MOLECULAR ECOLOGY
RESOURCES

Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip.

Journal:	*Molecular Ecology Resources*
Manuscript ID:	MER-10-0334.R2
Manuscript Type:	Permanent Genetic Resources Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Karlsson, Sten; Nofima Moen, Thomas; Aqua Gen AS Lien, Sigbjørn; Norwegian University of Life Sciences, Center for Integrative Genetics Glover, Kevin; Institute of Marine Research Hindar, Kjetil; Norwegian Institute for Nature Research
Keywords:	Conservation Genetics, Contemporary Evolution, Experimental Evolution, Fish, Aquaculture

Table1. Wild populations (a) and farmed populations (b) of Atlantic salmon and corresponding number of specimens successfully genotyped for 4514 SNPs. Numbers after population name is sampling year for the wild and hatching year for the farmed strains, respectively.

(a)

Population	N
Tana -89	40
Altaelva -81, -82, -95	40
Saltdalselva -77, -78	35
Namsen -78	48
Gaula -89, -90	44
Surna -77	30
Driva -77, -91	41
Rauma -74, -76, -77, -91, -95	48
Lærdalselva -77, -78, -97	61
Vosso -77, -78	18
Suldalslågen -79, -80	50
Figgjo -89	48
Numedalslågen -89	50
Total	**553**

(b)

Population	N
AG -98	48
AG -99	89
AG -00	58
AG -01	291
Mowi -04	20
Mowi -05	20
Mowi -08	20
Mowi -09	20
SB -04	47
SB -05	47
SB -06	48
SB -07	48
Total	**756**

Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip.

Sten Karlsson[1*], Thomas Moen[2, 3], Sigbjørn Lien[3, 4], Kevin A. Glover[5], and Kjetil Hindar[6]

1. Nofima Marine, Arboretveien 6, N-1432 Ås, Norway

2. Aqua Gen AS, P.O. Box 1240, N-7462 Trondheim, Norway

3. Center for Integrative Genetics, Norwegian University of Life Sciences, Arboretveien 6, N-1432 Ås, Norway.

4. Department of Animal and Aquaculture Sciences, Norwegian University of Life Sciences, Arboretveien 6, N-1432 Ås, Norway.

5. Institute of Marine Research, P.O.Box 1870 Nordnes, N-5817 Bergen, Norway

6. Norwegian Institute for Nature Research (NINA), P.O. Box 5685 Sluppen, N-7485 Trondheim, Norway,

*Correspondence: Sten Karlsson, Address: Norwegian Institute for Nature Research (NINA), P.O. Box 5685 Sluppen, N-7485 Trondheim, Norway,

Fax number: +47 64949502, e-mail: sten.karlsson@.nina.no

Running title: Farmed and Wild Atlantic salmon

This is the peer reviewed version of the following article, Karlsson, S. et al 2011 Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. Molecular Ecology Resources 2011 ;Volum 11. Suppl. 1 s. 247-253 , which has been published in final form at 10.1111/j.1755-0998.2010.02959.x .

Abstract

Genetic interactions between farmed and wild conspecifics are of special concern in fisheries where large numbers of domesticated individuals are released into the wild. In the Atlantic salmon (*Salmo salar*), selective breeding since the 1970´s has resulted in rapid genetic changes in commercially important traits, such as a doubling of the growth rate. Each year, farmed salmon escape from net pens, enter rivers, and interbreed with wild salmon. Field experiments demonstrate that genetic introgression may weaken the viability of recipient populations. However, due to the lack of diagnostic genetic markers, little is known about actual rates of gene flow from farmed to wild populations. Here we present a panel of 60 SNPs that collectively are diagnostic in identifying individual salmon as being farmed or wild, regardless of their populations of origin. These were sourced from a pool of 7000 SNPs comparing historical wild and farmed salmon populations, and were distributed on all but two of the 29 chromosomes. We suggest that the generic differences between farmed and wild salmon at these SNPs have arisen due to domestication. The identified panel of SNPs will permit quantification of gene flow from farmed to wild salmon populations, elucidating one of the most controversial potential impacts of aquaculture. With increasing global interest in aquaculture and increasing pressure on wild populations, results from our study have implications for a wide range of species.

Introduction

Rapid genetic improvements from selective breeding are expected in many aquaculture species having high fecundity and large phenotypic and genetic variation (Gjedrem & Baranski 2009). Large-scale Atlantic salmon (*Salmo salar*) breeding programs were

established in the early seventies with broodfish collected from a large set of Norwegian wild populations (Gjedrem *et al*. 1991, Gjøen & Bentsen 1997). Selective breeding of the Atlantic salmon has more than doubled the growth rate in five generations (Thodesen *et al.* 1999), implying a rapid change in the genetic makeup of farmed stocks compared to their wild origin. At the same time, many wild Atlantic salmon populations are at risk from high proportions of escaped farmed salmon (Hindar *et al.* 1991). In some rivers, escaped farmed salmon have outnumbered wild salmon for many years (Fiske *et al.* 2006, Thorstad *et al.* 2008). There is an urgent need for a tool that can be used for monitoring genetic introgression from farmed to wild salmon.

A prerequisite for estimating potential gene flow from farmed escapees to wild populations is that the genetic make-up of farmed and wild salmon is known and that there are sufficient genetic differences between them. Attempts at quantifying gene flow from farmed salmon to wild salmon have been limited to single wild populations receiving farmed salmon from well defined escapes, from which samples could be obtained (Crozier 1993, Clifford *et al.* 1998, Crozier 2000). Attempts have also been made to relate temporal genetic changes in wild salmon populations to farmed salmon escapes (Skaala *et al.* 2006). While these authors were able to document genetic changes in some wild populations affected by escaped farmed salmon, in other rivers where large numbers of escaped farmed salmon were observed, no genetic changes were detected in a panel of neutral microsatellite markers. It is likely that the lack of genetic markers between farmed and wild salmon limited the ability to accurately detect gene flow in some of the populations.

In general, identification of markers that are diagnostic on a farmed-wild boundary requires the availability of DNA markers in linkage disequilibrium (LD) with loci under selection. This requires that large numbers of DNA markers have been characterized, and

that these markers can be genotyped efficiently in large numbers of animals. Recently, a large number of single nucleotide polymorphisms (SNPs) have been identified in Atlantic salmon, and an Illumina 7k SNP-chip has been manufactured (S. Lien, unpublished). This SNP-chip covers a large fraction of the salmon genome, enabling genome-wide search for SNPs that can distinguish farmed from wild salmon. We hypothesized that since the breeding goal in several farmed populations are the same or similar, these strains should evolve in the same direction away from their wild origin. Consequently, some of the genes controlling traits important for aquaculture, or polymorphic genetic markers linked to such genes might therefore display similar changes in allele frequencies across isolated farmed strains. The main goal in this study was to apply the 7K SNP-chip to identify genetic markers for generic distinction between farmed and wild Atlantic salmon, enabling large scale studies of gene-flow from escaped farmed salmon to wild populations.

Materials and methods

Sample collection

Genomic DNA samples were obtained from breeding companies dominating production of farmed Atlantic salmon in Norway, while also covering a substantial fraction of the international market: Aqua Gen (AG), SalmoBreed (SB), and Marine Harvest (Mowi strain). Samples from each of these breeding companies included four year-classes which to varying degrees could be regarded as isolated populations. Genomic DNA from wild salmon (Table 1) was obtained from historical scale samples, to ensure that they represented truly wild salmon, unaffected by farmed salmon. A total of 13 wild populations was sampled, covering the distribution range in Norway (Fig. 1), and included large populations and populations that gave rise to the farmed strains.

SNP genotyping

Samples were genotyped using the 7K Atlantic salmon Illumina SNP-chip (CIGENE). Genotype clustering was performed using the Illumina©GenomeStudio 2008 software. Each SNP locus was inspected manually and clusters were adjusted when appropriate. Individuals with call rates (proportion of SNPs genotyped) < 90% were excluded from further analyses.

Identification of a diagnostic panel of SNPs

Genetic differentiation was measured with fixation index (Weir & Cockerham 1984) (F_{ST}) between pooled samples of wild and pooled samples of farmed salmon, for each locus, using Genepop v.4 (Raymond & Rousset 1995). To ensure reasonable independence between genetic markers, a threshold of an inter-marker distance of 5 centi-Morgan (cM) was chosen on the basis of published data on levels of LD in Atlantic salmon (Moen *et al.* 2008) and a newly developed genetic map including the SNPs used in the present study (Lien *et al*. unpublished). The 200 loci displaying the highest F_{ST} were ranked according to their assignment performance, using BELS (Bromaghin 2008), by arranging the populations in a farmed and a wild reporting group, maximizing mean individual assignment accuracy, re-sampling the baseline populations with 200 fish per population, simulating genotypes, creating 200 individuals per reporting group with equal population size within groups, and by performing permutations with 250 replicates. The method implemented in BELS was preferred because it exploits synergy among loci, while allowing individual assignment to groups of populations rather than to specific populations (Bromaghin 2008).

Test of assignment performance

The 200 SNPs showing the highest F_{ST} between farmed and wild salmon were evaluated by performing individual genetic assignment as of farmed or wild origin in GeneClass2 (Piry *et al*. 2004) using the self-assignment option and the Bayesian method (Rannala & Mountain 1997). Correct assignment was recorded whenever a wild specimen was assigned to any of the wild populations, and a farmed specimen was assigned to any of the farmed strains. These tests were performed with different numbers of loci. The 60 highest ranked SNPs were further evaluated for discrimination between farmed and wild salmon, as well as their simulated hybrids. First generation (F1) hybrids were generated from all pairs of wild and farmed populations (156 pairs) using Hybridlab (Nielsen *et al.* 2006). Individual discrimination to any of these three groups was tested in STRUCTURE ver. 2.3.1 (Pritchard *et al*. 2000), assuming two populations (K=2), with 10 000 repetitions as burn in, and 10 000 repetitions after burn in, and applying the admixture model with no *a priori* information of the origin of the individuals. In STRUCTURE, individuals are assigned probabilistically to populations based on their multi-locus genotypes, to obtain highest possible conformance to Hardy-Weinberg equilibrium and linkage equilibrium, within populations. Consequently, admixed individuals, like the F1-hybrids generated in the present study, are expected to show equal probabilities (or proportion of their genome) of belonging to one or the other of the two assumed populations.

Testing the universal property of the panel of diagnostic SNPs

An equal number of individuals (18) were randomly sampled from each population followed by a random assignment of the populations to one of two groups (wild/farmed). This was done 1000 times, yielding 1000 estimates of F_{ST} for each SNP and allowing us to estimate average F_{ST} and the 95% percentile for comparison with the F_{ST} values between the actual farmed and the wild group at the 4514 loci. F_{ST} was estimated in batch mode,

using Genepop v.4 (Raymond & Rousset 1995). To explore the possibility of introducing a bias when using the same populations for identification of the SNP-panel as those used to test its performance, wild populations were excluded one at a time, and for each exclusion, a new SNP-panel was identified based on the F_{ST} values between the farmed and the wild group. Thirteen (equal to the number of wild populations) new SNP-panels were generated, and for each one of these, the proportion of SNPs overlapping with the original SNP panel was estimated. Furthermore, each of the 13 generated SNP-panels were tested for their performance in GeneClass2 (Piry *et al*. 2004) by assigning individuals from the unsampled wild population which (1) had not been included when identifying the SNP-panel, and (2) was not included in the reference populations when doing the assignment.

SNP id and corresponding NCBI, dbSNP accession number for the diagnostic panel of SNPs may be found in Table S1 (Supplementary Information).

Results

A total of 756 farmed salmon and 553 wild salmon were assayed for genetic variation using the 7K SNP-chip. A total of 4514 SNPs showed reliable genotypes and were included in the analyses (Table 1). The overall genetic differentiation (F_{ST}) among farmed populations was 0.095 and among wild populations 0.038. Genetic differentiation (F_{ST}) between a pool of wild salmon samples and a pool of farmed salmon samples was on average 0.016 across all 4514 SNPs, and 0.075 (range = 0.04 to 0.21) for the 200 loci with the highest F_{ST} and an inter-locus distance of at least 5 centi-Morgan (cM). There was a significant difference between the observed F_{ST} distribution of 4514 SNPs for the wild/farmed grouping and the F_{ST} distributions from random allocation of populations into two groups (Fig. 2). Specifically, the 200 loci that showed the highest F_{ST} values between

the wild and the farmed groups, showed significantly higher F_{ST}-values than the F_{ST}-values obtained for the same loci from random allocation of populations into two groups (Fig. 3).

Individual genetic assignment to farmed or wild origin was very accurate, even with a diagnostic panel of only the top 10 highest ranked SNPs. Precision of assignment increased with number of loci, and was close to 100% using the highest-ranked 60 loci (Fig. 4), located on all but two chromosomes. When whole populations were excluded one at a time from the reference populations, and individuals belonging to each one of these excluded populations were assigned, a very similar result was obtained (Fig. 5), except for three farmed strains (AG98, AG99, and AG00).

To evaluate the possible bias introduced by using the same populations for identification and validation of the SNP-panel, we excluded each wild population one at a time, and identified a new SNP-panel for each exclusion. The proportion of SNPs shared between any of these 13 new SNP-panels and the original SNP-panel varied between 89% and 97%. Furthermore, each of the 13 new SNP-panels were tested for diagnostic power by assigning individuals from the unsampled wild population which (1) had not been included when identifying the SNP-panel, and (2) was not included in the reference populations when doing the assignment. For each SNP-panel, the difference in performance compared to the original SNP-panel was negligible (Fig. S1, Supporting Information).

The 60 highest ranking SNPs were further evaluated for discrimination between farmed and wild salmon, as were their *in silico* generated hybrids. Individual discrimination to any of these three groups was tested by a model-based clustering method implemented in STRUCTURE, assuming two populations. A high discrimination between individual farmed and wild salmon was obtained for all pairs of farmed and wild salmon populations, and also for the hybrids (Fig. 6), with an average proportion of the genome

belonging to one of the two populations being 0.76 - 0.94 for the wild fish, 0.07 - 0.33 for the farmed fish, and 0.40 - 0.68 for their F1 hybrids.

Discussion

We have identified a diagnostic panel of genetic markers that discriminate farmed and wild Atlantic salmon, regardless of their populations of origin. Individual genetic assignment to farmed or wild origin was very accurate and close to 100% using the highest-ranked 60 SNPs. For wild salmon, assignment success was high even when assigning individuals from unsampled wild populations. As Norwegian strains of farmed Atlantic salmon dominate salmon aquaculture worldwide, discrimination between farmed and wild Atlantic salmon is likely to be easier outside Norway where wild Atlantic salmon populations differ from Norwegian populations (Verspoor *et al.* 2007).

For farmed salmon, three out of 12 strains studied were not successfully assigned to the farmed group when these strains were not included in the reference panel. In future studies, this is unlikely to cause major problems since almost all farmed salmon in Norway and most farmed salmon elsewhere (Ferguson *et al*. 2007) originate from the 12 strains included in this study. Specifically, it is unlikely to sample an escaped farmed salmon of a different origin that that included in this study. Nevertheless, in future studies it is important to expand the number of farmed salmon strains to be included in the reference group, so that correct assignment is possible even for escaped farmed salmon having a different origin than those included in this study.

Using STRUCTURE and the panel of 60 SNPs, we obtained a clear separation of all possible pairs of wild and farmed salmon, and their *in silico* generated F1 hybrids. This suggests that first-generation farmed salmon immigrants into wild populations, as well as

first-generation hybrids, can be identified in the wild. This makes it possible to directly estimate levels of gene flow resulting from each spawning event. Moreover, temporal genetic changes at the 60 SNPs may be ascribed to introgression of farmed salmon by comparing historical and current levels of farmed salmon representation in the genomes of wild salmon. Until a complete baseline of farmed strains exists, however, farmed to wild salmon gene flow is likely to be underestimated in situations where unsampled farmed strains contribute to this gene flow.

The generic difference between farmed and wild Atlantic salmon seen at these SNPs likely reflects signatures of selection during the breeding programs and not a common origin of farmed strains. This contention is supported by two observations: First, a common shift in allele frequency level in farmed strains, away from allele frequencies in wild populations (Fig. S2 Supporting Information), indicates a parallel molecular evolution in different farmed strains, likely due to similar breeding goals and similar natural selection to the captive environment. Secondly, the overall genetic differentiation among different farmed strains was higher than that among the wild populations when all SNPs were used, consistent with information of the origin of the farmed strains (Gjedrem et al. 1991, Gjøen & Bentsen 1997). As the markers in the diagnostic panel are located on all but two chromosomes, we demonstrate that genome wide molecular genetic changes may happen after few generations in the domestication process of a new aquaculture species. The role of selection will be pursued in a separate study.

A challenge applicable to the present study is what has been called “high-grading bias” (Anderson 2010). This is a bias introduced when the same individuals are being used for identification of genetic markers for genetic assignment, and for testing the performance of these markers. An optimal approach for validating the genetic assignment

performance of a sub-set of selected loci is to test them on an independent data set, a so called "gold standard" (Waples 2010). An obvious conflict between high-grading bias and gold standard is that, while a gold standard procedure ensures unbiased testing, a split of the data set (leaving data out for independent testing), leaves less data, and hence lower power, for finding the most diagnostic panel. This is of particular importance in the present study, where we wanted to find a diagnostic panel for genetic assignment of individuals to two groups of populations, each having considerably genetic variation between populations within group. In our particular case we included all populations for the identification of the diagnostic panel. The high-grading biases were assessed by doing the exercise of excluding each population one at the time and repeating the procedure for identification of a diagnostic panel of SNPs. Each one of the SNP-sets was tested for individual genetic assignment on individuals from the excluded population. From this we could conclude that the high-grading bias was very small. The reason for the low observed high-grading bias is that exclusion of one population out of a total of 25 populations (13 wild and 12 farmed populations) is likely to only have a minor effect on the estimate of genetic differentiation (F_{ST}) between the wild and the farmed groups. Arguably, an even more important reason for the low high-grading bias is the underlying generic differences between wild and farmed salmon, i.e. the SNPs in the diagnostic panel are not collectively diagnostic by chance, but from parallel evolution in farmed salmon strains.

Atlantic salmon populations worldwide are regarded as threatened by aquaculture, including escaped farmed salmon (Hindar *et al.* 1991, Hutchings 1991, Hindar *et al.* 2006, Ford and Myers 2008, Vøllestad *et al.* 2009). Lower viability of wild populations receiving farmed immigrants has been experimentally demonstrated in whole-river experiments (Fleming *et al.* 2000, McGinnity *et al.* 2003). On the other hand, the low overall fitness of

farmed salmon, including low breeding success (Fleming *et al.* 1996) and reduced survival of offspring (McGinnity *et al.* 2003), tend to limit such gene flow. Here we present the tool needed to quantify gene flow from farmed to wild Atlantic salmon. This paper also illustrates how genome wide studies can be applied to farmed-wild genetic interactions for an increasing number of fish species being developed for aquaculture (Bert 2007, Svåsand *et al.* 2007).

Advances in molecular techniques now make it possible to conduct large scale screening of wild Atlantic salmon population to quantify gene flow from escaped farmed salmon, using the SNP-panel presented in this study. This will bring crucial information to a long lived debate regarding consequences on the genetic integrity of wild salmon populations from genetic introgression of farmed salmon.

Acknowledgements We gratefully acknowledge SALMOBREED, AQUA GEN, MARINE HARVEST (MOWI) for providing samples from farmed strains, NINA, Rådgivende Biologer AS and County Fishery Officers for providing samples from wild populations, Stig Omholt (CIGENE) for contributions to the initiation of the project, Torveig Balstad (NINA), Gunnel Østborg (NINA), Hege Munck (Nofima), Matthew Peter Kent (CIGENE) for technical assistance, Fabian Grammes (Nofima) and Kari Sivertsen (NINA) for generating figures, and Nils Ryman (Stockholm University) and Robin Waples (NOAA Fisheries Service) for valuable comments on an earlier version of this manuscript. This study was funded by the FUGE program of the Research Council of Norway (grant no. 175130).

References

Anderson EC (2010) Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. *Molecular Ecology Resources*, **10**, 701-710.

Bert TM (2007) *Ecological and Genetic Implications of Aquaculture Activities*. Kluwer Academic Publishers, Doordrecht.

Bromaghin JF (2008) BELS: backward elimination locus selection for studies of mixture composition or individual assignment. *Molecular Ecology Resources,* **8**, 568-571.

Clifford SL, McGinnity P, Ferguson A (1998) Genetic changes in Atlantic salmon (*Salmo salar*) populations of Northwest Irish rivers resulting from escapes of adult farm salmon. *Can*adian *Journal Fisheries and Aquatic Sciences,* **55**, 358-363.

Crozier WW (1993) Evidence of genetic interaction between escaped farmed salmon and wild Atlantic salmon (*Salmo salar* L.) in a Northern Irish river. *Aquaculture*. **113**, 19-29.

Crozier WW (2000) Escaped farmed salmon, *Salmo salar* L., in the Glenarm River, Northern Ireland: genetic status of the wild population 7 years on. *Fisheries Management and Ecology,* **7**, 437-446.

Ferguson A, Fleming IA, Hindar K *et al*. (2007) Farm Escapes. In: *The Atlantic Salmon: Genetics, Conservation and Management* (eds Verspoor E, Stradmeyer L, Nielsen JL). P. 363, Blackwell, Oxford.

Fiske P, Lund RA, Hansen LP (2006) Relationships between the frequency of farmed Atlantic salmon, *Salmo salar* L, in wild salmon populations and fish farming activity in Norway, 1989 – 2004. *ICES Journal of Marine Science,* **63**, 1182-1189.

Fleming IA, Jonsson B, Gross MR, Lamberg A (1996) An experimental study of the reproductive behaviour and success of farmed and wild Atlantic salmon (*Salmo salar*). *Journal of Applied Ecology,* **33**, 893-905.

Fleming IA, Hindar K, Mjølnerød ID *et al*. (2000) Lifetime success and interactions of farm salmon invading a native population. *Proceedings of the Royal Society London B.* **267**, 1517-1523.

Ford JS, Myers RA (2008) A global assessment of salmon aquaculture impacts on wild salmonids. *PLOS Biol.* **6**(2): e33. Doi: 10.1371/journal.pbio.0060033.

Gjedrem T, Gjøen HM, Gjerde B (1991) Genetic origin of Norwegian farmed salmon. *Aquaculture*, **98**, 41-50.

Gjedrem T, Baranski M (Eds) (2009). *Selective Breeding in Aquaculture: An Introduction.* Springer, London. 221pp. ISBN: 978-90-481-2773-3 (2009).

Gjøen HM, Bentsen HB (1997) Past, present, and future of genetic improvement in salmon aquaculture. *ICES Journal of Marine Science,* **54**, 1009-1014.

Hindar K, Ryman N, Utter F (1991) Genetic effects of cultured fish on natural fish populations. *Can*adian *Journal Fisheries and Aquatic Sciences,* **48**, 954-957.

Hindar K, Fleming IA, McGinnity P, Diserud O (2006) Genetic and ecological effects of salmon farming on wild salmon: modelling from experimental results. *ICES Journal of Marine Science,* **63**, 1234-1247.

Hutchings JA (1991) The threat of extinction to native populations experiencing spawning intrusions by cultured Atlantic salmon. *Aquaculture*, **98**, 119-132.

McGinnity P, Prodöhl P, Ferguson A *et al.* (2003) Fitness reduction and potential extinction of wild populations of Atlantic salmon, *Salmo salar*, as a result of interaction with escaped farm salmon. *Proceedings of the Royal Society London B,* **270**, 2443-2520.

Moen T, Hayes B, Baranski M *et al*. (2008) A linkage map of the Atlantic salmon (*Salmo salar*) based on EST-derived SNP markers. *BMC Genomics* **9,** 223.

Nielsen EE, Bach LA, Kotlicki P (2006) HYBRIDLAB (version 1.9): a program for generating simulated hybrids from population samples. *Molecular Ecology Notes,* **6**, 971-973.

Piry, S, Alapetite A, Cornuet JM *et al*. (2004) GeneClass2: A software for genetic assignment and first-generation migrant detection. *Journal of Heredity,* **95**, 536-539.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics,* **155**, 945-959.

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Science of the United States of America,* **94**, 9197-9201.

Raymond M, Rousset F (1995) Genepop (version 2.1): Population genetics software for exact tests and ecumenicism. *Journal of Heredity,* **86**, 248-249.

Skaala Ø, Wennevik V, Glover KA (2006) Evidence of temporal genetic change in wild Atlantic salmon, *Salmo salar* L., populations affected by farm escapees. *ICES Journal of Marine Science,* **63**, 1224-1233.

Svåsand T, Crosetti D, García-Vázquez E, Verspoor E (2007) *Genetic impact of aquaculture activities on native populations.* Genimpact final scientific report. EU contract no. RICA-CT-2005-022802). 176 pp. http://genimpact.imr.no/.

Thodesen J, Grisdale-Helland B, Helland, SJ, Gjerde B (1999) Feed intake, growth and feed utilization of offspring from wild and selected Atlantic salmon (*Salmo salar*). *Aquaculture*, **180**, 237-246.

Thorstad EB, Økland F, Aarestrup K, Heggberget TG (2008) Factors affecting the within-river spawning migration of Atlantic salmon, with emphasis on human impacts. *Reviews of Fish Biology and Fisheries*, **18**, 345-371.

Verspoor E, Stradmeyer L, Nielsen JL (2007). *The Atlantic Salmon: Genetics, Conservation and Management*. pp. 500, Blackwell, Oxford.

Vøllestad LA, Hirst D, L'Abée-Lund JH *et al.* (2009) Divergent trends in anadromous salmonid populations in Norwegian and Scottish rivers *Proceedings of the Royal Society London B,***276**, 1021-1027.

Waples RS (2010) High-grading bias: subtle problems with assessing power of selected subsets of loci for population assignment. *Molecular Ecology*, in press.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution,* **38**, 1358-1370.

368 **Figure Legends**

369 **Fig. 1** Map of Norway showing the sample sites of wild Atlantic salmon. For year of
370 sampling and sample sizes please see Table 1.

371 **Fig. 2** Observed (solid black line) F_{ST} distribution among 4514 loci between a pool of wild
372 and a pool of farmed Atlantic salmon, and the average (filled black) and upper 95%
373 percentile (filled grey) of F_{ST} estimated from a random allocation of populations into two
374 groups (1000 replicates). For the observed and simulated data, the loci are ordered from
375 largest to smallest F_{ST} value.

376 **Fig. 3** Observed F_{ST} (Obs) between a pool of wild and a pool of farmed Atlantic salmon for
377 each of 200 loci identified as discriminatory for wild and farmed salmon, and the average
378 (Mean) and upper 95% percentile (95% Upper) of F_{ST} estimated from a random sampling
379 of populations into two groups (1000 repetitions).

380 **Fig. 4** Individual genetic assignment of Atlantic salmon as of farmed or wild origin using
381 13 wild and 12 farmed populations. Proportions of correctly assigned individuals to farmed
382 or wild origin are plotted for different numbers of SNP loci and for each population.

383 **Fig. 5** Individual genetic assignment of Atlantic salmon as of farmed or wild origin using
384 13 wild and 12 farmed populations. Whole populations were excluded one at a time from
385 the reference populations, and individuals belonging to each one of these excluded
386 populations were assigned. Proportions of correctly assigned individuals are plotted for
387 different number of loci used and for each population.

388

389 **Fig. 6** Average proportion of genome membership for each pair of farmed and wild
390 populations and their hybrids, assuming two populations and applying the admixture model
391 in STRUCTURE. Each dot represents either wild ("Wild" column), hybrids ("F1 hybrid"
392 column), or farmed ("Farm" column) fish from each of 156 pairs of Farmed and wild
393 populations, and their hybrids.

394

395

Figure 1.

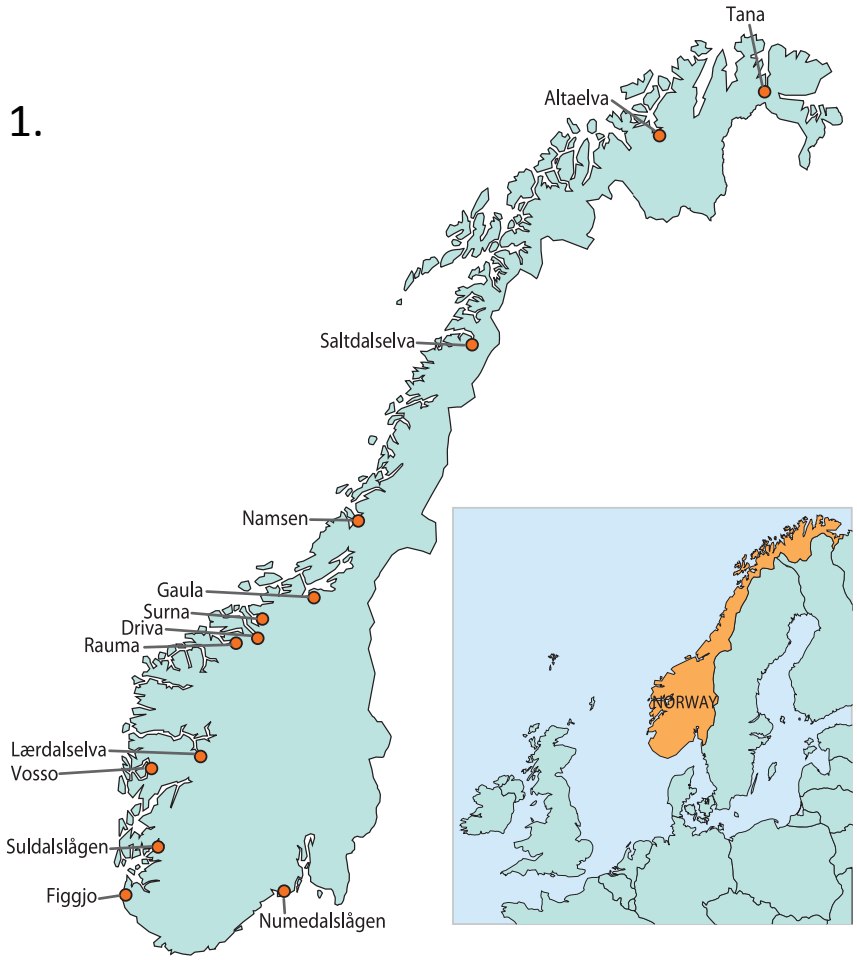

Figure 2.

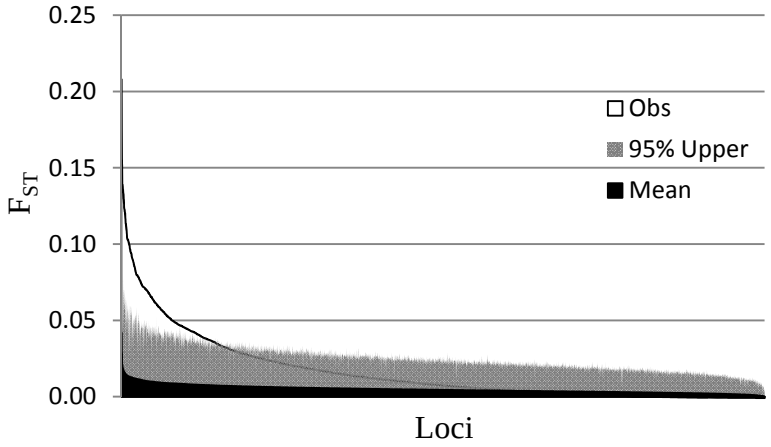

Figure 3.

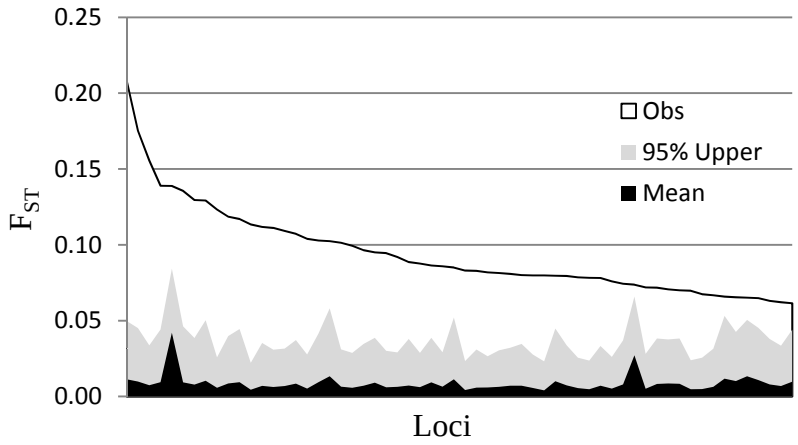

Figure 4.

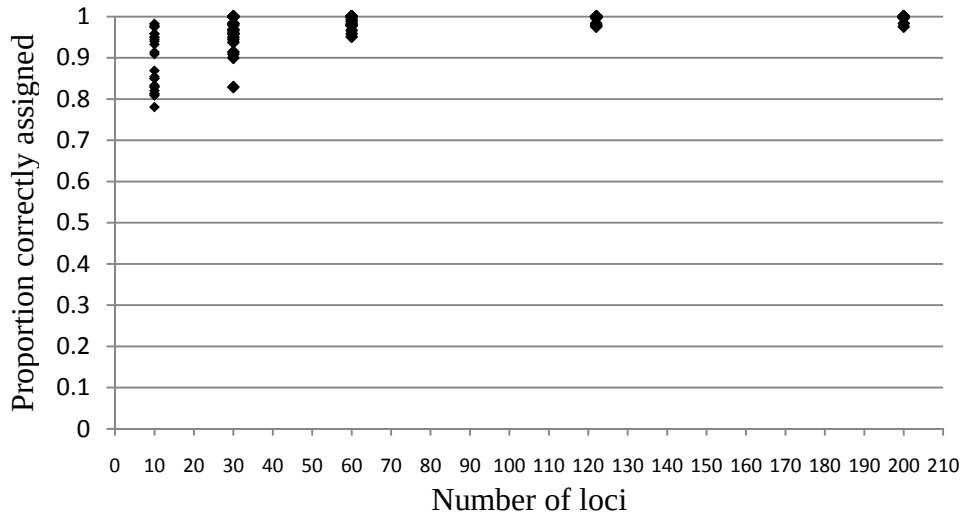

Figure 5.

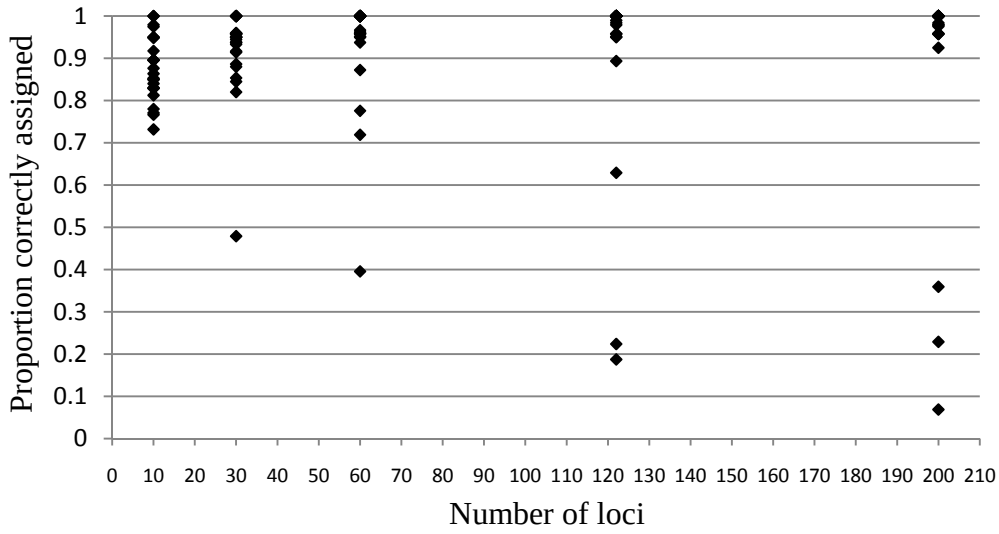

Figure 6.