. Initial qPCR tests with test primers that both overlapped the Cas9 cut site showed consistent artefact amplification in the control reactions without a DNA template. The double-overlap primer pairs were therefore not used to test getPCR, and new primer pairs with only one primer overlapping the target site were designed. One pair of the single-overlap primer pairs was selected as the test primer pair for each knockout gene (see Table 5 in method-section for full sequence).

Sequencing-based quantification of KO efficiency was done on 4 salmon from group *abcg2b_KO* and *bco1l_KO*, respectively. These fish were then used for getPCR, and the resulting KO efficiency was compared to the sequencing-based KO efficiency (Figure 9). Some samples of the *bco1l_KO* group did not amplify successfully in the first getPCR run (Figure 9B). A second getPCR was therefore run for both KO groups, where the gDNA of group *bco1l_KO* was diluted 1:5. All samples showed successful amplification in the second getPCR. Apart from one sample in the *bco1l_KO* group, the getPCR-based method estimated the KO efficiency to be higher than the sequencing-based method.

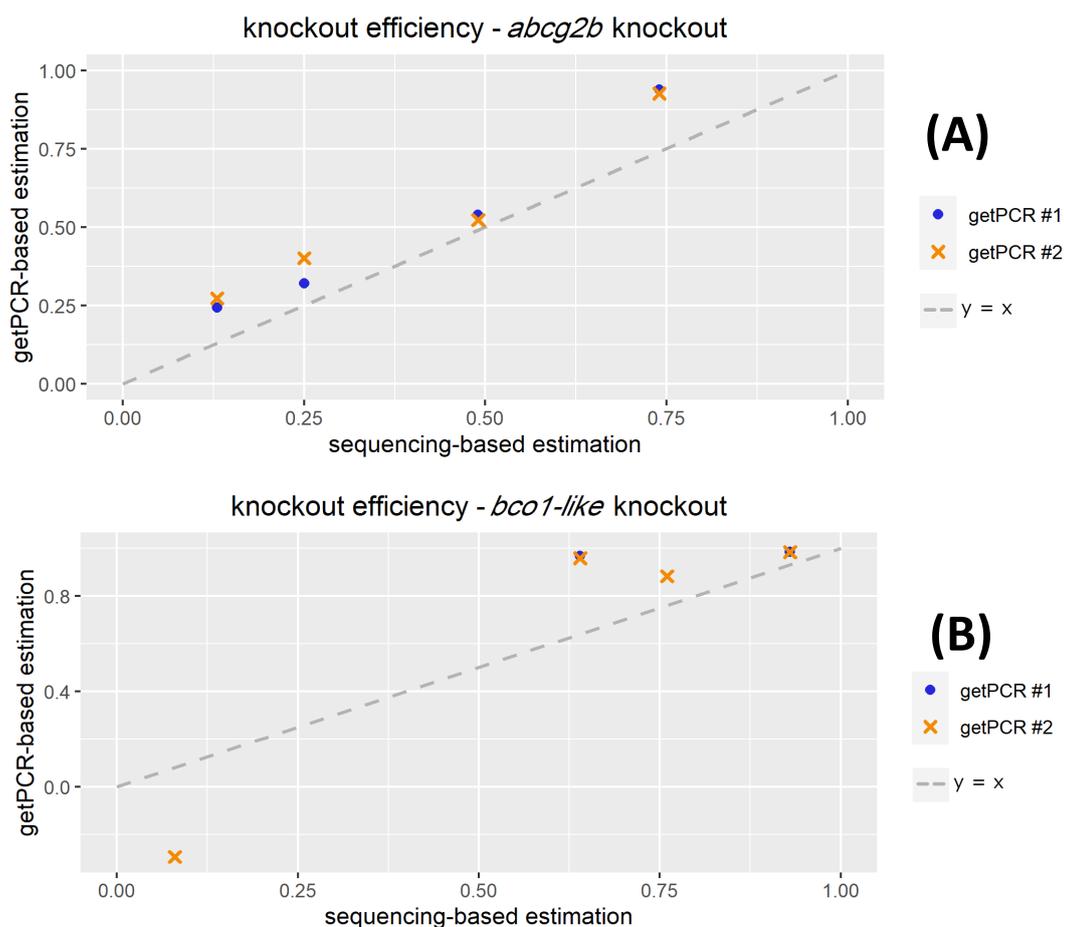


Figure 9: Comparison of sequencing-based and getPCR-based KO efficiency estimations.

Estimations were done on (A) group *abcg2b_KO* and (B) group *bco1l_KO*

The dashed grey line indicates a theoretical 1:1 correlation between the two methods. The first getPCR (getPCR #1) did not result in Ct-values for all samples.

Estimating KO efficiency in two tissues

A total of 14 salmon in the *abcg2b/slc45a2_KO* group were evaluated for *abcg2b* KO efficiency using getPCR. These fish displayed a varying degree of the albinism phenotype (rated as high, medium or low) (Supplemental figure 3). Samples of gDNA isolated from the mid intestine indicated a weak trend of individuals with high degree of albinism also having among the highest estimates of *abcg2b* KO efficiency (Figure 10A, exact values listed in Supplemental table 3). However, a significant difference in KO efficiency was found only in comparison between low- and high-level albino salmon (Wilcoxon test p-value = 0.0095). For five of the 14 fish, an additional independent getPCR on diluted gDNA from intestine and liver was performed (Figure 10B, exact values listed in Supplemental table 4). This was done to compare KO efficiency between tissues. The results showed great differences in KO efficiency between liver and intestine tissue, however no systematic difference in KO efficiency was present. The getPCR estimates on the intestine tissue did not vary much between the two independent getPCR experiments but did result in a slightly higher KO efficiency estimation on the more diluted gDNA in the second getPCR (Figure 10A vs B).

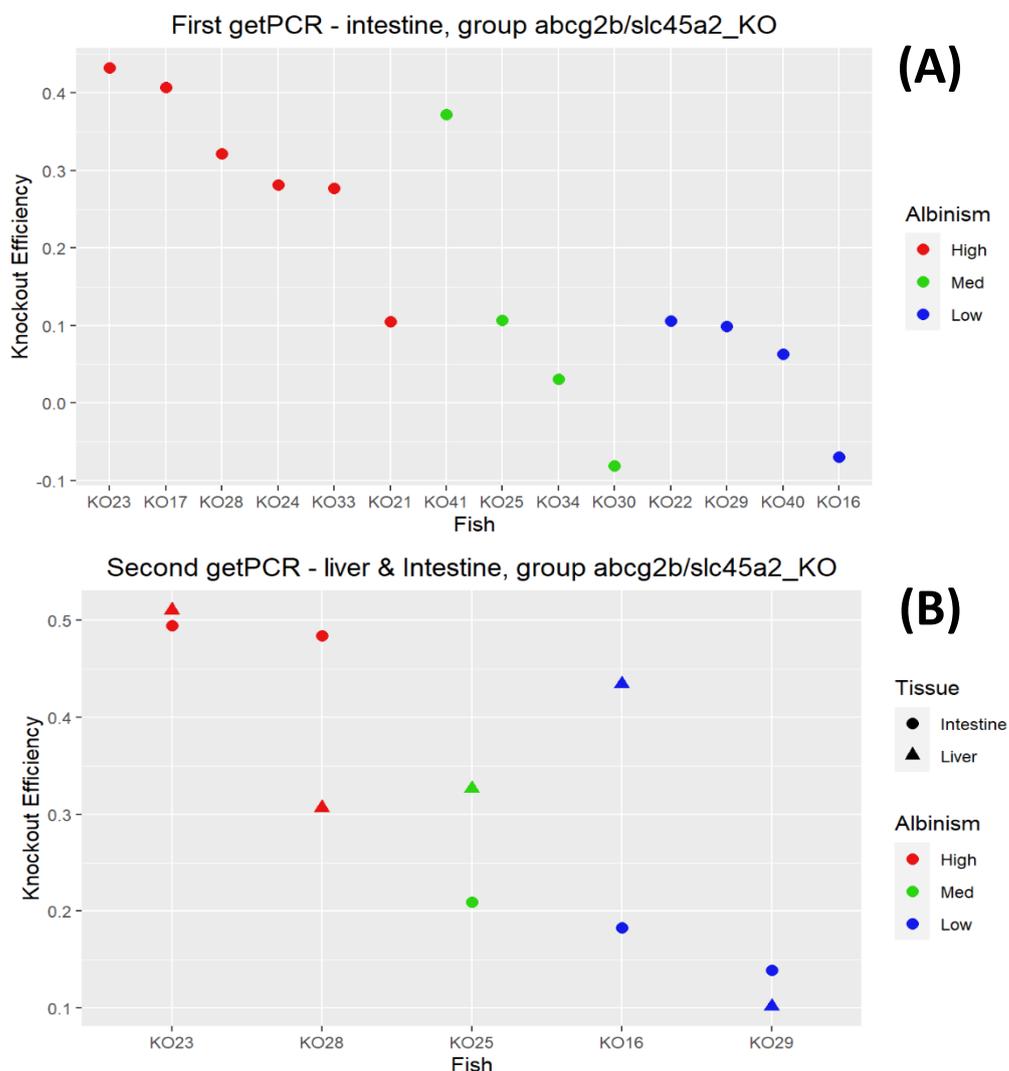


Figure 10: *abcg2b* KO efficiency estimated via getPCR on salmon from the *abcg2b/slc45a2_KO* group. Fish are tagged with ID (KO#) along the x-axis. The visible degree of albinism is divided into high (red), medium (green) or low (blue). **(A)** First getPCR, performed on all 14 salmon **(B)** second getPCR, performed on 5 of the 14 salmon. gDNA was isolated from liver (dot) or intestine (triangle) and diluted before running the getPCR.

Quantification of lipid droplets in mid-intestine salmon villi

Salmon from group *abcg2b/slc45a2_KO* were used to investigate the functional role of *abcg2b*. An image analysis of 300-400 μm thick slices of stained, PFA-fixed mid-intestine was carried out to visualize lipid histology and quantify differences between villi in knockout and wild type salmon. The images captured by the microscope each contained one villus with three superimposed channels, one for each stain (see example in **Figure 11A+C**). The cholesterol stain was included in the imaging process but not investigated further. Visual inspection of the microscope images of knockout villi revealed clearly visible, large lipid droplets clustered around the lacteal/centre of the villus (**Figure 11A** and Supplemental table 5). These few largest droplets were less common in control villi (**Figure 11C** and Supplemental table 6). Both groups however showed a trend of larger droplets towards the centre of the villus, and smaller droplets radiating outwards to the edges. A binary segmentation-image of the lipid-stain channel was produced (**Figure 11B+D**), and the region of interest was used for further analysis of the lipid-droplets in 8 control and 8 knockout salmon for which we had KO efficiency from getPCR (**Figure 10A**).

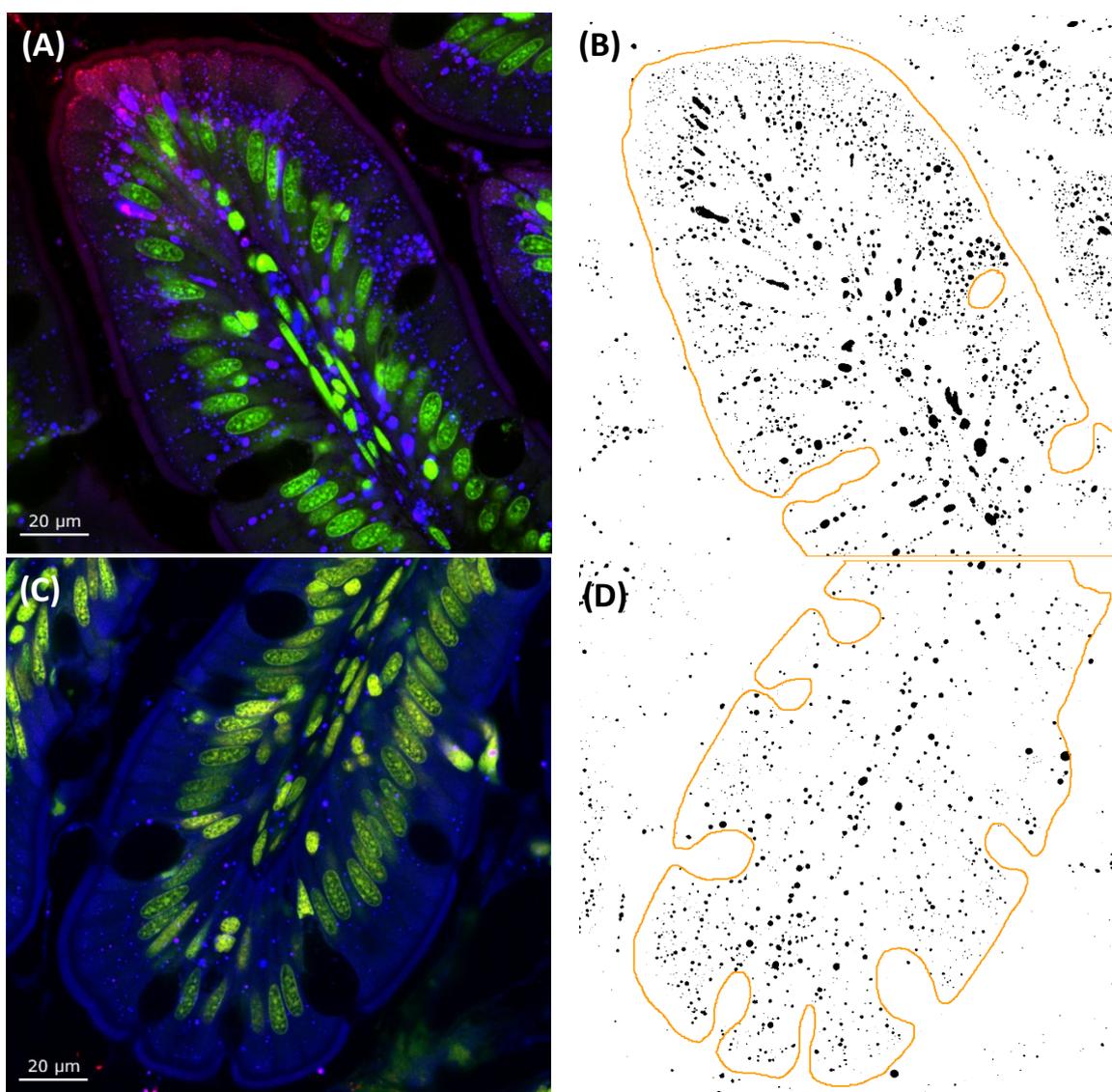


Figure 11: Representative images of a knockout and control salmon villus.

(A+C) Three-channel image produced by the microscope. The channels show staining of the nuclei (green), cholesterol (red) and lipids (blue) respectively. **(A)** Knockout villus **(C)** Control villus

(B+D) Binary segmentation image of the lipids (black) against the background (white). The region of interest is marked as an orange outline. **(B)** Knockout villus **(D)** Control villus

Firstly, the relative amount of lipids in each villus was compared. Figure 12A shows the fraction of the villus area covered by lipids for each individual fish. A clear difference can be seen between salmon from the control and knockout groups. Knockout salmon show higher median area coverage, as well as a wider range in values than control salmon. The area coverage of the *abcg2b* knockout group (mean of 7.01%) was significantly higher (Wilcoxon test p-value < $2.2e^{-16}$) than the control group (mean of 3.07%). No correlation between KO efficiency and lipid coverage was found within the knockout group (Kendall rank correlation, $\tau = -0.214$, p-value = 0.55).

Secondly, differences in lipid droplet number were investigated by comparing droplet number normalized to the area of the villus (i.e. droplets per area). Figure 12B shows the number of lipid droplets per μm^2 for each individual fish. As with the lipid proportion, the *abcg2b* knockout group also has significantly more droplets (Wilcoxon test p-value < $2.2e^{-16}$), with a mean value of 0.134 droplets per μm^2 compared to the mean 0.061 droplets per μm^2 in the control group. No correlation between KO efficiency and normalized droplet number was found within the knockout group (Kendall rank correlation, $\tau = -0.071$, p-value = 0.90).

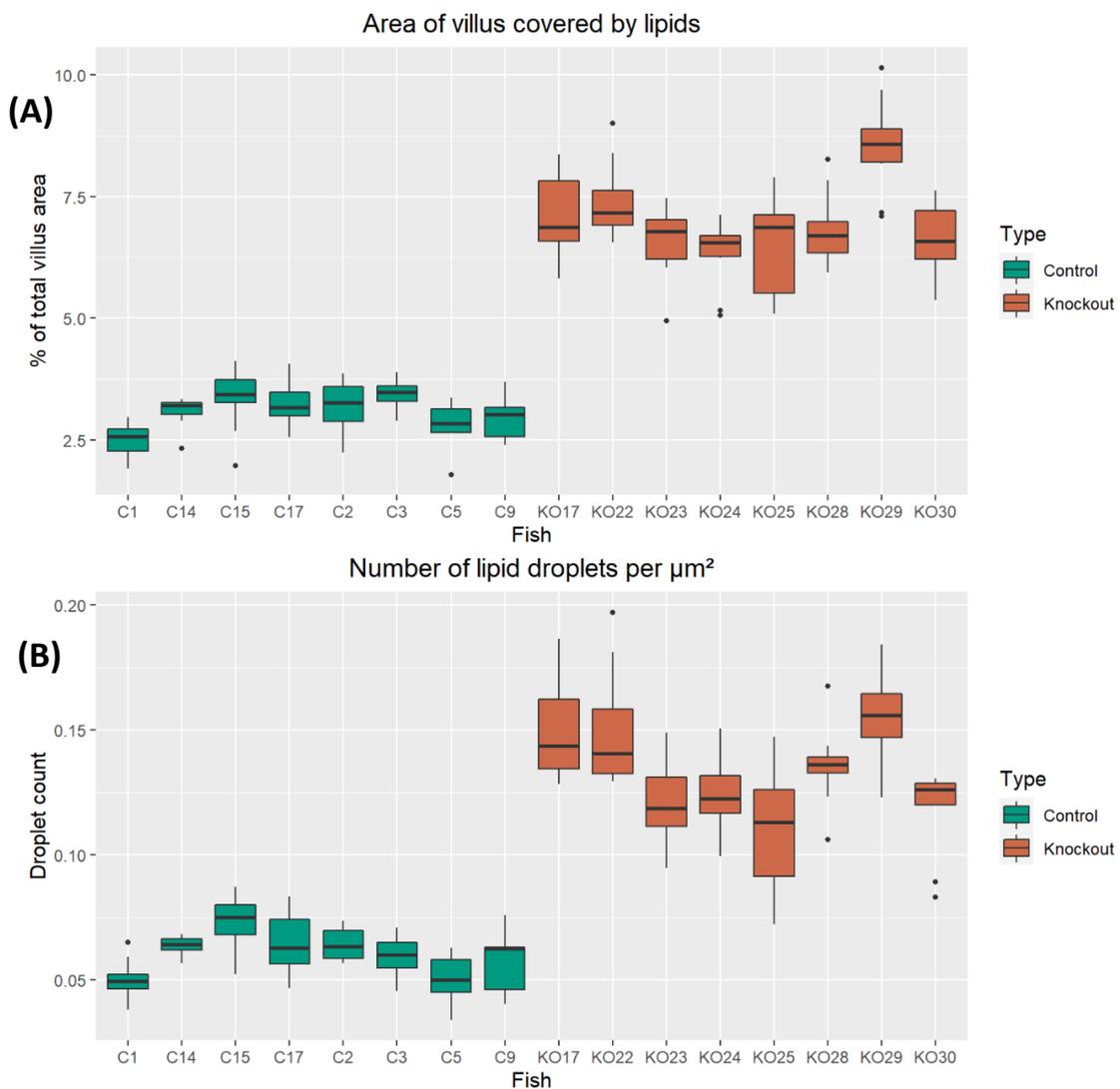


Figure 12: Area and quantity of lipid droplets found in mid-intestine villi of knockout and control salmon. Fish are tagged with ID along the x-axis. Control fish are labelled with a "C", while group *abcg2b/slc45a2_KO* fish are labelled with "KO".

(A) Fraction of the total villus area that is covered by lipids for each fish

(B) Average number of droplets per μm^2 for each fish

The sizes of individual droplets could possibly be impacted by variation in lipid transporter gene functions. Hence, differences in droplet size distribution were evaluated as well (Figure 13). The knockout group had a significantly larger median value for droplets above $1\mu\text{m}^2$ in size (Wilcoxon test $p\text{-value} = 2.686e^{-8}$). However, based on the entire size range the control group had a significantly larger median value than the knockout group ($0.265\mu\text{m}^2$ and $0.228\mu\text{m}^2$ respectively, Wilcoxon test $p\text{-value} = 2.409e^{-14}$). This is because $\sim 88\%$ of all droplets in both the wild type and the knockout group were smaller than $1\mu\text{m}^2$, and droplets within that size range were significantly bigger in the control group.

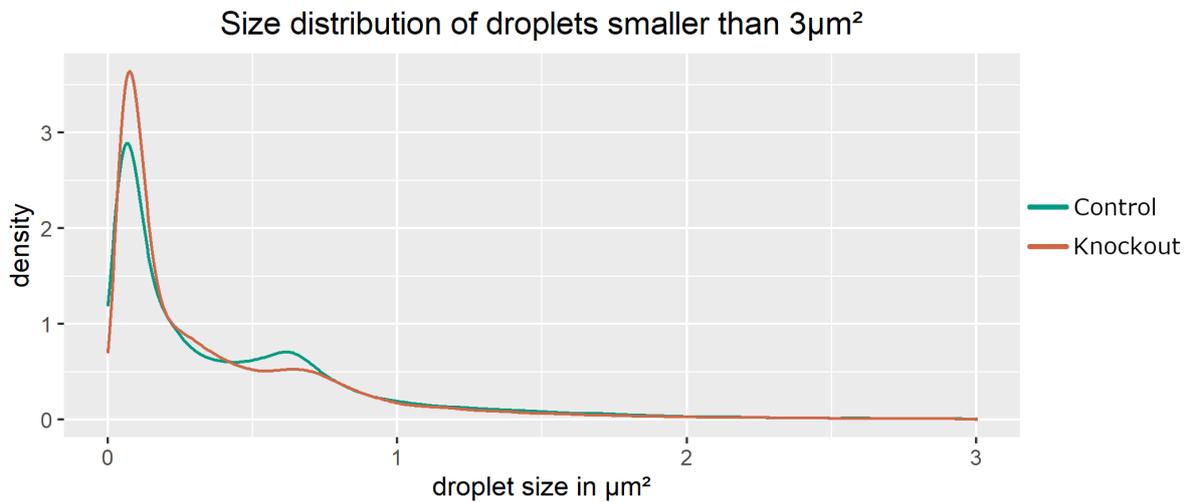


Figure 13: Density plot showing the distribution of droplets that range from 0.037 to $3\mu\text{m}^2$ in size.

Only a small fraction (less than 2%) of the droplets were bigger than $3\mu\text{m}^2$. However, these droplets represented 12.53% and 22.60% of the total lipid coverage in the control and knockout group, respectively. Knockout droplets above the size of $3\mu\text{m}^2$ were significantly larger than control droplets above $3\mu\text{m}^2$ (Wilcoxon test $p\text{-value} < 2.2e^{-16}$, Figure 14).

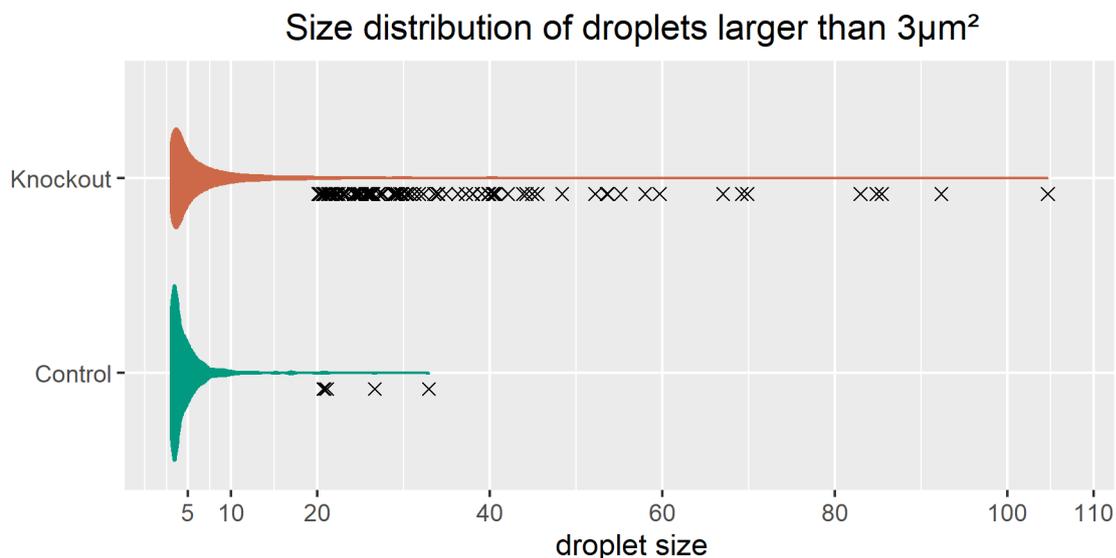


Figure 14: Violin plot showing the distribution of droplets larger than $3\mu\text{m}^2$ in size. Black "X" below the respective density marks droplets above $20\mu\text{m}^2$ in size.

Discussion

The two main objectives with this thesis were: (i) to evaluate a new method for estimating KO efficiency in mosaic gene edited fish, and (ii) to functionally characterize the *abcg2b* gene and its role in lipid transport and flesh coloration by using CRISPR gene editing. In the following discussion, I will first focus on estimation of KO efficiency and leave the role of *abcg2b* in flesh coloration to be discussed at the end.

getPCR estimates high KO efficiencies well

Estimation of CRISPR KO efficiency by high throughput sequencing is the gold standard method used to characterize the mosaic levels in F0 gene edited animals (Edwardsen et al., 2014; Jin, Liao, Migaud & Davie, 2020). However, this method is expensive and time consuming. Here we tested a simpler and cheaper method called getPCR as an alternative.

The comparison of getPCR with sequencing-based knockout estimation (Figure 9) shows that the two methods sometimes differ quite substantially in KO efficiency estimates. In the case of *bco1-like* knockout salmon, the getPCR-based KO efficiency estimate was drastically smaller when sequencing-based KO efficiencies were low (~20% or less). This could indicate that the getPCR method is too inaccurate in low knockout individuals. However, it is likely that getPCR can still be useful. For all sequencing-based KO efficiencies above 20% getPCR estimations are slightly higher, but quite comparable to sequencing-based estimations. It is in fact possible that the getPCR estimates are more accurate than the sequencing-based estimates when KO efficiency is high. The sequencing-based KO efficiency estimation is inferred from the base with the highest mismatch percentage around the Cas9 cut site (see Figure 7 in the Method-section). Edits that lead to a phenotypic knockout but do not affect the base with the highest mismatch percentage will thereby not be included in the estimate. The KO efficiency estimation based on sequencing data therefore represents an estimate of the “minimum KO efficiency”, but the true KO efficiency could be higher. The getPCR-based method on the other hand includes all edits around the Cas9 cut site, regardless of which base is affected.

Contamination can impact getPCR

A major drawback of the getPCR estimates are the poor results in samples with a low KO efficiency (<20%). However, the use of gDNA with better quality could potentially reduce this problem substantially. The few failed amplifications in the first getPCR (Figure 9B) indicate that the *bco1-like* KO gDNA likely contained PCR-inhibiting contamination. This hypothesis is strengthened by the second getPCR (containing 5x diluted gDNA) showing successful amplification for all *bco1-like* KO samples. Figure 10, which shows the KO efficiency of the *abcg2b/slc45a2_KO* group, also shows a consistently higher efficiency estimation with increasing intestine gDNA dilution. The higher KO efficiency in diluted samples could indicate that contamination decreases getPCR-based KO efficiency estimations. However, the calculation used to estimate getPCR KO efficiency (Equation 1) is purely based on relative differences within the sample and between the wild type and knockout fish. This means that the contamination would have to impact the control primer pair differently than the test primer pair, thereby skewing the relative difference between the primers within the knockout sample. Some contaminants, like bile salts and urea, could cause such a relative difference. They are known to impact amplicons of different lengths and GC-content differently, while others can alter melting temperatures and block access to the gDNA (Pionzio & McCord, 2014). A getPCR could be impacted more severely by these contaminants than a classic qPCR, since the extra high annealing temperature in a getPCR already leads to weaker binding between the primers and their target sites. Contamination could thereby indeed impact the primers and amplicons differently and skew the results.

Important steps in designing a good getPCR

The process of developing and testing the getPCR method for KO estimation underlined the importance of high purity gDNA, but the design of correct primers is also essential. In this project we struggled with primer-dimer amplification when both test primers covered the Cas9 cut site, to a higher degree than the article by Li et al. (2019) reported. Therefore, we recommend that only one of the primers overlaps the Cas9 cut site. Since the overlapping primer must cover the Cas9 cut site with 2-4 overhanging bases, its sequence and melting temperature (T_m) is predetermined by the sequence around the cut site. All other primers and PCR parameters should therefore be designed to fit the cut site primer. The annealing temperature of the getPCR should optimally lie 4°C above the T_m . This raises the specificity of the cut site primer but could also lead to primers with lower melting temperatures to not anneal to their target. We therefore recommend designing all other primers to have a melting temperature identical to or slightly higher than the T_m of the cut site primer. One primer design recommendation from Li et al. (2019) suggests to aim for adenine to be the terminal base on the 3' end of the cut site primer, as this showed best specificity compared to the other three bases.

getPCR cannot replace sequence based knockout estimation, but can supplement it

The getPCR method carried out in this thesis varies too greatly in accuracy to fully replace sequencing-based knockout estimation. A sequencing-based method can also provide additional information on the indel mutations induced by the gene editing event. This additional sequence information is however, apart from being more informative, not directly necessary for KO efficiency estimation. Knockout estimation by getPCR should despite its inaccuracies not be disregarded entirely, as it shows promise for improvement, and could be used to investigate relative KO efficiencies in a group. In all tested groups, even though the calculated KO efficiency was inaccurate, the trend was similar between getPCR and sequencing-based KO estimates. If the aim is to for example screen many individuals to identify particularly high KO efficiencies, getPCR could be a valuable tool. Fish showing the highest KO efficiencies could then be selected for more accurate sequence-based knockout-estimation. This reduces the total cost by reducing the amount of salmon that need to get evaluated by the more expensive sequencing-based method. It does not compromise the quality of the end-result, as the selected high KO salmon would still be evaluated with high quality sequencing-based data. For the selection of low-knockout salmon one should tread carefully when using getPCR, as the estimated KO efficiencies were highly inaccurate. In conclusion, we argue that getPCR is a powerful, fast and cheap complement to sequencing-based KO efficiency estimation methods and will be useful for cost-efficient screening of large numbers of fish when the KO efficiency is reasonably high.

Can a visual marker gene replace genetically based KO efficiency assays?

Using a visual marker to quickly estimate a KO efficiency is common practice (Jao, Wentz & Chen, 2013; Thakare et al., 2020). In this study we targeted *slc45a2* to induce albinism. Since both marker gene and target gene gRNAs are co-injected simultaneously this requires little extra effort. Overall, we found that salmon rated “high” in albinism generally show the highest KO efficiencies with getPCR as well (Figure 10A). Nevertheless, if we look closer at specific cases the picture is more nuanced and reveals several challenges with using a visual marker.

Firstly, performing two gene-editing events simultaneously poses a greater risk for the health of the gene edited organism. Any gene editing event bears the risk of off-target effects, which could turn toxic to the organism (Koo, Lee & Kim, 2015; Zhang et al., 2015). Two gene-editing events thereby increase the risk of dangerous side-effects.

Secondly, albinism as a marker gene relies on two assumptions: 1) That KO efficiency in the pigment cells is representative for the KO efficiency of all tissues. 2) That KO efficiency of the marker gene is representative for the KO efficiency of the target gene. In many cases these assumptions could be misleading, since the stem cells of the early embryonic stage might not all be equally affected by CRISPR. The KO efficiency in different tissues is therefore dependent on how many of their “ancestral” stem cells were successfully knocked out in each gene. Figure 10B visualizes the issue of differing tissue-specific KO efficiencies clearly, as three of five tested salmon had vastly different *abcg2b* KO efficiency estimations for liver and intestine. Figure 10A on the other hand underlines the issue of differing gene-specific KO efficiencies, with fish #30 and #41 both classified as “medium” albino, while getPCR estimated their *abcg2b* KO efficiency in intestine to differ by 45%.

Lastly, estimating KO efficiency by visually inspecting the degree of albinism in the skin is highly subjective. The degree of albinism in group *abcg2b/slc45a2_KO* is therefore not defined more precisely than “high”, “medium” or “low”. A more precise estimation can arguably also not be done when using getPCR, but the other problems bound to marker gene assisted knockout estimation do not exist. Genetically based methods therefore remain a more accurate and direct way to estimate KO efficiency and can be supported, but not replaced, by marker gene assisted KO efficiency estimation.

abcg2b knockout salmon show a clear increase in villus lipid content

The second half of this thesis focused on *abcg2b* and its functional connection to salmon flesh pigmentation. The enterocyte membrane transporter produced by *abcg2b* is suspected to act as an exporting channel for lipids, and thereby also astaxanthin. A knockout of *abcg2b* was therefore investigated by imaging villi in the mid-intestine and analysing their lipid content on a 2D plane.

Minimizing errors caused by image manipulation

Many tools exist for performing a binary image segmentation and choosing the right one for our specific image type was crucial for the quality of the resulting binary image. We aimed to find a method that produced few false positives without producing too many false negatives. The simplest way of achieving a binary segmentation from a fluorescence microscope image is global thresholding, which classifies all pixels above a set light intensity threshold as a positive signal (Rogowska, 2009). For the images taken in this thesis global thresholding was not accurate enough due to strong background luminescence (particularly in control salmon villi). Excluding the background signal would have led to the loss of too many small droplets with low light intensity, which were abundant in our samples (Figure 13). The segmentation toolkit Ilastik, which uses annotations provided by the user and machine learning to perform a binary segmentation (Berg et al., 2019), was therefore used instead. However, Ilastik relies on the user to know what should be classified as a positive signal and when the program is “trained enough” to use for final segmentation. In this thesis the impact of human error was minimized by annotating both background and lipid classification thoroughly, thereby giving a possibly erroneous annotation less weight. Training the binary segmentation algorithm on the whole image quality range was also emphasized, so that the segmentor could be verified to work well on the whole dataset. Additionally, we chose to exclude particles below 4 pixels in size as any lipid droplet that small could not confidently be distinguished from background noise by visual inspection. No further manipulation was done to ensure that all segmentation in the binary image had a direct connection to the original microscope image. The region of interest that was used during pixel classification in Fiji (see Figure 11B+D) was also linked directly to the microscope image, as it was hand-drawn around the villus depicted in it. We are confident to have chosen a suitable image manipulation path that produces a high-quality binary image without diverging too much from the original source image. Any skewing of results through image manipulation should therefore be negligible.

Lipid content and droplet sizes

The lipid area coverage was more than twice as high in *abcg2b* knockout villi than in the control group (Figure 12A). This shows that knocking out the *abcg2b* gene caused an increased lipid content in the enterocyte, and that the membrane transporter is likely connected to lipid export in these cells. The knockout group however also displayed a change in lipid droplet sizes. Figure 12B indicates that most of the increase in area coverage was much due to an increase in droplet number rather than an increase in mean droplet size. The density plot shown in Figure 13 also confirms this by showing that the largest fraction of droplets was below $1\mu\text{m}^2$ in size for both groups. The median droplet size was in fact larger in the control group. However, even though most droplets were relatively small, a few droplets in the knockout group were much larger than any droplets in the control group (see Figure 14). Even though these few droplets did not impact the median droplet size very much, they stood for a considerable portion of the lipid coverage and thereby impacted the overall lipid content in the villus greatly. The high lipid coverage in the knockout villi is thereby composed of a highly increased number of small droplets, as well as the formation of a few large droplets. Visual inspection of the knockout villi shows most of these larger droplets to be situated close to the centre of the villus (see Figure 11B and Supplemental table 5). We believe these larger droplets form as smaller droplets travel from the edges of the villus towards the centre, creating high droplet density until the smaller droplets finally fuse to larger droplets. Since the *abcg2b* transporter is situated on the outer edges of the villus, its knockout is believed to block the export of smaller droplets in the outer edges, creating a build-up. The knockout thereby directly increases the number of droplets in the villus but is likely not directly linked to the formation of large droplets. In conclusion, *abcg2b* activity is linked to lipid export in the salmon enterocyte, as a knockout of this gene leads to a considerable increase in both lipid droplet number and over-all lipid coverage. The increased lipid content also caused the formation of a few large lipid droplets, but *abcg2b* activity is likely not directly responsible for their formation.

Low KO efficiency is enough to create a high difference in lipid content

Low-knockout individuals carry a high fraction of wild type cells with functioning *abcg2b* transporter proteins. We therefore expected the lipid content of low-knockout individuals to be closer to the control group than high-knockout individuals. The data however shows no correlation between KO efficiency (Figure 10A) and the measured lipid content (Figure 12A), even though the differences between the control and knockout group are highly significant. This indicates that a small reduction in export efficiency can lead to a much higher lipid content in the villus, and that reducing it any further does not increase the lipid content more. One possible explanation for this could be that a reduction in *abcg2b* transporter efficiency leads to a slow build-up because the import of lipids becomes greater than the export. Over time, the lipids build up until the observed ~7% lipid coverage seen in the knockout group is reached. Reaching the 7% threshold would likely take longer for low-knockout individuals, as their export capacity is only slightly reduced, but was still reached before the fish were sampled (4 months after start-feeding). Four months to reach maximum lipid build-up would be early enough for the aquaculture industry, as the tested salmon still are far from slaughter weight at that time. The reason lipid coverage does not build up past 7% could be due to regulatory factors that either directly increase *abcg2b* export efficiency and/or increase lipid export through other transporters.

A confident functional link between *abcg2b* and flesh coloration requires more research

In this thesis the *abcg2b* gene was confirmed to be edited and a corresponding phenotype was observed in the edited salmon. However, we did not provide evidence for causation between the *abcg2b* gene knockout and depletion of functional *abcg2b* protein in the cells. This is still a necessary step to confidently link the gene knockout to the observed functional change. The presence of the transporter protein could for example be investigated through a western blot of intestine tissue in knockout and control fish. Initially we planned to confirm the histology and relative presence of the transporter protein through immunostaining, but the currently ongoing Covid-19 pandemic led to delays and restrictions that made it impossible to perform this experiment in time.

Furthermore, we have not confirmed that the increased lipid content in the villus leads to an increased astaxanthin content in the muscles. This is the last step to directly show causation between *abcg2b* and flesh coloration. Previous research suggests that a higher lipid content in the feed leads to higher redness in the muscle (B. Bjerkeng et al., 1997). The GWAS-mediated connection between *abcg2b* and flesh coloration also strengthens the assumption that the observed lipid increase also increases the amount of astaxanthin in the villus, and subsequently in the muscle. A measurement of increased astaxanthin content in the villus is however still necessary, as there is no direct evidence that the import of more lipids into the enterocyte leads to the import of more astaxanthin as well. In this thesis we assume that lipid droplets within the enterocyte have a near constant fraction of dissolved astaxanthin, but this could prove untrue. The astaxanthin content could be investigated through fluorescence microscopy of the mid-intestine villi, just like this thesis did for the investigation of lipid content. For investigating astaxanthin content this microscopy should however be performed on *abcg2b* KO salmon that are fed a higher astaxanthin content than the salmon used in this thesis. Coloration could also be measured in the muscle tissue of older *abcg2b* KO salmon that have had enough time to deposit astaxanthin in the flesh. Measuring muscle redness (for example by using a spectrometer on fresh muscle tissue) would create a direct connection between the knockout and flesh coloration. However, it would not elucidate the functional role *abcg2b* plays in achieving this colour. Muscle colour measurements should therefore be used in combination to a direct measurement of astaxanthin in mid-intestine villi to further elucidate the functional role of *abcg2b* in flesh coloration.

Other genes connected to flesh coloration should be investigated as well

The GWAS-mediated correlation between *abcg2b* and flesh coloration makes it probable that the naturally occurring genotype found in red salmon leads to increased lipid content in the villus, and that this simultaneously leads to an increased astaxanthin content in the muscle. However, if the amount of one lipid soluble compound increases with lipid content, other lipid soluble compounds could increase too. Many well-known xenobiotics (such as PCB and dioxins) are lipid soluble, bioaccumulate over time, and can be toxic to both salmon and humans (Isosaari et al., 2004; Opperhuizen & Sijm, 1990). Since *abcg2b* also has been linked to export of xenobiotics (Sarkadi, Özvegy-Laczka, Németh & Váradi, 2004) one should make sure that the concentration of toxic chemicals does not increase along with the concentration of astaxanthin. A previous investigation found no significant difference in xenobiotic build-up between the two *abcg2b* genotypes, but also stated that “contaminants were very low for both genotypes, which could possibly mask a genotype effect” (see p. 21 of Zoric, Moen, et al. (2017)). Increased uptake of xenobiotics should therefore still be kept in mind as a possible unwanted side effect when using *abcg2b* to select for flesh redness, especially since a small xenobiotic increase in salmon can lead to a larger bioaccumulation in humans that consume the fish over a long period of time. Increasing astaxanthin uptake purely by simultaneously increasing lipid uptake is therefore not ideal. Other genes that could be connected to the import or metabolism of

astaxanthin should be investigated as well. The *bco1* and *bco1-like* genes are a good example of this, as they have been connected to flesh coloration and could impact astaxanthin metabolism rate (Helgeland et al., 2019; Zoric, Torgersen, et al., 2017). Decreasing the intracellular breakdown of astaxanthin could increase astaxanthin deposition in the muscle without simultaneously increasing xenobiotic deposition in the body.

Conclusion

In this thesis we tested the use of getPCR to estimate KO efficiency in gene-edited salmon and used fluorescence microscopy to quantify lipid content in mid intestine salmon villi.

The testing of the getPCR method for estimating KO efficiency underlines the importance of using high purity gDNA and well-designed primers for this method to work ideally. It proved to work best on high-knockout individuals, and could thereby be used as a fast and cost-efficient way to screen large numbers of fish and select the individuals with highest KO efficiency for further evaluation. This method is thereby a good alternative to using CRISPR based visual markers which firstly could lead to unwanted off-target effects in the genome and secondly only provides an indirect measure of the KO efficiency for the gene of interest.

The *abcg2b* knockout salmon showed a clear increase in lipid content in villi of the mid-intestine, which indicates that the *abcg2b* membrane transporter plays a role in lipid export from the enterocyte. If the increased lipid content in the enterocyte can be confidently linked to an increase in flesh coloration still has to be determined. Previous research however strongly indicates that higher lipid content in the enterocyte also leads to higher astaxanthin content in the muscle. In addition, *abcg2b* has been linked to flesh coloration via GWAS, implying a link between the observed effect *abcg2b* activity has on villus lipid content and the muscle coloration caused by astaxanthin. Therefore our model suggests that the *abcg2b* gene contributes to variation in flesh coloration by mediating the export of astaxanthin, along with lipids, from the enterocyte to the intestinal lumen. Higher activity of this transporter causes decreased astaxanthin uptake from the feed, lower pigment levels in the blood, and thereby reduced astaxanthin deposition in the muscle.

REFERENCES

- Alfnes F., Guttormsen A. G., Steine G. & Kolstad K. (2006). Consumers' Willingness to Pay for the Color of Salmon: A Choice Experiment with Real Economic Incentives. *American Journal of Agricultural Economics*, 88(4), 1050-1061. doi:<https://doi.org/10.1111/j.1467-8276.2006.00915.x>
- Arnesen P., Krogdahl A. & Sundby A. (1995). Nutrient digestibilities, weight gain and plasma and liver levels of carbohydrate in Atlantic salmon (*Salmo salar*, L.) fed diets containing oats and maize. *Aquaculture Nutrition*, 1(3), 151-158. doi:<https://doi.org/10.1111/j.1365-2095.1995.tb00039.x>
- Baranski M., Moen T. & Våge D. (2010). Mapping of quantitative trait loci for flesh colour and growth traits in Atlantic salmon (*Salmo salar*). *Genetics, selection, evolution : GSE*, 42, 17. doi:<https://doi.org/10.1186/1297-9686-42-17>
- Barrangou R., Fremaux C., Deveau H., Richards M., Boyaval P., Moineau S., . . . Horvath P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*, 315(5819), 1709. doi:<https://doi.org/10.1126/science.1138140>
- Berg S., Kutra D., Kroeger T., Straehle C. N., Kausler B. X., Haubold C., . . . Kreshuk A. (2019). ilastik: interactive machine learning for (bio)image analysis. *Nature Methods*, 16(12), 1226-1232. doi:<https://doi.org/10.1038/s41592-019-0582-9>
- Bessen J. L., Afeyan L. K., Dančík V., Koblan L. W., Thompson D. B., Lechner C., . . . Liu D. R. (2019). High-resolution specificity profiling and off-target prediction for site-specific DNA recombinases. *Nature Communications*, 10(1), 1937. doi:<https://doi.org/10.1038/s41467-019-09987-0>
- Bjerkeng B. (2008). Carotenoids in Aquaculture: Fish and Crustaceans. In G. Britton, S. Liaaen-Jensen, & H. Pfander (Eds.), *Carotenoids: Volume 4: Natural Functions* (pp. 237-254). Basel: Birkhäuser Basel.
- Bjerkeng B. & Berge G. M. (2000). Apparent digestibility coefficients and accumulation of astaxanthin E/Z isomers in Atlantic salmon (*Salmo salar* L.) and Atlantic halibut (*Hippoglossus hippoglossus* L.). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 127(3), 423-432. doi:[https://doi.org/10.1016/S0305-0491\(00\)00278-9](https://doi.org/10.1016/S0305-0491(00)00278-9)
- Bjerkeng B., Refstie S., Fjalestad K. T., Storebakken T., Rødbotten M. & Roem A. J. (1997). Quality parameters of the flesh of Atlantic salmon (*Salmo salar*) as affected by dietary fat content and full-fat soybean meal as a partial substitute for fish meal in the diet. *Aquaculture*, 157(3), 297-309. doi:[https://doi.org/10.1016/S0044-8486\(97\)00162-2](https://doi.org/10.1016/S0044-8486(97)00162-2)
- Bohn T. (2008). Bioavailability of Non-Provitamin A Carotenoids. *Current Nutrition & Food Science*, 4(4), 240-258. doi:<https://doi.org/10.2174/157340108786263685>
- Borel P., Grolier P., Armand M., Partier A., Lafont H., Lairon D. & Braesco V. (1996). Carotenoids in biological emulsions: Solubility, surface-to-core distribution, and release from lipid droplets. *Journal of lipid research*, 37, 250-261. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/9026524/>
- Burma S., Chen B. P. C. & Chen D. J. (2006). Role of non-homologous end joining (NHEJ) in maintaining genomic integrity. *DNA Repair*, 5(9), 1042-1048. doi:<https://doi.org/10.1016/j.dnarep.2006.05.026>
- Christiansen R., Glette J., Lie Ø., Torrissen O. J. & Waagbø R. (1995). Antioxidant status and immunity in Atlantic salmon, *Salmo salar* L., fed semi-purified diets with and without astaxanthin supplementation. *Journal of Fish Diseases*, 18(4), 317-328. doi:<https://doi.org/10.1111/j.1365-2761.1995.tb00308.x>
- Christiansen R., Lie Ø. & Torrissen O. J. (1994). Effect of astaxanthin and vitamin A on growth and survival during first feeding of Atlantic salmon, *Salmo salar* L. *Aquaculture Research*, 25(9), 903-914. doi:<https://doi.org/10.1111/j.1365-2109.1994.tb01352.x>

- Christiansen R. & Torrissen O. J. (1996). Growth and survival of Atlantic salmon, *Salmo salar* L. fed different dietary levels of astaxanthin. Juveniles. *Aquaculture Nutrition*, 2(1), 55-62. doi:<https://doi.org/10.1111/j.1365-2095.1996.tb00008.x>
- Cohen S. N. & Chang A. C. (1973). Recircularization and autonomous replication of a sheared R-factor DNA segment in *Escherichia coli* transformants. *Proceedings of the National Academy of Sciences of the United States of America*, 70(5), 1293-1297. doi:<https://doi.org/10.1073/pnas.70.5.1293>
- Correa K., Bangera R., Figueroa R., Lhorente J. P. & Yáñez J. M. (2017). The use of genomic information increases the accuracy of breeding value predictions for sea louse (*Caligus rogercresseyi*) resistance in Atlantic salmon (*Salmo salar*). *Genetics Selection Evolution*, 49(1), 15. doi:<https://doi.org/10.1186/s12711-017-0291-8>
- Dang Y., Jia G., Choi J., Ma H., Anaya E., Ye C., . . . Wu H. (2015). Optimizing sgRNA structure to improve CRISPR-Cas9 knockout efficiency. *Genome Biology*, 16(1), 280. doi:<https://doi.org/10.1186/s13059-015-0846-3>
- Deltcheva E., Chylinski K., Sharma C. M., Gonzales K., Chao Y., Pirzada Z. A., . . . Charpentier E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340), 602-607. doi:<https://doi.org/10.1038/nature09886>
- Desmarchelier C. & Borel P. (2017). Overview of carotenoid bioavailability determinants: From dietary factors to host genetic variations. *Trends in Food Science & Technology*, 69, 270-280. doi:<https://doi.org/10.1016/j.tifs.2017.03.002>
- Dose J., Matsugo S., Yokokawa H., Koshida Y., Okazaki S., Seidel U., . . . Esatbeyoglu T. (2016). Free Radical Scavenging and Cellular Antioxidant Properties of Astaxanthin. *International journal of molecular sciences*, 17(1), 103. doi:<https://doi.org/10.3390/ijms17010103>
- Edvardsen R. B., Leininger S., Kleppe L., Skaftnesmo K. O. & Wargelius A. (2014). Targeted Mutagenesis in Atlantic Salmon (*Salmo salar* L.) Using the CRISPR/Cas9 System Induces Complete Knockout Individuals in the F0 Generation. *PLOS ONE*, 9(9), e108622. doi:<https://doi.org/10.1371/journal.pone.0108622>
- Elliott N. & Kube P. (2009). Development and early results of the Tasmanian Atlantic Salmon breeding program. *Proc Assoc Advmt Anim Breed Genet*, 18. Retrieved from [https://www.researchgate.net/publication/267416861_Aquaculture DEVELOPMENT AND EARLY RESULTS OF THE TASMANIAN ATLANTIC SALMON BREEDING PROGRAM](https://www.researchgate.net/publication/267416861_Aquaculture_DEVELOPMENT_AND_EARLY_RESULTS_OF_THE_TASMANIAN_ATLANTIC_SALMON_BREEDING_PROGRAM)
- Espe M. (2008). 9 - Understanding factors affecting flesh quality in farmed fish. In Ø. Lie (Ed.), *Improving Farmed Fish Quality and Safety* (pp. 241-264): Woodhead Publishing.
- Ferreira M., Costa J. & Reis-Henriques M. A. (2014). ABC transporters in fish species: a review. *Frontiers in Physiology*, 5, 266. doi:<https://doi.org/10.3389/fphys.2014.00266>
- Fielding C. J. & Fielding P. E. (2008). CHAPTER 19 - Dynamics of lipoprotein transport in the circulatory system. In D. E. Vance & J. E. Vance (Eds.), *Biochemistry of Lipids, Lipoproteins and Membranes (Fifth Edition)* (pp. 533-553). San Diego: Elsevier.
- Flister M. J., Tsaih S.-W., O'Meara C. C., Endres B., Hoffman M. J., Geurts A. M., . . . Moreno C. (2013). Identifying multiple causative genes at a single GWAS locus. *Genome Research*, 23(12), 1996-2002. doi:<https://doi.org/10.1101/gr.160283.113>
- Foote C., Brown G. & Hawryshyn C. (2004). Female colour and male choice in sockeye salmon: Implications for the phenotypic convergence of anadromous and nonanadromous morphs. *Animal Behaviour*, 67, 69-83. doi:<https://doi.org/10.1016/j.anbehav.2003.02.004>
- Furr H. C. & Clark R. M. (1997). Intestinal absorption and tissue distribution of carotenoids. *The Journal of Nutritional Biochemistry* 8(7), 364-377. doi:[https://doi.org/10.1016/S0955-2863\(97\)00060-0](https://doi.org/10.1016/S0955-2863(97)00060-0)
- Gasiunas G., Barrangou R., Horvath P. & Siksnys V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences*, 109(39), E2579. doi:<https://doi.org/10.1073/pnas.1208507109>

- Gjedrem T. (2010). The first family-based breeding program in aquaculture. *Reviews in Aquaculture*, 2(1), 2-15. doi:<https://doi.org/10.1111/j.1753-5131.2010.01011.x>
- Heidenreich E., Novotny R., Kneidinger B., Holzmann V. & Wintersberger U. (2003). Non-homologous end joining as an important mutagenic process in cell cycle-arrested cells. *The EMBO Journal*, 22(9), 2274-2283. doi:<https://doi.org/10.1093/emboj/cdg203>
- Helgeland H., Sodeland M., Zoric N., Torgersen J. S., Grammes F., von Lintig J., . . . Våge D. I. (2019). Genomic and functional gene studies suggest a key role of beta-carotene oxygenase 1 like (bco1l) gene in salmon flesh color. *Scientific Reports*, 9(1), 20061. doi:<https://doi.org/10.1038/s41598-019-56438-3>
- Hendel A., Bak R. O., Clark J. T., Kennedy A. B., Ryan D. E., Roy S., . . . Porteus M. H. (2015). Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nature Biotechnology*, 33(9), 985-989. doi:<https://doi.org/10.1038/nbt.3290>
- Higuera-Ciapara I., Felix-Valenzuela L Fau - Goycoolea F. M. & Goycoolea F. M. (2006). Astaxanthin: a review of its chemistry and applications. (1040-8398 (Print)). doi:<https://doi.org/10.1080/10408690590957188>
- Houston R. D., Taggart J. B., Cézard T., Bekaert M., Lowe N. R., Downing A., . . . Hamilton A. (2014). Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC Genomics*, 15(1), 90. doi:<https://doi.org/10.1186/1471-2164-15-90>
- Hsu P. D., Scott D. A., Weinstein J. A., Ran F. A., Konermann S., Agarwala V., . . . Zhang F. (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology*, 31(9), 827-832. doi:<https://doi.org/10.1038/nbt.2647>
- Iorio A. L., Ros M. d., Fantappiè O., Lucchesi M., Facchini L., Stival A., . . . Sardi I. (2016). Blood-Brain Barrier and Breast Cancer Resistance Protein: A Limit to the Therapy of CNS Tumors and Neurodegenerative Diseases. *Anti-cancer agents in medicinal chemistry*, 16(7), 810-815. doi:<https://doi.org/10.2174/1871520616666151120121928>
- Isosaari P., Kiviranta H., Lie O., Lundebye A. K., Ritchie G. & Vartiainen T. (2004). Accumulation and distribution of polychlorinated dibenzo-p-dioxin, dibenzofuran, and polychlorinated biphenyl congeners in Atlantic salmon (*Salmo salar*). *Environ Toxicol Chem*, 23(7), 1672-1679. doi:<https://doi.org/10.1897/03-367>
- Jaenisch R. & Mintz B. (1974). Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 71(4), 1250-1254. doi:<https://doi.org/10.1073/pnas.71.4.1250>
- Jao L.-E., Wente S. R. & Chen W. (2013). Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proceedings of the National Academy of Sciences*, 110(34), 13904. doi:<https://doi.org/10.1073/pnas.1308335110>
- Jin Y. H., Liao B., Migaud H. & Davie A. (2020). Physiological impact and comparison of mutant screening methods in piwil2 KO founder Nile tilapia produced by CRISPR/Cas9 system. *Scientific Reports*, 10(1), 12600. doi:<https://doi.org/10.1038/s41598-020-69421-0>
- Jinek M., Chylinski K., Fonfara I., Hauer M., Doudna J. A. & Charpentier E. (2012). A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, 337(6096), 816. doi:<https://doi.org/10.1126/science.1225829>
- Kidd P. (2011). Astaxanthin, Cell Membrane Nutrient with Diverse Clinical Benefits and Anti-Aging Potential. *Alternative medicine review : a journal of clinical therapeutic*, 16, 355-364. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/22214255/>
- Kim Y., Kweon J. & Kim J.-S. (2013). TALENs and ZFNs are associated with different mutation signatures. *Nature Methods*, 10(3), 185-185. doi:10.1038/nmeth.2364
- Kim Y. G., Cha J. & Chandrasegaran S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, 93(3), 1156-1160. doi:<https://doi.org/10.1073/pnas.93.3.1156>

- Koo T., Lee J. & Kim J.-S. (2015). Measuring and Reducing Off-Target Activities of Programmable Nucleases Including CRISPR-Cas9. *Molecules and cells*, 38(6), 475-481.
doi:<https://doi.org/10.14348/molcells.2015.0103>
- Korte A. & Farlow A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, 9(1), 29. doi:<https://doi.org/10.1186/1746-4811-9-29>
- Leng P.-f., Lübberstedt T. & Xu M.-l. (2017). Genomics-assisted breeding – A revolutionary strategy for crop improvement. *Journal of Integrative Agriculture*, 16(12), 2674-2685.
doi:[https://doi.org/10.1016/S2095-3119\(17\)61813-6](https://doi.org/10.1016/S2095-3119(17)61813-6)
- Li B., Ren N., Yang L., Liu J. & Huang Q. (2019). A qPCR method for genome editing efficiency determination and single-cell clone screening in human cells. *Scientific Reports*, 9(1), 18877.
doi:<https://doi.org/10.1038/s41598-019-55463-6>
- Lien S., Koop B. F., Sandve S. R., Miller J. R., Kent M. P., Nome T., . . . Davidson W. S. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature*, 533(7602), 200-205.
doi:<https://doi.org/10.1038/nature17164>
- Livak K. J. & Schmittgen T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2- $\Delta\Delta$ CT Method. *Methods*, 25(4), 402-408.
doi:<https://doi.org/10.1006/meth.2001.1262>
- Lobo G. P., Amengual J., Palczewski G., Babino D. & von Lintig J. (2012). Mammalian Carotenoid-oxygenases: Key players for carotenoid function and homeostasis. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1821(1), 78-87.
doi:<https://doi.org/10.1016/j.bbalip.2011.04.010>
- Lorenz R. T. & Cysewski G. R. (2000). Commercial potential for Haematococcus microalgae as a natural source of astaxanthin. *Trends in Biotechnology*, 18(4), 160-167.
doi:[https://doi.org/10.1016/S0167-7799\(00\)01433-5](https://doi.org/10.1016/S0167-7799(00)01433-5)
- Makarova K. S., Haft D. H., Barrangou R., Brouns S. J. J., Charpentier E., Horvath P., . . . Koonin E. V. (2011). Evolution and classification of the CRISPR-Cas systems. *Nature Reviews Microbiology*, 9(6), 467-477. doi:<https://doi.org/10.1038/nrmicro2577>
- Miki W. (1991). Biological Functions and Activities of Animal Carotenoids. *Pure and Applied Chemistry - PURE APPL CHEM*, 63, 141-146. doi:<https://doi.org/10.1351/pac199163010141>
- Miller J. C., Tan S., Qiao G., Barlow K. A., Wang J., Xia D. F., . . . Rebar E. J. (2011). A TALE nuclease architecture for efficient genome editing. *Nature Biotechnology*, 29(2), 143-148.
doi:<https://doi.org/10.1038/nbt.1755>
- Moe E. (2017). *The Norwegian aquaculture analysis 2017*. Retrieved from https://assets.ey.com/content/dam/ey-sites/ey-com/no_no/topics/consumer-products/pdfs/ey-the-norwegian-aquaculture-analysis-2017.pdf
- Moen T. & Ødegård J. (2014). Genomics in Selective Breeding of Atlantic Salmon. *10th World Congress of Genetics Applied to Livestock Production, Species Breeding: Breeding in aquaculture species*. Retrieved from <http://www.wcgalp.org/proceedings/2014/genomics-selective-breeding-atlantic-salmon>
- Norwegian Seafood Council. (2020). Nøkkeltall for norsk sjømateksport. Retrieved from <https://nokkeltall.seafood.no/>
- Opperhuizen A. & Sijm D. T. H. M. (1990). Bioaccumulation and biotransformation of polychlorinated dibenzo-p-dioxins and dibenzofurans in fish. *Environmental Toxicology and Chemistry*, 9(2), 175-186. doi:<https://doi.org/10.1002/etc.5620090207>
- Page G. I. & Davies S. J. (2003). Hepatic carotenoid uptake in rainbow trout (*Oncorhynchus mykiss*) using an isolated organ perfusion model. *Aquaculture*, 225(1), 405-419.
doi:[https://doi.org/10.1016/S0044-8486\(03\)00305-3](https://doi.org/10.1016/S0044-8486(03)00305-3)
- Parker R. S. (1996). Absorption, metabolism, and transport of carotenoids. (0892-6638 (Print)). Retrieved from <https://pubmed.ncbi.nlm.nih.gov/8621054/>

- Pionzio A. M. & McCord B. R. (2014). The effect of internal control sequence and length on the response to PCR inhibition in real-time PCR quantitation. *Forensic Science International: Genetics*, 9, 55-60. doi:<https://doi.org/10.1016/j.fsigen.2013.10.010>
- Porteus M. H. & Baltimore D. (2003). Chimeric Nucleases Stimulate Gene Targeting in Human Cells. *Science*, 300(5620), 763. doi:<https://doi.org/10.1126/science.1078395>
- Rajasingh H., Gjuvsland A. B., Våge D. I. & Omholt S. W. (2008). When parameters in dynamic models become phenotypes: a case study on flesh pigmentation in the chinook salmon (*Oncorhynchus tshawytscha*). *Genetics*, 179(2), 1113. doi:<https://doi.org/10.1534/genetics.108.087064>
- Rajasingh H., Øyehaug L., Våge D. I. & Omholt S. W. (2006). Carotenoid dynamics in Atlantic salmon. *BMC Biology*, 4(1), 10. doi:<https://doi.org/10.1186/1741-7007-4-10>
- Reboul E. (2013). Absorption of Vitamin A and Carotenoids by the Enterocyte: Focus on Transport Proteins. *Nutrients*, 5, 3563-3581. doi:<https://doi.org/10.3390/nu5093563>
- Reboul E. & Borel P. (2011). Proteins involved in uptake, intracellular transport and basolateral secretion of fat-soluble vitamins and carotenoids by mammalian enterocytes. *Prog Lipid Res*, 50(4), 388-402. doi:<https://doi.org/10.1016/j.plipres.2011.07.001>
- Richardsen R., Myhre M. S. & Tyholt I. L. (2019). *Nasjonal betydning av sjømatnæringen*. Retrieved from https://www.sintef.no/globalassets/sintef-ocean/pdf/nasjonal-verdiskapning_2018tall_endelig_200619.pdf
- Robey R. W., To K. K. K., Polgar O., Dohse M., Fetsch P., Dean M. & Bates S. E. (2009). ABCG2: A perspective. *Advanced Drug Delivery Reviews*, 61(1), 3-13. doi:<https://doi.org/10.1016/j.addr.2008.11.003>
- Rogowska J. (2009). Chapter 5 - Overview and Fundamentals of Medical Image Segmentation. In I. N. Bankman (Ed.), *Handbook of Medical Image Processing and Analysis (Second Edition)* (pp. 74-75). Burlington: Academic Press.
- Sae-Lim P., Kause A., Lillehammer M. & Mulder H. A. (2017). Estimation of breeding values for uniformity of growth in Atlantic salmon (*Salmo salar*) using pedigree relationships or single-step genomic evaluation. *Genetics Selection Evolution*, 49(1), 33. doi:<https://doi.org/10.1186/s12711-017-0308-3>
- San Filippo J., Sung P. & Klein H. (2008). Mechanism of Eukaryotic Homologous Recombination. *Annual Review of Biochemistry*, 77(1), 231-233. doi:<https://doi.org/10.1146/annurev.biochem.77.061306.125255>
- Sander J. D. & Joung J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology*, 32(4), 347-355. doi:<https://doi.org/10.1038/nbt.2842>
- Sarkadi B., Özvegy-Laczka C., Német K. & Váradi A. (2004). ABCG2 – a transporter for all seasons. *FEBS Letters*, 567(1), 116-120. doi:<https://doi.org/10.1016/j.febslet.2004.03.123>
- Schiedt K., Leuenberger F. J., Vecchi M. & Glinz E. (1985). Absorption, retention and metabolic transformations of carotenoids in rainbow trout, salmon and chicken. *Pure and Applied Chemistry*, 57(5), 685-692. doi:<https://doi.org/10.1351/pac198557050685>
- Schindelin J., Arganda-Carreras I., Frise E., Kaynig V., Longair M., Pietzsch T., . . . Cardona A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7), 676-682. doi:<https://doi.org/10.1038/nmeth.2019>
- Skarstein F. & Folstad I. (1996). Sexual Dichromatism and the Immunocompetence Handicap: An Observational Approach Using Arctic Charr. *Oikos*, 76, 359. doi:<https://doi.org/10.2307/3546208>
- Sternberg S. H., Redding S., Jinek M., Greene E. C. & Doudna J. A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507(7490), 62-67. doi:<https://doi.org/10.1038/nature13011>
- Storebakken T. & No H. K. (1992). Pigmentation of rainbow trout. *Aquaculture*, 100(1), 209-229. doi:[https://doi.org/10.1016/0044-8486\(92\)90372-R](https://doi.org/10.1016/0044-8486(92)90372-R)

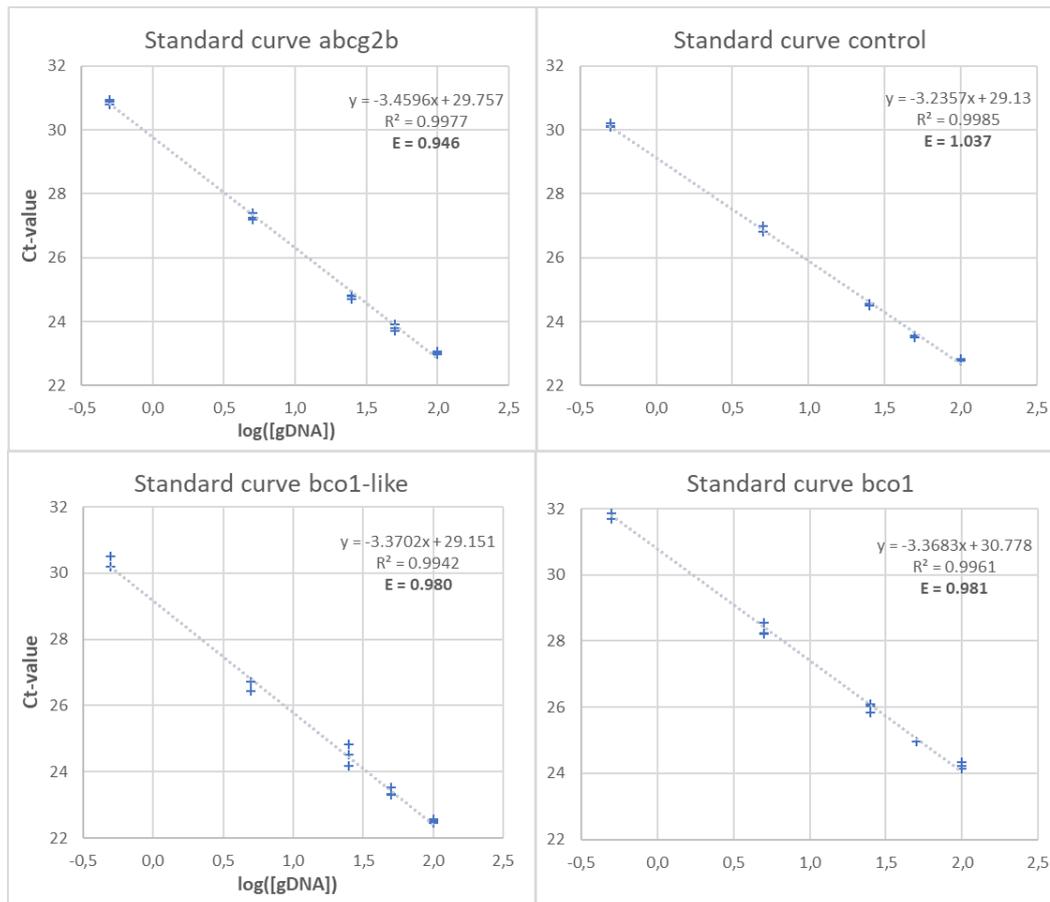
- Szczepek M., Brondani V., Büchel J., Serrano L., Segal D. J. & Cathomen T. (2007). Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nature Biotechnology*, 25(7), 786-793. doi:<https://doi.org/10.1038/nbt1317>
- Thakare S. S., Bansal N., Vanchinathan S., Rama Prashat G., Krishnan V., Sachdev A., . . . Vinutha T. (2020). GFP tagging based method to analyze the genome editing efficiency of CRISPR/Cas9-gRNAs through transient expression in *N. benthamiana*. *Journal of Plant Biochemistry and Biotechnology*, 29(2), 183-192. doi:<https://doi.org/10.1007/s13562-019-00540-0>
- The International HapMap Consortium. (2005). A haplotype map of the human genome. *Nature*, 437(7063), 1299-1320. doi:<https://doi.org/10.1038/nature04226>
- Thodesen J. & Gjedrem T. (2006). Breeding programs on Atlantic salmon in Norway: lessons learned. In R. W. Ponzoni, B. O. Acosta, & A. G. Ponniah (Eds.), *Development of aquatic animal genetic improvement and dissemination programs: current status and action plans* (pp. pp. 22-26).
- Torrissen O. J. (1989). Pigmentation of salmonids: Interactions of astaxanthin and canthaxanthin on pigment deposition in rainbow trout. *Aquaculture*, 79(1), 363-374. doi:[https://doi.org/10.1016/0044-8486\(89\)90478-X](https://doi.org/10.1016/0044-8486(89)90478-X)
- Tsai H.-Y., Matika O., Edwards S. M., Antolín-Sánchez R., Hamilton A., Guy D. R., . . . Houston R. D. (2017). Genotype Imputation To Improve the Cost-Efficiency of Genomic Selection in Farmed Atlantic Salmon. *G3: Genes/Genomes/Genetics*, 7(4), 1377. doi:<https://doi.org/10.1534/g3.117.040717>
- Part V – Exclusive Economic Zone, Article 56, (1982).
- Urnov F. D., Rebar E. J., Holmes M. C., Zhang H. S. & Gregory P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics*, 11(9), 636-646. doi:<https://doi.org/10.1038/nrg2842>
- Varshney R. K. (2016). Exciting journey of 10 years from genomes to fields and markets: Some success stories of genomics-assisted breeding in chickpea, pigeonpea and groundnut. *Plant Science*, 242, 98-107. doi:<https://doi.org/10.1016/j.plantsci.2015.09.009>
- Venvik T. (2005). *National Aquaculture Sector Overview. Norway. National Aquaculture Sector Overview Fact Sheets*. Retrieved from Food and Agriculture Organization of the United Nations; Fisheries and Aquaculture Department: http://www.fao.org/fishery/countrysector/naso_norway/en
- von der Ecken J., Müller M., Lehman W., Manstein D. J., Penczek P. A. & Raunser S. (2015). Structure of the F-actin–tropomyosin complex. *Nature*, 519(7541), 114-117. doi:<https://doi.org/10.1038/nature14033>
- White D. A., Ørnsrud R. & Davies S. J. (2003). Determination of carotenoid and vitamin A concentrations in everted salmonid intestine following exposure to solutions of carotenoid in vitro. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 136(3), 683-692. doi:[https://doi.org/10.1016/S1095-6433\(03\)00222-8](https://doi.org/10.1016/S1095-6433(03)00222-8)
- Yang Jin (CIGENE) (2020). [Report - Characterization of abcg2 genes in salmon and zebrafish].
- Young A. J., Pritchard J., Lowe G. M., Crampton V. & Buttle L. (2017). Reconstitution of muscle F-actin from Atlantic salmon (*Salmo salar* L.) with carotenoids—binding characteristics of astaxanthin and canthaxanthin. *Aquaculture Nutrition*, 23(6), 1296-1303. doi:<https://doi.org/10.1111/anu.12504>
- Ytrestøyl T. & Bjerkeng B. (2007). Dose response in uptake and deposition of intraperitoneally administered astaxanthin in Atlantic salmon (*Salmo salar* L.) and Atlantic cod (*Gadus morhua* L.). *Aquaculture*, 263(1), 179-191. doi:<https://doi.org/10.1016/j.aquaculture.2006.10.021>
- Ytrestøyl T., Dikiy A., Shumilina E., Bæverfjord G., Krasnov A., Ciampa A., . . . Ruyte B. (2019). *Effekt av fôr, temperatur og stress på pigmentering i laks*. Retrieved from Nofima: <https://nofima.no/publikasjon/1715939/>

- Ytrestøl T., Aas T. S. & Åsgård T. (2015). Utilisation of feed resources in production of Atlantic salmon (*Salmo salar*) in Norway. *Aquaculture*, 448, 365-374.
doi:<https://doi.org/10.1016/j.aquaculture.2015.06.023>
- Zaripheh S. & Erdman J. W., Jr. (2002). Factors that influence the bioavailability of xanthophylls. (0022-3166 (Print)). doi:<https://doi.org/10.1093/jn/132.3.531S>
- Zhang X.-H., Tee L. Y., Wang X.-G., Huang Q.-S. & Yang S.-H. (2015). Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Molecular Therapy - Nucleic Acids*, 4, e264.
doi:<https://doi.org/10.1038/mtna.2015.37>
- Zhu H., Li C. & Gao C. (2020). Applications of CRISPR–Cas in agriculture and plant biotechnology. *Nature Reviews Molecular Cell Biology*, 21(11), 661-677.
doi:<https://doi.org/10.1038/s41580-020-00288-9>
- Zoric N. (2017). *Characterization of genes and gene products influencing carotenoid metabolism in Atlantic salmon*. (Ph.D.). Norwegian University of Life Sciences, Ås, Retrieved from <http://hdl.handle.net/11250/2497990>
- Zoric N., Moen T., Korsvoll S. A., Kjølglum S., Santi N., Lien S., . . . Torgersen J. (2017). A missense mutation in the *abcg2-1a* gene (ATP-binding cassette sub-family G member 2) is strongly associated with muscle color in Atlantic salmon (*Salmo salar*). Retrieved from <http://hdl.handle.net/11250/2497990>
- Zoric N., Torgersen J., Grammes F., Lintig J. v. & Våge D. I. (2017). *Functional divergence of beta-carotene oxygenase 1 enzymes after gene duplication in salmon (Salmo salar)*. Retrieved from <http://hdl.handle.net/11250/2497990>
- Aas G. H., Bjerkeng B., Storebakken T. & Ruyter B. (1999). Blood appearance, metabolic transformation and plasma transport proteins of 14C-astaxanthin in Atlantic salmon (*Salmo salar* L.). *Fish Physiology and Biochemistry*, 21(4), 325-334.
doi:<https://doi.org/10.1023/A:1007890224389>
- Aas T. S., Ytrestøl T. & Åsgård T. (2019). Utilization of feed resources in the production of Atlantic salmon (*Salmo salar*) in Norway: An update for 2016. *Aquaculture Reports*, 15, 100216.
doi:<https://doi.org/10.1016/j.aqrep.2019.100216>

Appendix

Supplemental table 1: List of all primers that were tested for the use in getPCR.

Target gene	Primer name	Direction	Primer sequence
LOC106577560	Ctr_F	Forward	TCAGCGTTCATTGGCTTTGAGACCA
LOC106577560	Ctr_R	Reverse	CAGGCTGTTCTTGATCTCGTACTGGTT
<i>abcg2b</i>	A1_F	Forward	GATCCAGAGGTGATGTTCCAGGTGC
<i>abcg2b</i>	A1_F2	Forward	GTCCCAGCAGCAGATGATCCAGA
<i>abcg2b</i>	A1_F3	Forward	CAGAGGTGATGTTCCAGGTGCC
<i>abcg2b</i>	A1_F4	Forward	GAGGTGATGTTCCAGGTGCCTG
<i>abcg2b</i>	A1_R	Reverse	GGAGACGGTTGGACCAGGCAC
<i>abcg2b</i>	A1_R2	Reverse	GGTTGGACCAGGCACCTGGA
<i>abcg2b</i>	A1_R3	Reverse	GTGTAGTCTGGAGAAGGAGACGGTT
<i>bco1-like</i>	A3_F	Forward	TGATGCCATGTCTCAGTCCCTCATC
<i>bco1-like</i>	A3_F2	Forward	GCCTTAATAGCTGATGCCATGTCTCA
<i>bco1-like</i>	A3_F3	Forward	CATGTCTCAGTCCCTCATCGGC
<i>bco1-like</i>	A3_R	Reverse	ATTCAGTCCCGTTCTTGCCGATGA
<i>bco1-like</i>	A3_R2	Reverse	CTGCTTTGACAGGTTCCGGTGA



Supplemental figure 1: Standard curves for all four getPCR primer pairs, made on a dilution series of wild type gDNA (in triplicate). The values are visualised by plotting Ct-value over log starting quantity of gDNA. The slope of the linear trendline "y" stands proportional to the primer efficiency and was used to calculate the efficiency E (bold).

Supplemental table 2: getPCR values used to estimate the KO efficiency (ko getPCR). The given Ct-values for the test and control primer constructs are an average of three and measured on the respective KO gDNA. ΔCt was calculated by subtracting the Ct-value of the primer run on wt gDNA from the Ct-value of the same primer run on the respective KO gDNA. Rows marked with “-“ indicate that the sample never reached the threshold within the 40 getPCR cycles.

Group	Ct test	Ct ctr	ΔCt test	ΔCt ctr	ko getPCR	ko MiSeq
abcg2b_KO, getPCR#1	24,399	24.087	-1.230	-1.543	0.243	0.13
	26.437	25.842	0.809	0.212	0.321	0.25
	24.450	23.439	-1.178	-2.191	0.539	0.49
	30.186	25.970	4.558	0.340	0.939	0.74
bco1l_KO, getPCR#1	-	-	-	-	-	0.08
	31.688	26.110	5.547	0.480	0.968	0.64
	-	-	-	-	-	0.76
	33.809	27.134	7.668	1.504	0.985	0.93
abcg2b_KO, getPCR#2	24.856	24.460	-0.772	-1.171	0.273	0.13
	25.344	24.643	-0.284	-0.987	0.401	0.25
	24.843	23.854	-0.786	-1.776	0.523	0.49
	31.031	27.013	5.403	1.383	0.927	0.74
bco1l_KO, getPCR#2	29.438	29.049	2.020	2.303	-0.295	0.08
	33.168	27.807	5.751	1.061	0.958	0.64
	33.970	30.038	6.552	3.292	0.882	0.76
	34.501	27.844	7.084	1.098	0.983	0.93



Supplemental figure 2: Image of typical wild type salmon used as a control for both getPCR and lipid staining. The fish were bred and sampled at the same time as the ko group. Average weight $9,64 \pm 1,0g$



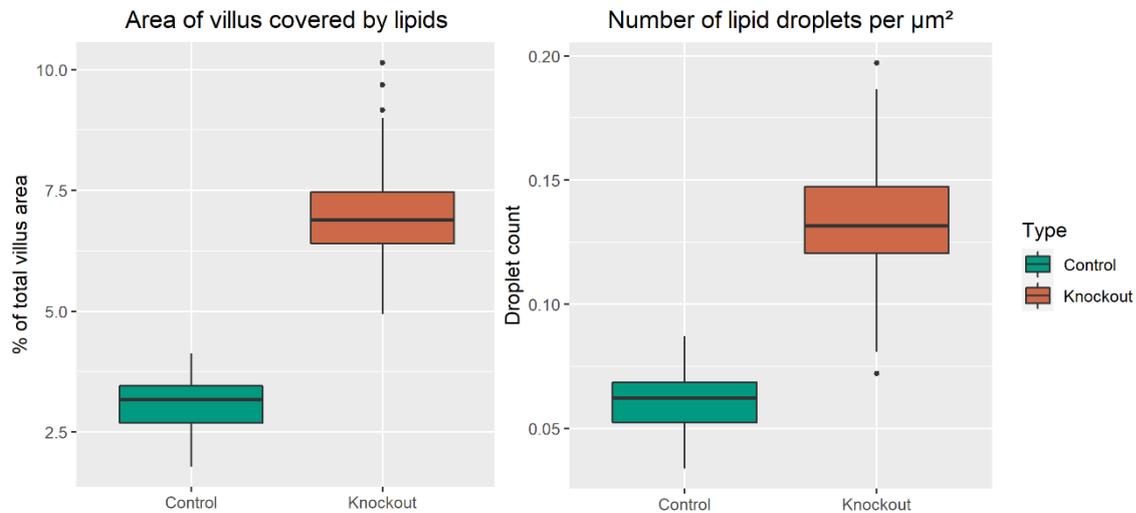
Supplemental figure 3: Photographs of the fourteen *abcg2b/slc45a2*_KO salmon used in *getPCR* knockout estimation. The body weight in grams is given to the right of the ID number the fish was given during sampling. Fish with a green star in the top left corner were used for lipid stain microscopy as well.

Supplemental table 3: KO efficiencies (getPCR KO) estimated by the first getPCR of the *abcg2b/slc45a2_KO* group. Fourteen fish were tested on gDNA isolated from the mid-intestine.

ID number	Albinism	getPCR KO	ID number	Albinism	getPCR KO
KO16	low	-0.070	KO28	high	0.321
KO17	high	0.407	KO29	low	0.098
KO21	high	0.105	KO30	medium	-0.080
KO22	low	0.106	KO33	high	0.278
KO23	high	0.432	KO34	medium	0.031
KO24	high	0.281	KO40	low	0.063
KO25	medium	0.107	KO41	medium	0.372

Supplemental table 4: KO efficiencies (getPCR KO) estimated by the second getPCR of the *abcg2b/slc45a2_KO* group. Five fish were tested on both liver and mid-intestine gDNA for comparison. The gDNA was diluted to 5ng/ μ L in advance.

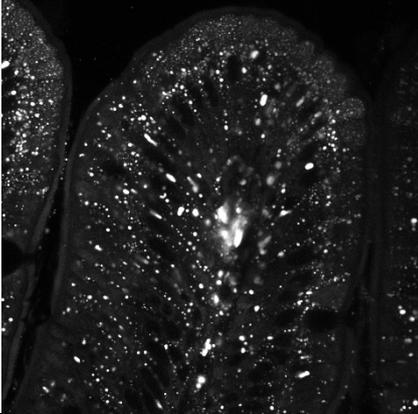
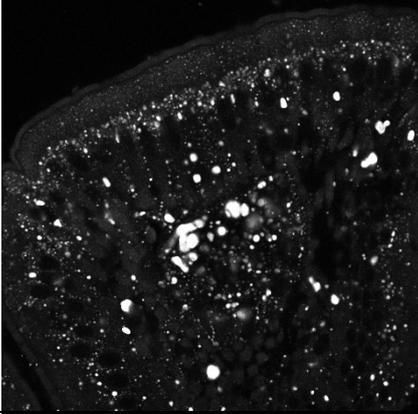
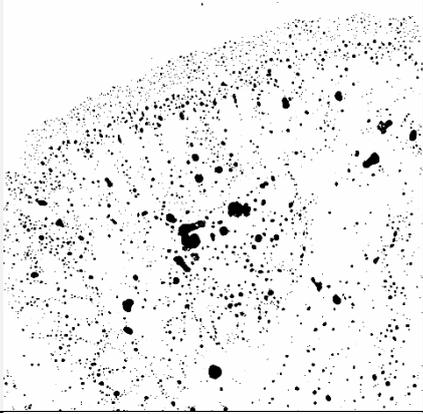
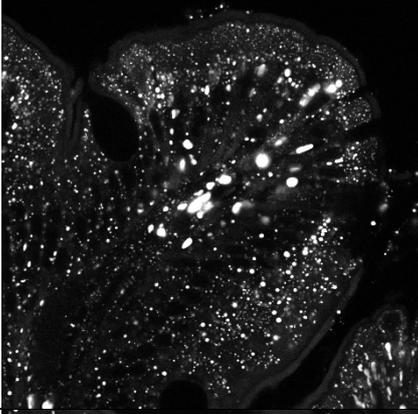
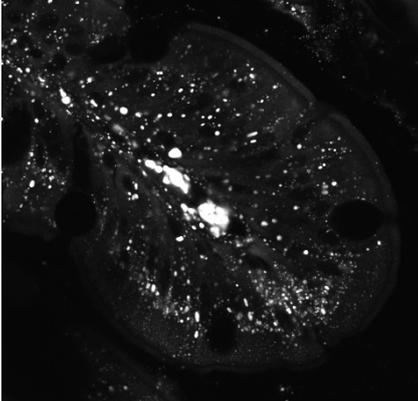
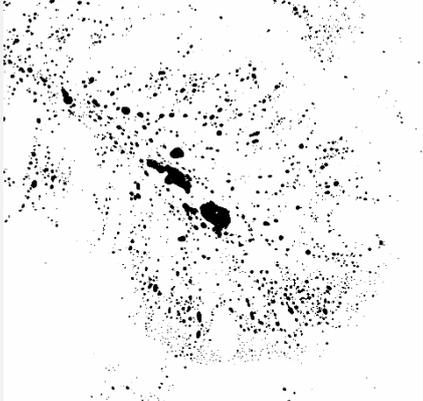
ID number	Albinism	getPCR KO, intestine	getPCR KO, liver
KO16	low	0.183	0.434
KO23	high	0.495	0.511
KO25	medium	0.209	0.326
KO28	high	0.484	0.306
KO29	low	0.139	0.102



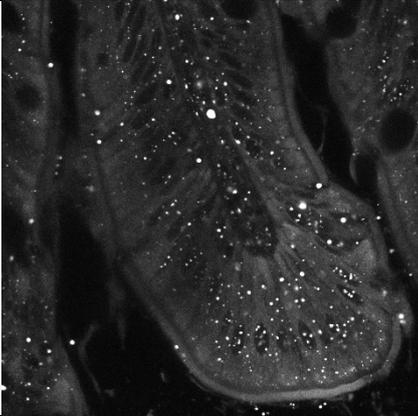
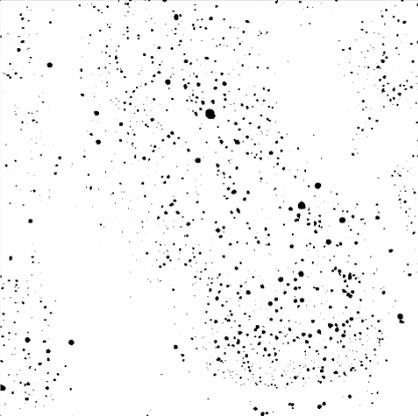
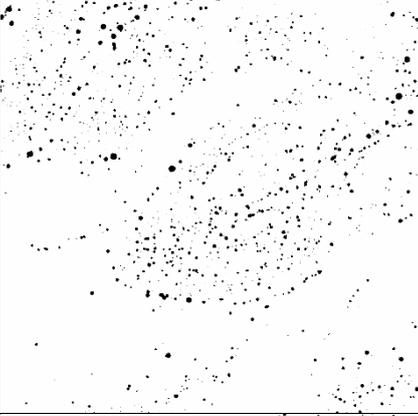
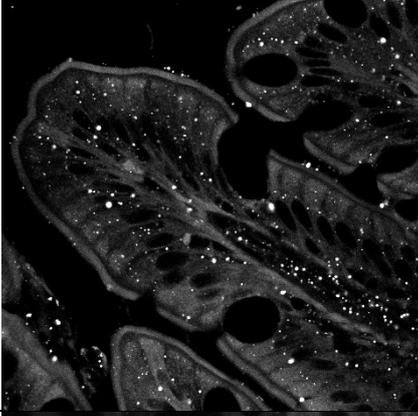
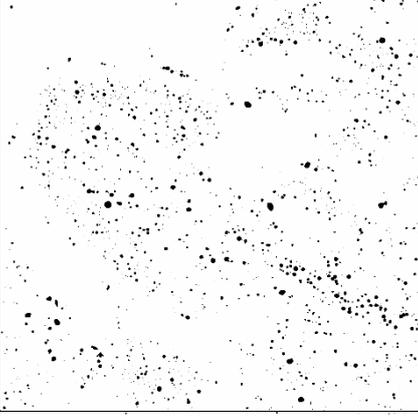
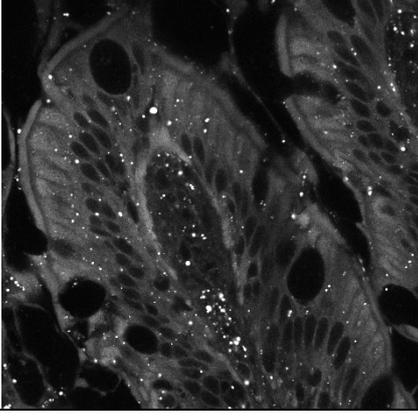
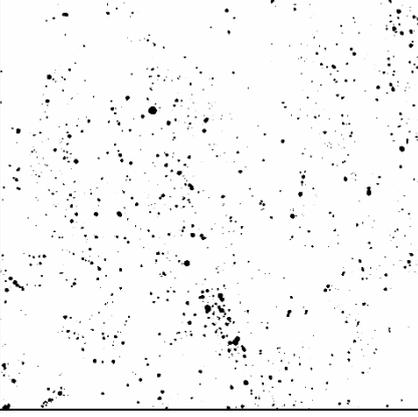
Supplemental figure 4: Lipid droplet measurements collected for the whole group.

(A) Percentage of villus area that is covered by lipids (B) Average number of lipid droplets found per μ m² in the villus

Supplemental table 5: Four example images of lipid histology in Knockout-villi and their resulting binary segmentation.

Microscope image (lipid channel only)	Binary segmentation
	
	
	
	

Supplemental table 6: Four example images of lipid histology in control villi and their resulting binary segmentation.

Microscope image (lipid channel only)	Binary segmentation
	
	
	
	



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitenskapelige universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway