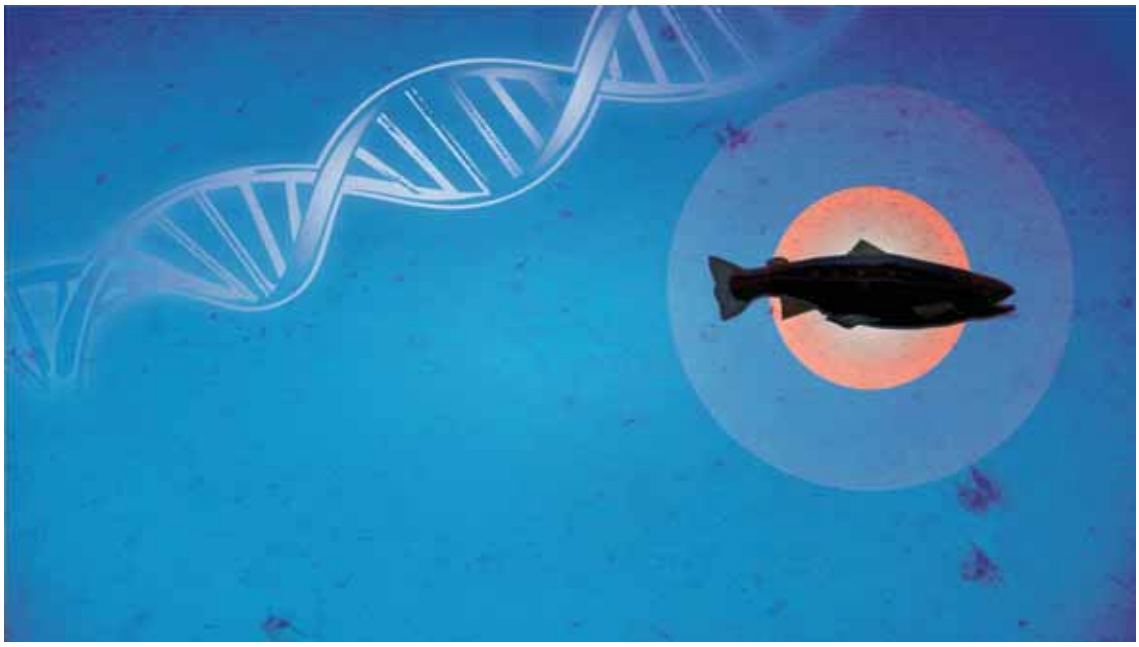
RAPPORT FRA HAVFORSKNINGEN
Nr. 18-2013

MENTOR (NRC project no: 190467)

Mitigating the effects of escaped farmed Atlantic salmon:

Combining single nucleotide polymorphisms, lipid acid profiling, and statistical methods to trace escaped salmon back to origin

By Øystein Skaala (project manager), Francois Besnier and Kevin A. Glover



Juni 2013

www.imr.no

HAVFORSKNINGSINSTITUTTET
INSTITUTE OF MARINE RESEARCH

MENTOR (NRC project no: 190467)
Sluttrapport

Mitigating the effects of escaped farmed Atlantic salmon: combining single nucleotide polymorphisms, lipid acid profiling, and statistical methods to trace escaped salmon back to origin

Øystein Skaala (project manager), Francois Besnier and Kevin A. Glover

Bergen, 8.02.2013

Content

Samandrag

Målet med MENTOR prosjektet er å auka presisjonen i metodar som blir brukt for identifisering av opphavet til rømt oppdettslaks, gjennom ein kombinasjon av SNP markørar, lipidprofilar og statistiske metodar.

Samanlikning mellom 300 polymorfe SNP og 14 STR markørar vart gjennomført på eit datasett med 10 populasjonar og 500 Atlantisk laks. Global Fst varierte frå 0.033 til 0.115 og frå -0.002 til 0.316 for dei 14 STR og dei 300 SNP markørane respektivt. Dei 15 beste SNP (30 allel) markørane gav ein presisjon som tilsvarte dei 4 beste STR (83 allel) markørane. Tillegg av fleire STR markørar gav kun liten auke i presisjon medan tillegg av opptil 100 SNP loci auka presisjonen i identifiseringa.

Eit common garden feltforsøk vart initiert i 2003 i Guddalselva for å samanlikna overleving, vekst mv hos avkom av oppdettslaks, villaks og kryssingar frå rogn til smolt i naturen. Signifikante skilnadar vart registrert i vekst, kroppsfasong (kondisjonsfaktor) og overleving mellom dei 69 familiane og mellom dei tre hovudgruppene. I MENTOR prosjektet har vi gått vidare for å gje ei meir inngåande forståing av det genetiske grunnlaget for desse observerte skilnadane i vekst, kroppsfasong og overleving mellom oppdrettslaks, villaks og kryssingar. Genom scan for QTL som påverkar lengde (L), vekt (W) og kondisjonsfaktor (CF) avdekka fleire genom regionar med signifikant påverknad på desse tre fenotypane. Genom posisjonane for desse QTL'ane og andel av variansen i karakteren er identifisert. Vekt er signifikant påverka av to QTL'ar: ein på kromosom 2, som forklarar 8.4% av observert varians i vekt i populasjonen, og ein på kromosom 11 som forklarar 7.7% av variansen i vekt. Kollektivt er desse to QTL'ane ansvarlege for 14.8% av vekt variansen. QTL posisjonen på kromosom 2 er ein særleg interessant region som påverkar både vekst og overleving i naturen. Interessant er det også at denne genom regionen viser signifikant redusert variasjon i oppdrettsmaterialet, medan variasjonen framleis er høg i villaks materialet. QTL på kromosom 2 framstår difor som ein genom region som potensielt viser fotavtrykk av pågåande domestiserings hos Atlantisk laks. Prosjektet omfattar den første undersøkinga av den genetiske arkitekturen som ligg til grunn for skilnadane som er observert mellom oppdrettslaks og villaks i naturen.

Summary

Comparisons between 300 polymorphic SNPs and 14 short tandem repeats (STRs) were conducted on a data set consisting of approximately 500 Atlantic salmon arranged in 10 samples/populations. Global FST ranged from 0.033-0.115 and -0.002-0.316 for the 14 STR and 300 SNP loci respectively. The best 15 SNPs (30 alleles) gave a similar level of self-assignment to the best 4 STR loci (83 alleles), however, addition of further STR loci did not lead to a notable increase assignment whereas addition of up to 100 SNP loci increased assignment.

In 2003, a common garden experiment was initiated in the Guddal river to evaluate and compare the performance of Farm, Wild and Hybrid Salmon in the river habitat. Significant differences in growth, body shape (condition factor) and mortality were reported between these three types (Skaala et al. 2012). The aims of this two years post-doctoral project was to provide better understanding on the genetic bases for the observed differences in growth, body shape and mortality between wild, farm, and hybrid salmon. Genome scan for QTL affecting length (L), weight (W), and condition factor (CF) revealed several genomic regions significantly affecting these three phenotypes. The genomic positions of these QTL as well as proportion of the trait variance explained are identified.

Weight is significantly affected by two QTL: one on chromosome 2 that explains 8.4% of the observed weight variance in our population, and one on chromosome 11 responsible for 7.7% of observed weight variance. Collectively, these two QTL are responsible for 14.8% of the weight variance.

QTL position on chromosome 2 appears to be a particularly interesting region that both affects growth and survival in the river habitat. Moreover, this genomic region displays a strong reduction in genetic variability in the farm population whereas genetic variability is still very high in the wild population. QTL on chromosome 2 appears thus to be a genomic region potentially bearing the footprint of ongoing domestication in salmon. This study represents the first investigation into the genetic architecture underlying the relative performance of farmed and wild Atlantic salmon in the natural habitat.

Objective

The main objective of the project is to improve the precision in methods for identification of origin of farmed salmon escaping from sea cages and smolt farms, by a combination of SNP markers and lipid acid profiles and adapted current statistical techniques.

The project

Four different work packages have been designed in order to gain the information required to meet the objectives of the project:

1: SNPs for tracing escaped salmon back to farm of origin
2: Quantification of natural selection on SNPs
3: Investigation of new statistical methods for identifying the origin for farmed escaped salmon and detecting offspring of farmed salmon in wild populations
4: Lipid acid profiling as a supplementary identification tool for farmed escaped salmon

1 SNPs for tracing escaped salmon back to farm of origin

Technological advances have lead to the rapid increase in availability of single nucleotide polymorphisms (SNPs) in a range of organisms, and there is a general optimism that SNPs will become the marker of choice for a range of evolutionary applications. Comparisons between 300 polymorphic SNPs and 14 short tandem repeats (STRs) were conducted on a data set consisting of approximately 500 Atlantic salmon arranged in 10 samples/populations. Global FST ranged from 0.033-0.115 and -0.002-0.316 for the 14 STR and 300 SNP loci respectively. Global FST was similar among 28 linkage groups when averaging data from mapped SNPs. With the exception of selecting a panel of SNPs taking the locus displaying the highest global FST for each of the 28 linkage groups, which inflated estimation of genetic differentiation among the samples, inferred genetic relationships were highly similar between

SNP and STR data sets and variants thereof. The best 15 SNPs (30 alleles) gave a similar level of self-assignment to the best 4 STR loci (83 alleles), however, addition of further STR loci did not lead to a notable increase assignment whereas addition of up to 100 SNP loci increased assignment. Whilst the optimal combinations of SNPs identified in this study are linked to the samples from which they were selected, this study demonstrates that identification of highly informative SNP loci from larger panels will provide researchers with a powerful approach to delineate genetic relationships at the individual and population levels.

2 Quantification of natural selection on SNPs

In 2003, a common garden experiment was initiated in the Guddal river to evaluate and compare the performance of Farm, Wild and Hybrid Salmon in the river habitat. Significant differences in growth, body shape (condition factor) and mortality were reported between these three types (Skaala et al. 2012). The aims of this two years post-doctoral project was to provide better understanding on the genetic bases for the observed differences in growth, body shape and mortality between wild, farm, and hybrid salmon.

During the year 2004 and 2005, the eggs from fifty-one full sib families of Atlantic salmon were obtained from either the MOWI farm strain, or from a wild population kept in the Norwegian gene bank (Skaala et al. 2012). In total 167742 eggs from farm, wild or hybrid families were planted in winter 2004 and 2005 in the river Guddal. Every year following the planting of the eggs, a wolf trap was installed on the river before the migration season, and dismounted several weeks after the termination of the smolt migration. All captured smolt were tranquilized, and length (L), weight (W), and condition factor ($CF = 1000 \times \frac{W}{L^3}$) were measured for each fish. After identification of the parentage (Skaala et al. 2012), the two parents and three offspring in each full sib family were genotyped for 5650 SNP present in a chip developed in CIGENE (Lien et al. 2012). The position of each SNP on the salmon genome was determined from the linkage map recently developed on this set of markers (Lien et al. 2012). Out of 5650 SNP, 272 were selected for genotyping all the captured fishes from cohort 2004 and 2005.

Genotypes at 272 markers were used together with family information in a linkage-mapping framework. In short, a mixed linear model with random genetic effect was used to establish statistical correlation between a trait variation (e.g. Growth) and genotype at regular interval along the genome. Genomic regions significantly affecting a trait are thereafter referred to as Quantitative Trait Locus or QTL.

Quantitative traits (length (L), weight (W), and condition factor (CF))
Genome scan for QTL affecting length (L), weight (W), and condition factor (CF) revealed several genomic regions significantly affecting these three phenotypes. The genomic positions of these QTL as well as proportion of the trait variance explained are summarized in Table 1,

Table 1. Position and associated variance of QTL affecting weight, length and cf.

Phenotype	Linkage group	Position confidence interval (cM)	Proportion of phenotype variance (%)	Collective proportion of phenotype variance (%)
W	2	0-90	8.4	14.8
W	11	30-70	7.7	
L	2	0-90	5.6	11.2
L	11	10-70	5.5	
CF	11	10-70	5.7	9.8
CF	20	30-60	4.6	

e.g: Weight is significantly affected by two QTL: one on chromosome 2 that explains 8.4% of the observed weight variance in our population, and one on chromosome 11 responsible for 7.7% of observed weight variance. Collectively, these two QTL are responsible for 14.8% of the weight variance. In a second step, the genetic variability of the parental alleles within type

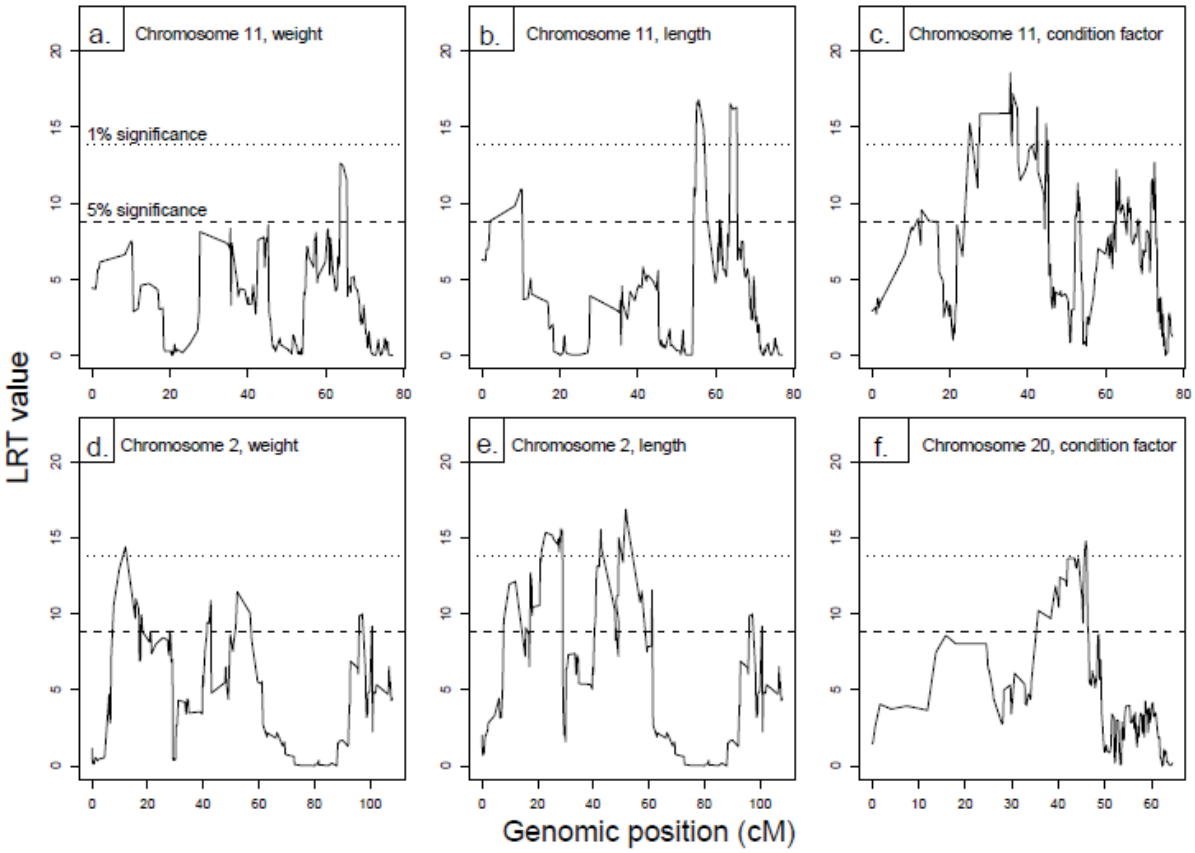
(Wild or Farm) at QTL position revealed that the genetic variability could largely differ among type and among QTL. These results are summarized in Table 2.

Table 2. QTL allelic fixation within type at QTL position.

	Weight		Length		Condition Factor	
Chromosome	**2**	**11**	**2**	**11**	**11**	**20**
WILD	**0.08 (ns)**	**0.71***	**0.01 (ns)**	**0.46 (ns)**	**0.25 (ns)**	**0.53 ***
FARM	**0.92 *****	**0.01 (ns)**	**0.97 *****	**0.01 (ns)**	**0.01 (ns)**	**0.46 ***

*Significance code: ns: p>0.05, *: p<0.05, **: p<0.01, ***: p<0.001*
Correspond to the test H1"allele correlation >0" versus H0 "Allele correlation =0"

On chromosome 2, the QTL affecting weight displays a high genetic diversity for the alleles coming from the wild population (allele correlation = 0.08 (ns)), whereas genetic diversity is much lower for the alleles coming from the farm population (allele correlation = 0.92***. Knowing that farm salmon were selected for about 10 generations for production related traits such as growth, it is likely that the loss of genetic diversity at QTL position reported here are related to domestication by selective breeding in the farm strains.



Genome wide scan based on dense inferred marker genotype at exact marker position for (a) weight on chromosome 11, (b) length on chromosome 11, (c) condition factor on chromosome 11, (d) weight on chromosome 2, (e) length on chromosome 2, (f) condition

factor on chromosome 20. Horizontal dashed line and pointed line indicate the 5% and 1% genome wide significance threshold respectively.

Mortality trait

Mortality is a more delicate phenotype to study than continuous traits *i*- because the trait distribution is binomial (Dead / Alive) whereas statistical methods to correlate genotype and phenotype are adapted for continuous normally distributed traits, *ii*- because dead fishes can not be captured in the experiment, only the genotype of surviving individuals can be accessed, whereas genotype of dead individuals have to be inferred from the parental population.

Computer simulation were used to generate a large number of hypothetic offspring from the parental populations of the experiment. The allelic distribution of the simulated populations was compared to the one observed in the genotyped fishes that were captured in the trap and thus considered as the surviving population. The R package HGLM (Ronnegard et al. 2010) was used to correlate genotype and mortality in our population. One locus on chromosome 2 appeared to be strongly correlated with fish mortality in the river habitat. QTL for mortality seem to be located at the same chromosome position as QTL affecting weight and length.

QTL position on chromosome 2 appears to be a particularly interesting region that both affects growth and survival in the river habitat. Moreover, this genomic region displays a strong reduction in genetic variability in the farm population whereas genetic variability is still very high in the wild population. QTL on chromosome 2 appears thus to be a genomic region potentially bearing the footprint of ongoing domestication in salmon.

Publications

Preliminary results from the quantitative trait study were presented in QTLMAS conference in 2011 in Rennes (France) under the title “QTL detection for survival related traits in atlantic salmon” (https://colloque4.inra.fr/qtlmas/Media/QTLMAS-program). Final results from the quantitative trait study are submitted to the journal “Molecular Ecology”.

3 Investigation of new statistical methods for identifying the origin for farmed escaped salmon and detecting offspring of farmed salmon in wild populations

In this study, >2200 salmon were collected from 44 cages located on 26 farms in the Hardangerfjord, western Norway. This fjord represents one of the major salmon farming areas in Norway. Based upon genetic data from 17 microsatellite markers, significant but highly variable differentiation was observed among the 44 samples (cages), with pair-wise FST values ranging between 0.000 and 0.185. Bayesian clustering of the samples revealed five major genetic groups, into which the 44 samples were re-organised. Bayesian clustering also identified two samples consisting of fish with mixed genetic background. Performing self-assignment simulations with the data divided into different subsets, overall accuracy of assignment was 44% within the entire material (44 samples), 44% for the 28 spring samples, 59% for the 16 autumn samples, and 70% for 8 autumn samples collected from a geographically restricted area. Accuracy of assignment varied greatly among the individual

samples. For the Bayesian clustered data set consisting of five genetic groups, overall accuracy of self-assignment was 99%, demonstrating the effectiveness of this strategy to significantly increase accuracy of assignment, albeit at the expense of precision. This study demonstrates the potential to identify the farm of origin for escapees in a region with a large number of salmon farms. The approaches described here will be of relevance to a range of other species reared in culture where identification of escapees may be required.

A 'DNA stand-by method' for identification of escaped Atlantic salmon, back to the cage and farm of origin, was established. In addition, proof-of-concept for the method has been demonstrated to be able to trace rainbow trout and Atlantic cod escapees back to their farm source. The combined sampling, genotyping and statistical analysis on which the method is based has been implemented successfully in the identification of fish farm escapees in Norway, resulting in fines for companies found in breach of regulations. It is concluded that as the method has been successful for the 3 major species farmed in Norway, each with contrasting production logistics including breeding programs, state of domestication, and magnitude of production, the DNA stand-by method can be applicable to identification of fish farm escapees for a wide range of aquaculture species in all regions of the world.

4 Lipid acid profiling as a supplementary identification tool for farmed escaped salmon

This study represents the first determination of lipids and fatty acids in fish scales. Scales collected from groups of Atlantic salmon reared on fish farms and in experimental tanks were analyzed by chromatography. The complete suite of fatty acids normally found in marine organisms was detected in the scales, with the following fatty acids dominating: 16:0, 18:0, 18:1n9, 20:5n3, 22:6n3 and 24:1n9. Scales contained relatively high levels of furan fatty acids, and the level of cholesterol (2.5–5 mg/g tissue) was much higher than the levels found in the edible parts of marine fishes (0.2–1 mg/g tissue). The fatty acid profile of scales was distinct between groups of salmon originating from different commercial strains reared on the same farm, between salmon groups originating from the same strains but reared at different farms, and between groups of fed and unfed salmon in experimental tanks. Together, these data indicate that the fatty acid composition of fish scales is dependent upon both environmental and genetic factors. The fatty acid composition of fish scales may be used in stock/population identification, for example identification of escaped Atlantic salmon to farm of origin.

Publications

Besnier F, Glover KA, Skaala Ø. 2011. Investigating genetic change in wild populations: modelling gene flow from farm escapees Aquacult Environ Interact Vol. 2: 75–86.

Besnier F, Glover K.A, Lien S, Kent M, Hansen M.M, Skaala Ø. Identifying genomic regions linked to domestication and fitness in Atlantic salmon: a QTL study in a natural habitat. Submitted Mol Ecol.

Glover KA. 2010. Forensic identification of fish farm escapees: the Norwegian experience. Aquacult Environ Interact Vol. 1: 1–10.

Glover, K.A., M M. Hansen , Ø. Skaala. 2009 .Identifying the source of farmed escaped Atlantic salmon (Salmo salar): Bayesian clustering analysis increases accuracy of assignment. Aquaculture 290 (2009) 37–46.

Glover, K.A., M M Hansen, S. Lien, T. D. Als, B. Høyheim, Ø. Skaala. 2010. A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. BMC Genetics 2010, 11:2.

Glover KA, Quintela M, Wennevik V, Besnier F, Sørvik AGE, Skaala Ø. 2012 Three Decades of Farmed Escapees in the Wild: A Spatio-Temporal Analysis of Atlantic Salmon Population Genetic Structure throughout Norway. PLoS ONE 7(8): e43129. doi:10.1371/journal.pone.0043129.

Glover KA, Skaala Ø, Sørvik, AGE & Helle TA. 2010. Genetic differentiation among Atlantic salmon reared in sea-cages reveals a non-random distribution of genetic material from a breeding programme to commercial production. Aquaculture Research, 1-9.

Grahl-Nielsen O, Glover KA. 2010. Fatty acids in fish scales. Mar Biol. DOI 10.1007/s00227-010-1430-8.

Skaala Ø, Glover KA, Barlaup BT, Svåsand T, Besnier F, Hansen MM, Borgstrøm R. 2012. Performance of farm, hybrid and wild Atlantic salmon (Salmo salar) families in a natural river environment. Can. J. Fish. Aquat. Sci. 69: 1994-2006.