npg

ORIGINAL ARTICLE

Identification of quantitative genetic components of fitness variation in farmed, hybrid and native salmon in the wild

F Besnier[1], KA Glover[1], S Lien[2], M Kent[2], MM Hansen[1,3], X Shen[4,5] and Ø Skaala[1]

Feral animals represent an important problem in many ecosystems due to interbreeding with wild conspecifics. Hybrid offspring from wild and domestic parents are often less adapted to local environment and ultimately, can reduce the fitness of the native population. This problem is an important concern in Norway, where each year, hundreds of thousands of farm Atlantic salmon escape from fish farms. Feral fish outnumber wild populations, leading to a possible loss of local adaptive genetic variation and erosion of genetic structure in wild populations. Studying the genetic factors underlying relative performance between wild and domesticated conspecific can help to better understand how domestication modifies the genetic background of populations, and how it may alter their ability to adapt to the natural environment. Here, based upon a large-scale release of wild, farm and wild x farm salmon crosses into a natural river system, a genome-wide quantitative trait locus (QTL) scan was performed on the offspring of 50 full-sib families, for traits related to fitness (length, weight, condition factor and survival). Six QTLs were detected as significant contributors to the phenotypic variation of the first three traits, explaining collectively between 9.8 and 14.8% of the phenotypic variation. The seventh QTL had a significant contribution to the variation in survival, and is regarded as a key factor to understand the fitness variability observed among salmon in the river. Interestingly, strong allelic correlation within one of the QTL regions in farmed salmon might reflect a recent selective sweep due to artificial selection.
Heredity (2015) 115, 47–55; doi:10.1038/hdy.2015.15; published online 18 March 2015

INTRODUCTION

Artificial selection in captive wild animals represents an extreme example of evolutionary response to directional selection and therefore, has long been used as a model to understand the mechanisms behind evolution (Darwin, 1868; Lewin, 2009). Animal domestication defined here as an evolutionary process involving the genotypic adaptation of animals to the captive environment (Price and King, 1968) is the result of both selection for desired traits, relaxation of selection for non-desired traits and non-selective forces, such as genetic drift. A side effect of animal domestication is often a loss of adaptive variation that is important for persistence in wild habitats. Various phenotypic alterations of the wild type make most domestic animals less fit in the natural environment than their wild counterparts. This is, for example, the case of body shape (Pulcini et al., 2013), behavior (Price, 1999) or coat coloration (Belyaev et al., 1981), linked to mimicry in the wild. Alteration of coat coloration in domestic animals is also often linked with pathogenic alleles (Webb and Cullen, 2010). However, for a variety of organisms, wild populations are regularly exposed to gene flow from domesticated individuals, either through accidental escapes (Clifford et al., 1998), or voluntarily introduced as part of large-scale translocation efforts (Fischer and Lindenmayer, 2000), or restocking (Koljonen et al., 2002).

The Atlantic salmon (Salmo salar L.) represents an outstanding example of large-scale accidental escapes of farmed strains that subsequently interbreed with wild populations (Clifford et al., 1998;

McGinnity et al., 2003). In Norway, a study of 22 rivers revealed that the amount of genetic introgression from farm salmon in wild populations was between 2% in the most preserved rivers and 47% in the most affected ones (Glover et al., 2013). Wild Atlantic salmon populations are genetically structured (Verspoor et al., 2005), and may be locally adapted to the environments that they inhabit (Garcia de Leaniz et al., 2007). However, in Norway, many populations have been exposed to farm escapees during the past 40 years, and the introduction of farm fish in those rivers represents a conservation problem through the risk of erosion of the genetic structure, loss of local adaptive genetic variation or total replacement of the wild population by domesticated strains (Rhymer and Simberloff, 1996; Allendorf et al., 2001). Simultaneously, farmed salmon have undergone 40 years (approximately 10 generations) of selective breeding for production-related traits, and now differ from wild salmon in many aspects such as allelic frequencies at molecular markers (Skaala et al., 2004), behavior (Einum and Fleming, 1997; Fleming and Einum, 1997), growth (Solberg et al., 2013) and gene expression (Debes et al., 2012). Experimental studies in rivers have revealed that offspring of farmed salmon display significantly reduced survival when compared with offspring of native wild salmon (McGinnity et al., 1997; Skaala et al., 2012). In addition, important differences in other life history traits such as growth or age at maturity have been demonstrated between farmed and wild salmon (McGinnity et al., 2003).

[1]Department of Population Genetics and Ecology, Institute of Marine Research, Bergen, Norway; [2]Department of Animal and Aquacultural Sciences, Centre for Integrative Genetics, Norwegian University of Life Sciences, Ås, Norway; [3]Department of Bioscience, Aarhus University, Aarhus C, Denmark; [4]Department of Medical Epidemiology and Biostatistics (MEB), Karolinska Institutet, Stockholm, Sweden and [5]MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine (IGMM), University of Edinburgh, Edinburgh, UK
Correspondence: Dr F Besnier, Department of Population Genetics and Ecology, Institute of Marine Research, Postbox 1870 Nordnes, Bergen 5817, Norway.
E-mail: francois.besnier@imr.no

Genetic differences between farm and wild Atlantic salmon have been investigated at various levels: (i) allele frequencies of molecular genetic markers to establish a diagnostic identification tool (Karlsson *et al.*, 2011), (ii) among markers under selection and selective sweeps, that is, reduction in genetic variation in DNA that surrounds a locus that is under strong directional selection, to assess the footprint of selection in commercial populations (Martinez *et al.*, 2013) or (iii) by identifying genomic regions associated with phenotypic traits under selection in hatchery condition (Houston *et al.*, 2009; Baranski *et al.*, 2010). While there is a considerable interest in investigating the genetic architecture of fitness in salmonids (Carlson and Seamons, 2008), this latter aspect remains poorly documented. A better documentation of genetic architecture of fitness of wild and farm salmons in the natural habitat would provide crucial information to predict the long-term evolutionary trajectories of wild populations invaded by farm escapees. In such context, combining approaches (ii) and (iii) could provide a more powerful approach than each method separately. Commercial salmon strains were indeed actively selected for a set of traits, and genes controlling these traits are expected to display both association with the phenotypic variance of the trait, and bear the footprint of directional selection. For this reason, we propose here to combine a quantitative trait locus (QTL) scan with a search for loss of genetic diversity.

QTL studies encompass a set of genetic and statistical approaches aiming to identify genomic regions associated with the phenotypic variation of complex traits. However, genotype-by-environment (G × E) interaction is an important factor when detecting QTLs. Ideally, QTLs should be identified in a relevant environmental setting, which for farmed-wild salmon interactions would mean identifying QTLs in a natural environment. To date, most QTL detection studies in salmonid fishes have taken place in captive environments (Reid *et al.*, 2005; Baranski *et al.*, 2010). Investigation of the genetic architecture of a given phenotype in natural environments is, however feasible, either through genome-wide association studies or through variance component QTL mapping. Most such studies have focused on intensively monitored and pedigreed populations (Slate *et al.*, 2002), though examples of experimental crosses conducted in natural environments and opportunistic use of admixed populations for admixture mapping also appear in the literature (Malek *et al.*, 2012).

In the present study, we used a variance component mapping approach to identify QTLs for mortality rate in the river, and three other important fitness-related morphological traits that diverge between wild and farmed Atlantic salmon, that is, length, weight and condition factor (CF). Those traits are related to growth and body conformation. They are expected to be differentially selected for in hatchery and river conditions, and thus, provide a suitable proxy for some component of the fitness. We made use of an experiment where fertilized eggs from wild, farmed and wild x farmed crosses were transplanted into a wild environment as part of a large-scale common garden experiment (Skaala *et al.*, 2012). Our aim was to address the following questions: (i) Can we identify QTL regions for the aforementioned traits?; (ii) Do the QTL regions identified in a natural environment correspond to QTL for the same traits previously identified in hatchery environments?; (iii) Does strong selection at the traits in farmed salmon manifest itself at the genomic level, for example, as loss of variation around QTLs due to selective sweeps?

MATERIALS AND METHODS
Study area and experimental populations
The Guddal, river, which is the location of the present study, drains into the middle region of the Hardangerfjord on the west coast of Norway. The length

of the river available for the anadromous species, Atlantic salmon and sea trout (*Salmo trutta* L.), is approximately 2 km, from the sea up to a waterfall (Liarefossen), which acts as a barrier for ascending fish. Above the waterfall, resident brown trout is the only naturally occurring fish species. The river supports a downstream wolf trap (Millis, 1991), which permits the capture of most smolts migrating from the river into the sea as part of their anadromous lifestyle.

In 2004 and 2005, a total of 167 742 eggs from respectively 20 farmed, 14 wild and 17 hybrid families were planted above the waterfall, into the non-anadromous area (15 000 m^2) of the river. Parents of the hybrid families were sampled from the parents of the farm and wild families. This was part of a long-term study investigating survival in a natural habitat (Skaala *et al.*, 2012). The farmed fish were provided from Marine Harvest fish farm, from a widely used domesticated Norwegian strain (Mowi), while the wild salmon eggs were from the Lærdal river, also from the west of Norway. Despite not being a natural population for the Guddal river, salmons from the Lærdal population are expected to perform better in this environment than a commercial strain that had been subject to approximately 10 generations of domestication selection in the absence of predation and competition for resources. Detail regarding crosses and family data are given in Supplementary Table 1. Earlier studies have demonstrated that this farmed strain displays significantly higher growth rates compared with wild salmon under identical conditions (Solberg *et al.*, 2013), and reduced survival in the Guddal river itself when compared with the river Lærdal population (Skaala *et al.*, 2012).

Genotype and phenotype
In Atlantic salmon, smolt migration (migration from fresh water to the sea) typically involves juvenile fishes of 2–4 years after hatching. This migration is seasonal and occurs in spring, between April and May in the study region. Smolts migrating from the river Guddal in the years following egg planting were captured in the downstream trap. All captured smolt on the trap were tranquilized, and length (*L*), weight (*W*) and condition factor $CF = \frac{1000 \times W}{L}$ were measured for each fish. Mean ± s.d. of each phenotype is reported for each fish type in Table 1. Part of the adipose fin was cut and stored in a 2-ml tube with absolute ethanol for DNA extraction. Upon recovery, smolts were released in the river below the trap. In total, 2411 smolts, respectively 802 from cohorts 2004 and 1609 from cohort 2005 were captured in the trap in the five years following egg planting.

Parentage identification was obtained by genotyping parents and captured offspring using four microsatellite markers Ssa85, Ssa202, Ssa197 (O'Reilly *et al.*, 1996) and SSOSL85 (Slettan *et al.*, 1995). Using the family assignment program (Taggart, 2007), the four markers permitted > 95% of the offspring to be unambiguously identified to their family of origin. After identification of the parentage (Skaala *et al.*, 2012), all stocked families had surviving smolts that were recaptured. The average sample size (number of re-captured smolts) per family was 47 ± 25, with family survival rates varying from 0.5 to 6% of the planted eggs.

The genotyping was performed on an Illumina assay (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. Initially, the parents of each full sib family were genotyped for the 5650 SNPs available on the salmon 6 K SNP chip (Lien *et al.*, 2011; Bourret *et al.*, 2013). The position of each SNP on the salmon genome was determined from the linkage map developed on this set of markers (Lien *et al.*, 2011). From the initial 5650 SNPs genotyped, 272 were selected for genotyping all the captured fishes from cohorts 2004 and 2005

Table 1 Summary statistics (mean ± s.d.) for each of the three studied phenotypes and for the three types

	Length (mm)	Weight (g)	Condition factor
Farmed	145.0 ± 10.6	25.3 ± 5.2	0.82 ± 0.07
Hybrid	140.3 ± 10.5	23.0 ± 4.9	0.83 ± 0.07
Wild	140.2 ± 11.7	22.9 ± 5.3	0.82 ± 0.08
total	140.9 ± 15.2	23.5 ± 5.5	0.82 ± 0.09

(Supplementary Table 2). The selection of those 272 SNP was conducted to optimize information content for QTL mapping, and was based on both the marker position on the salmon genome and the allelic frequencies in the parental generation. SNPs were selected for providing genotype information at regular intervals of 20–30 CM in the female recombination map, and for being polymorphic within full sib families, with, when possible, heterozygous parental genotypes in each family.

After genotyping all captured offspring at 272 SNP markers, the family assignment initially established from microsatellite data was verified based on the SNP information. The family assignments from the two marker sets were concordant with more than 99% overlap.

QTL analysis

A genome-wide scan for QTL was performed via a two-step variance component method (George et al., 2000). The QTL scan was based on the marker information from the 272 SNPs typed in all individuals. A polygenic model with fixed cohort, smolt age effects and random polygenic effects (null model) was compared with a model including a random QTL effect (model 1),

$$y = X\beta + a + e \ \text{(null model)}$$

$$y = X\beta + a + v + e \ \text{(model 1)}$$

where y is a vector of individual phenotypes (length $n = 2411$), β is a vector of fixed effects accounting for cohort (2004 or 2005) and smolt age, X is the corresponding design matrix, a is a vector of random polygenic effects (length n), v is a vector of random QTL effect (length n), and e is a vector of residuals (length n).

The variance covariance matrix for the random QTL effect was estimated via a deterministic method (Pong-Wong et al., 2001) as the Identity By Descent (IBD) matrix at each genomic position tested for QTL.

The parameters of each model were estimated by an average information restricted maximum likelihood algorithm (Johnson and Thompson, 1995) as implemented by Rönnegård and Carlborg (2007) in the R programming environment (Team, 2013)

The presence of QTLs at each genomic location was tested by LRT (Likelihood Ratio Tests):

$$\text{LRT} = -2(\text{llh}_0 - \text{llh}_1)$$

where llh_0 and llh_1 are the log-likelihood of the polygenic and QTL model, respectively.

The significance threshold was determined by permutation (Churchill and Doerge, 1994). Ten thousand permutations were performed at 5 centimorgan (cM) intervals on the genome. The distribution of the maximum test statistic over all analysis points for each of the 10 000 shuffled analyses represents an empirical distribution of the test statistic under the null hypothesis of absence of QTL. Our 5 and 1% significance values were thus set as the 95 and 99 percentile of the empirical distribution of the test statistic under the null hypothesis (Churchill and Doerge, 1994). When two QTLs were significantly associated with the same trait separately, a two-QTL model was fitted to determine the proportion of phenotypic variance collectively explained by the two QTLs.

Estimation of allele fixation within type

The variance component approach that was used for QTL detection assumed that the alleles present in the parental generation are independent, that is, inherited from a different ancestor. This hypothesis is often adopted by default in variance component QTL mapping, as it is conservative with respect to the power to detect significant QTLs. However, in the present experiment, it is likely that within farmed and wild populations, some parents will be related and share some alleles from a common ancestor.

To investigate possible degree of allelic fixation within types, we used an extension of the variance component approach where the alleles within each types are correlated by a coefficient ρ to be estimated. For each detected QTL, the flexible intercross analysis method (Ronnegard et al., 2008) was employed to estimate potential allelic fixation within type ρ, that is, the probability that all parental alleles within wild or within farmed population are inherited from a common ancestor. This method relies on a different estimation of the IBD

matrices, assuming complete fixation of the alleles within each type. First, IBD matrix Π was re-calculated at each QTL position assuming fixation of the alleles within farmed type Π_F or within wild-type Π_W. Matrices were estimated from the genotype information of the 2411 individuals, then, a variance component model was fitted as

$$y = X\beta + a + v + u + e \ \text{(model 2)}$$

where u is a third random variable accounting for the effect of correlation between alleles from the same type, and where the phenotypic variance can be decomposed as:

$$\text{var}(y) = A\sigma_a^2 + \Pi\sigma_v^2 + (\Pi_W - \Pi)\sigma_u^2 + I\sigma_e^2 \ \text{ or}$$

$$\text{var}(y) = A\sigma_a^2 + \Pi\sigma_v^2 + (\Pi_F - \Pi)\sigma_u^2 + I\sigma_e^2$$

Correlation of the alleles within farmed type (ρ_F) or within wild type (ρ_W) was then estimated as $\rho = \frac{\sigma_u^2}{\sigma_v^2}$ (Ronnegard et al., 2008). The alternative hypothesis of allelic correlation (ρ_F or ρ_W) being greater than zero was tested by likelihood ratio between model 1 and model 2. The likelihood ratio was then compared with a Chi-square distribution with one degree of freedom, which corresponds to the difference in degrees of freedom between model 1 and model 2. We then performed a one-tailed test, that is, testing whether the likelihood of model 2 was greater than likelihood of model 1.

Scan for genomic regions associated with mortality

Previous studies of the same experimental population revealed significant differences in survival among half sibs (Skaala et al., 2012), suggesting a genetic contribution to survival between egg and smolt stage. Therefore, here, we also searched for genomic regions associated with survival between egg stage and smolt migration, where the mortality rate varied between 96 and 99% among families. Because the genotype of non-surviving individual was not accessible, an additional set of individuals were simulated by gene dropping from the parental genotypes. For each sampled fish that had survived and was genotyped, a full sibling was simulated by sampling alleles from the parental gametes. Recombinations were simulated according to the linkage map in Atlantic salmon (Lien et al., 2011), and the simulated full sibling was assigned a surviving status (dead or alive) according to the family survival rate (between 4 and 1%). The resulting data were then a set of 2411 observed individuals that had all survived until smolt stage, and a set of 2411 simulated individuals with surviving proportion equal to their family survival rate. The second set was thus a neutral data set assumed to represent a realization of the F1 population if no selective mortality occurs in the river. Simulated and observed data were then merged to crate a 4822 individuals data set, and the influence of genotype on survival was investigated by fitting a hierarchical generalized linear model at each of the 272 genotyped marker positions with the R package hglm (Ronnegard et al., 2010). At each genomic position, a binomial hierarchical generalized linear model was fitted with survival status (dead or alive) as binary response to one fixed full sib family effect and one random genetic effect. The dispersion coefficient of the random genetic effect was kept as an indicator of the genetic contribution to the variance in survival. The incidence matrix Z for the genetic effect was built in a similar manner as the IBD matrix calculated for QTL analyses, and consisted of the first N columns of the gametic IBD matrix, where N is the number of gametes in the parental generation. For genomic regions significantly associated with mortality, a five-markers haplotype window was reconstructed, and the number of parental haplotypes observed in the F1 population was compared with those expected under the neutral hypothesis of no selective mortality in the river via chi-square test.

RESULTS

QTL scan

Based on the information from 272 SNPs, a genome-wide QTL scan was conducted at 1 CM intervals for three phenotypic traits: body weight (W), body length (L) and CF. Permutation testing indicated that 5 and 1% significance threshold were achieved with an LRT value of 8.8 and 13.8, respectively.

The results from the three QTL scans are represented in Figures 1–3 for W, L and CF, respectively. One QTL was significant at 1%
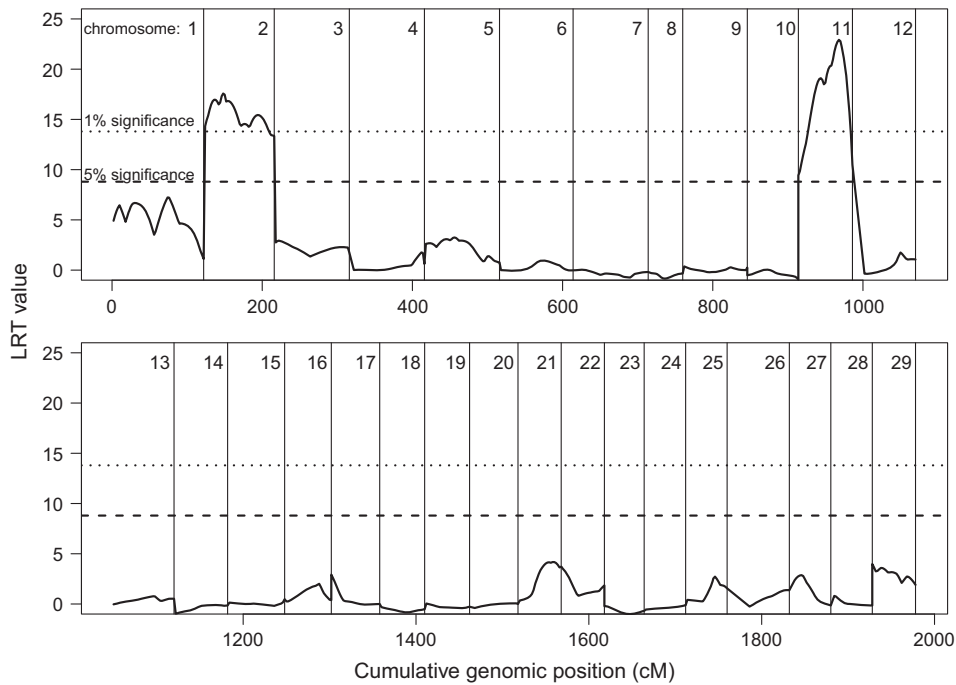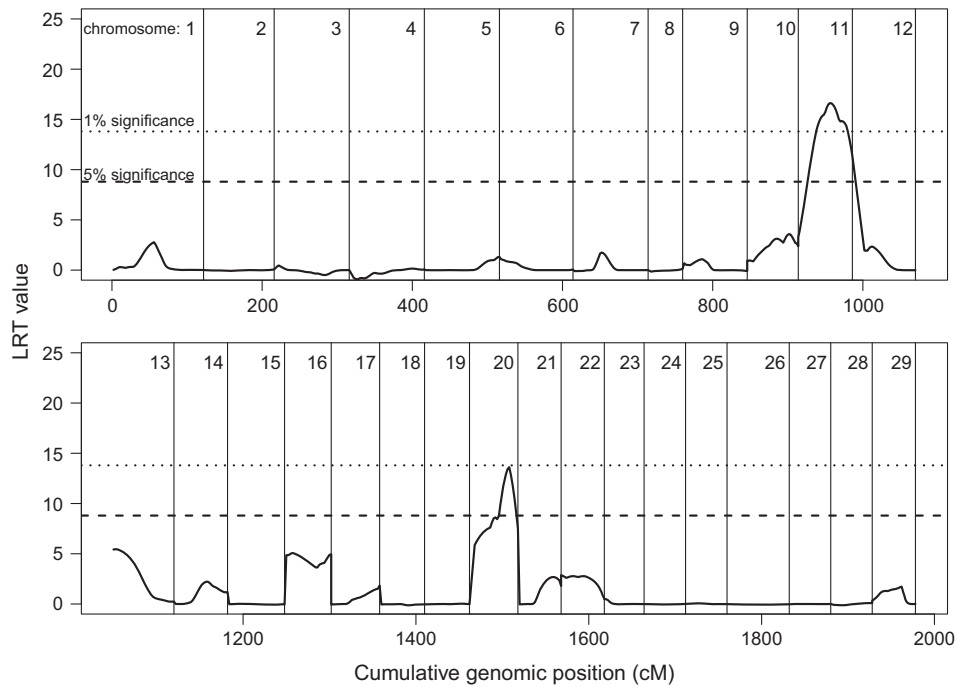
Figure 1 Genome wide scan every 1 cm of the female recombination map for QTL affecting length. Horizontal dashed line and pointed line indicate the 5 and 1% genome-wide significance threshold, respectively. Vertical lines separate chromosomes.
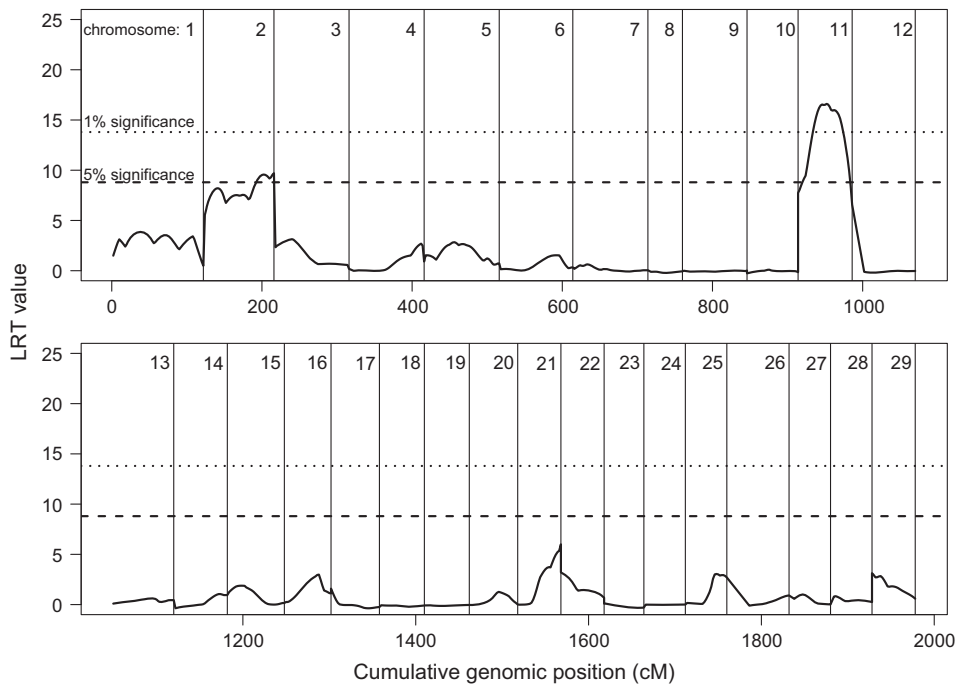
Figure 2 Genome-wide scan every 1 cm of the female recombination map for QTL affecting weight. Horizontal dashed line and pointed line indicate the 5 and 1% genome-wide significance threshold, respectively. Vertical lines separate chromosomes.

threshold on chromosome 11, and was highly correlated with all the three phenotypes (Figures 1–3). A second QTL on chromosome 2 was also highly correlated with W and L (Figures 1 and 2), matching 1% significance threshold for length (Figure 1), and 5% significance for body weight (Figure 2). Additionally, QTL scan for CF revealed one original QTL on chromosome 20 that matched 1% significance

threshold. This last QTL was not linked to W or L individually (Figure 3). Two QTLs situated on chromosomes 2 and 11 (Figures 1 and 2), explained respectively 8.4 and 7.7% of the length variance, and 5.6 and 5.5% of the weight variance (Table 2). Two CF QTLs were detected on chromosomes 11 and 20 (Figure 3), and explained 5.7 and 4.6% of the variance, respectively (Table 2).

Figure 3 Genome-wide scan every 1 cM of the female recombination map for QTL affecting CF. Horizontal dashed line and pointed line indicate the 5 and 1% genome-wide significance threshold, respectively. Vertical lines separate chromosomes.

Table 2 Position and associated variance of QTL affecting weight (W), length (L) and condition factor (CF)

Phenotype	Linkage group	Position confidence interval (cM) on the female map	Position confidence interval (cM) on the sex averaged map	Proportion of phenotype variance (%)	Collective proportion of phenotype variance (%)
W	2	0–90	0–70	8.4	14.8
W	11	60–70	31–36	7.7	
L	2	0–90	0–70	5.6	11.2
L	11	60–70	31–36	5.5	
CF	11	20–50	10–30	5.7	9.8
CF	20	30–50	15–25	4.6	

Abbreviation: QTL, quantitative trait loci.

No correlation was found among QTLs linked to the same trait. Therefore, the phenotypic variance explained by two QTLs was nearly cumulative. A two QTL model revealed that chromosome 2 and 11 QTLs were collectively responsible for 14.8% of the variance of L (Table 2). Similarly, QTLs on chromosomes 2 and 11 were responsible for 11.2% of the variance of W, while QTLs on chromosome 11 and 20 were responsible for 9.8% of the variance of CF (Table 2).

Allelic fixation within type
Estimating the degree of fixation of the alleles within the parental generation revealed a strong correlation pattern among alleles from the farmed parents ($\rho_F = 0.92$ $P < 0.001$) under the weight and length QTL on chromosome 2 (Table 3). No significant correlation could be established among the alleles from the wild parents on the same QTL. Moderate correlation was also detected for the alleles from the wild parents ($\rho_W = 0.71$ $P < 0.01$) under the weight and length QTL on chromosome 11 (Table 3), whereas no significant correlation could be established for the alleles from the farmed parents on the same QTL. Moderate allele correlation was also detected among alleles from both the farmed and wild populations ($\rho_F = 0.46$ $P < 0.05$, $\rho_W = 0.53$ $P < 0.05$) for the CF QTL on chromosome 20. On the polygenic

scale, the correlation among alleles was not significantly different from zero in any of the parental lines. However, the estimate of allelic fixation was constantly smaller among wild alleles than among farm alleles (Table 3), suggesting a generally higher allelic diversity within wild population than within farm strains for the loci contributing to the studied phenotypes.

Scan for genomic regions associated with mortality
The presence of locus associated with survival between egg and smolt stage was investigated by comparing the genotypes of observed surviving individuals with the genotypes of simulated dead individuals in a binomial hierarchical generalized linear model framework. From this analysis, one large genomic region covering chromosome 2 was significantly linked to the variation in survival ($P < 0.05$) (Figure 4). The individual effect of each parental gamete was estimated after reconstructing a five-markers haplotype under mortality-associated region on chromosome 2. Comparing the observed number of parental haplotypes in the F1 population with those expected under the null hypothesis revealed that one wild male individual had a strong contribution to the survival variance. This individual 2005_M2 had indeed a significant and heterogeneous contribution to the survival

Table 3 Estimated allelic fixation within type at QTL on chromosome (Chr.) 2, chromosome 11 and at the polygenic scale (P)

Phenotype	Weight			Length			Condition factor		
	Chr.2	Chr.11	P	Chr.2	Chr.11	P	Chr.2	Chr.11	P
Wild	0.08 (n.s.)	0.71*	0.02 (n.s.)	0.01 (n.s.)	0.46 (n.s.)	0.01 (n.s.)	0.25 (n.s.)	0.53*	0.01 (n.s.)
Farm	0.92***	0.01 (n.s.)	0.20 (n.s.)	0.97***	0.01 (n.s.)	0.31 (n.s.)	0.01 (n.s.)	0.46*	0.12 (n.s.)

Abbreviation: QTL, quantitative trait loci.
***$P<0.001$, **$P<0.01$, *$P<0.05$, n.s., Fixation not significantly different from zero.
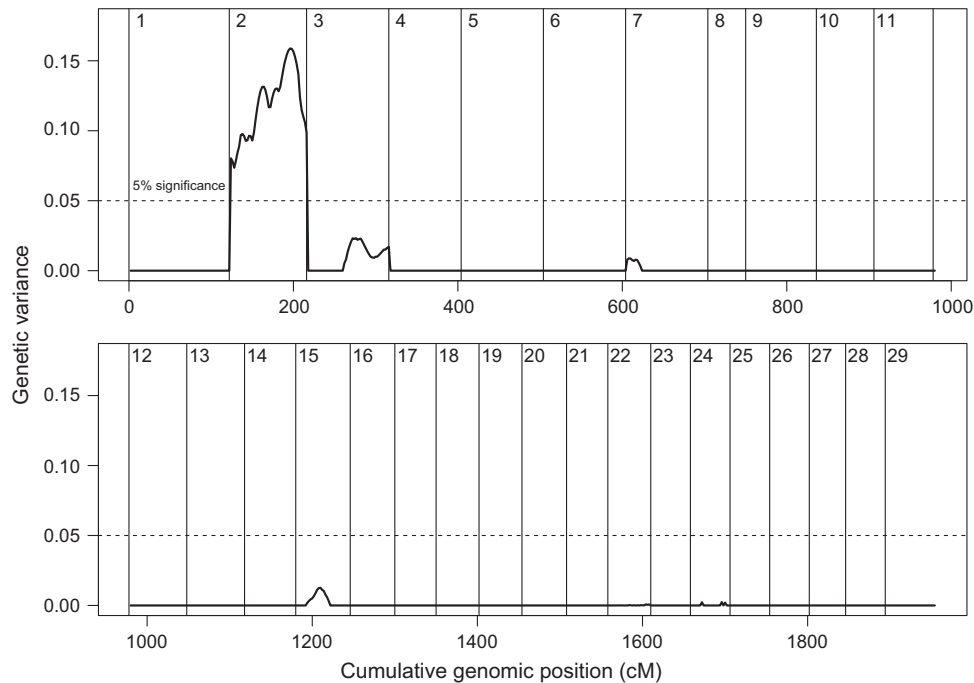
Figure 4 Genome-wide scan every 1 cm of the female recombination map for QTL affecting mortality. Horizontal dashed line and pointed line indicate the 5 and 1% genome-wide significance threshold, respectively. Vertical lines separate chromosomes.

Table 4 Haplotypes at five markers[a], on the genomic region of chromosome 2 associated with survival

	2004_L12 (P = 0.001)	2004_L18 (P = 0.001)	2005_M2 (P = 5.10⁻⁸)	2005_M9 (P = 0.001)
Beneficial haplotype	12 121 (56)	21 222 (53)	12 121 (84)	21 121 (109)
Alternative haplotype	21 121 (17)	22 112 (14)	11 211 (11)	11 211 (50)

Haplotypes are given for the four individuals that most contributed to the survival variance. The number between brackets indicates the number of living offspring that inherited the corresponding haplotype. *P*-values correspond to the chi-square test comparing the repartition of parental haplotypes among surviving offspring with the theoretical number if no selective mortality occurred in the river.
[a]ESTNV_33243_1092, BASS13_B7_B07_761, GCR_cBin34406_Ctg1_88, ESTV_21702_136 and ESTNV_36751_1595.

variance, with two alleles having opposite effects in the survival of the offspring. From the 95 surviving offspring of 2005_M2, 84 inherited the favorable allele, while only 11 inherited the alternative allele (Table 4). In addition to 2005_M2, three wild males, 2004_L12, 2004_L18 and 2005_M9 also had a significant contribution to the variance of survival. In the surviving offspring of those parents, we observed more than a twofold difference between the number of favorable haplotype and the alternative one (Table 4).

DISCUSSION
In the present study, we show that QTLs could indeed be identified for all four studied phenotypes, and that the QTLs linked to body size and

shape coincided with QTLs for the same traits previously identified in farmed environments. In addition, there was evidence for reduced genetic variation in some of QTLs in farmed fish, indicating possible selective sweeps as a result of the domestication process and selection program. One genomic region on chromosome 2 was strongly associated with mortality in the river stage, illustrating thus a significant genetic contribution to a key element of the fitness.

The results presented here provide important information about fitness variation among wild and farmed salmonid fishes and indicate that artificial selection affects ecologically important traits when farmed fish are reintroduced into a natural setting. The detection of survival QTL is also an important step toward a better understanding

of the genetic architecture of salmonid fitness in the natural environment, and a better prediction of the evolutionary trajectories of wild populations invaded by farm escapees.

Ecological relevance of the studied traits
Three of the studied traits (body length, weight and CF) are related to growth. After 10 generations of selection in aquaculture for commercially important traits, farmed strains grow several times faster than wild salmons under controlled hatchery conditions, revealing large genetic differences as a result of domestication (Solberg et al., 2013). Such difference in growth is, to a large extent, achieved through a more intense feeding behavior in farmed salmon. Offspring of escaped farmed salmon may therefore have a competitive advantage in natural environments relative to wild salmon, at least through the early (fry and parr) part of their life cycle (McGinnity et al., 2003). Yet, intensive feeding behavior in aquaculture environments is achieved at the cost of other traits that are ecologically important in the wild, such as shyness and predator avoidance (Einum and Fleming, 1997). To summarize, the available evidence suggests that farmed salmon may perform better than wild salmon at some life stages and worse at others, with a net result of decreased population size (McGinnity et al., 2003). Although farmed and wild salmon differ in a number of biological traits (Jonsson and Jonsson, 2006), feeding behavior and growth rate are probably linked to farm VS wild differences, due to directional selection for those traits in hatchery. It is also likely that the studied traits are linked with fitness differences between the two groups, due to competition for food and predator avoidance. It is however important to note that fitness is a highly complex trait that is influenced by a number of genetic factor, some of which, like for example immune capacity, are not considered in the present work. In addition, the environmental component of fitness is expected to be large, for example, farm fish fitness in the wild has been shown to benefit from enrichment of rearing conditions in hatchery (Roberts et al., 2014).

Growth and survival are also expected to result from a mixture of polygenic and environmental factors. The environmental component is expected to be particularly important in the present study where the environment is heterogeneous and uncontrolled. The effect of reported QTLs range between 4.8 and 8.4% of the phenotypic variance, with cumulative effect of 9.8–14.8%. This indicates that growth at the river stage is determined by a combination of environmental effects jointly with several genes with small individual contributions, rather than a few genes with large contributions.

Condition factor is a function of both fish's length and weight, and cannot be predicted from weight or length alone. This phenotype is nevertheless a very relevant variable to study as it may be linked to fitness. Among fish of similar weight, a large value of CF reflects a robust body shape with a shorter distance between head and tail compared to fish with small CF value that have a more slender shape and longer head to tail distance. Body conformation may thus be linked to adaptation to living conditions and life history trait, such as migratory behavior (Pulcini et al., 2013) or sexual selection (Roberts et al., 2014). In the studied population, the correlation between CF and weight and CF and length was respectively 0.20 and 0.22, while weight-length correlation was 0.85. This correlation between the studied phenotypes is reflected in the results of the QTL scan where QTLs linked to weight and length were nearly identical, whereas the scan for CF revealed a region on chromosome 20 that was linked to CF, but was not linked to weight nor length. Those results partially overlap with those from a former study (Reid et al., 2005) where signal for QTL linked to weight and CF was found on salmon linkage group 11.

QTL identification and dependence of the environment
The Guddal river does not host any natural population of salmon. Thus, in the experimental study upon which these QTL analyses are based (Skaala et al., 2012), the wild population originated from another river located on the west of Norway (Laerdal). While this donor population cannot be specifically regarded as adapted to the river Guddal, the offspring resulting from the wild experimental crosses were nevertheless expected to perform better in the river than the farmed strain that had been subject to approximately 10 generations of domestication selection in the absence of predation and competition for resources. We assumed here that the joint study of the farmed strain, the wild Laerdal population and their F1 hybrids in the river Guddal provided a suitable experimental setting in which to study the genetic basis of adaptation in the natural environment.

Several QTLs linked to weight, length and CF were detected on chromosomes 2, 11 and 20. These identified regions overlap with the results of previous QTL studies performed with different populations of farmed salmon reared in an aquaculture environment (Reid et al., 2005; Baranski et al., 2010). The proportion of variance explained by QTLs may vary depending on the statistical method for QTL detection. Here, with variance component QTL mapping, the estimated variance contribution of each QTL ranged from 4.6 to 8.4%, which was comparable to those from previous studies where QTLs explained 3–6% of weight and CF variance using half-sib regression analysis (Baranski et al., 2010), or 11–17% of weight and CF variance (Reid et al., 2005). It is well known that QTLs identified for a given trait may not be consistent across different populations and environments, the former due to the possibility of parallel evolution and the latter due to genotype × environment interactions (Slate et al., 2010). Our study provides some evidence that the same QTLs underlie the same traits in different populations, though the number of involved populations is small. This result is not surprising given that Norwegian commercial salmon strains were initially founded in the start of the 1970s by individuals from local wild populations (Gjedrem et al., 1991). Other studies that have found different genetic architecture for the same traits in different populations have studied more genetically diverged population with divergent geographical origin (Thurber et al., 2013).

Here, QTL detection relied on variance component linkage mapping, in a population that consisted of two generations (parents and offspring) of Atlantic salmon from either farmed or wild origin. In such populations, the second generation of intercross (F2) is usually a more appropriate mapping population as it gathers the maximum genetic polymorphism. However, because the present parent populations are not inbred, it is also possible to use the first-generation intercross (F1) as mapping population. The statistic model that was employed for QTL detection was adapted to this particular situation as it assumes that all alleles present in the parental generation are inherited from a different ancestor. This approach is thus particularly suitable for the analysis of outbred populations. It is however expected to be conservative, as it does not make use of the expected genetic similarities within parental lines. This approach nevertheless identified several QTL linked to body size and body conformation. For each of these QTLs, the flexible intercross analysis (Ronnegard et al., 2008) method was employed to estimate a posteriori genetic similarities within parental lines. This estimation indicated higher genetic diversity within wild population than within farm strain, probably revealing the consequence of selection and/or genetic drift within the farm strain.

In addition, the QTL chromosome 2 displayed a stronger reduction of the genetic diversity than observed at the polygenic level. This may reflect a selective sweep due to directional selection on chromosome 2 in hatchery conditions.

The use of a first-generation intercross (F1) also implied some limitations in the search for survival QTL. It was already established that in this data set, the families had significant differences in the number of offspring surviving until smolt migration stage (Skaala et al., 2012). The model utilized to detect genomic regions associated with survival was thus correcting for survival variation due to family effect. In this model, only heterozygous parents with gametes providing significantly different survival rate to the offspring had a significant contribution to the genetic component of survival. This is expected to considerably reduce the number of individuals that contribute to the genetic component of survival, and thus reduce the power of detection.

Nevertheless, this approach allowed the detection of one genomic region on chromosome 2 that was significantly linked to the survival rate in the population. As expected, the main contributor of mortality variance was a wild parent bearing two alleles with significantly different effect on offspring survival. In this particular case, the beneficial allele was found in 84 of the surviving offspring while the alternative allele was only present in 11 individuals on the F1 generation. As reported in Table 4, in chromosome 2, the main contributors to the survival variance were all wild males. This may again indicate a greater genetic diversity in the wild population. The reason why males would have a greater contribution than females is unclear. There are however several experimental constraints that could explain this. First, the same males were used in the wild-wild and farm-wild crosses. As a consequence, wild males have on average twice as many offspring than farm males, and females from both types. This larger sample size logically gives greater power to detect survival discrepancies in the offspring of wild males. Second, in salmon, the chromosome recombinations are more frequent in females than in males (Lien et al., 2011), making thus the haplotype reconstruction more accurate in males than in females.

As QTL areas are generally broad, it is not possible to assess whether the co-location of growth and survival QTL on chromosome 2 is coincidental, or whether the two phenotypes are affected by the same locus. Correlation between growth and survival has however been documented in salmonids (Vehvilainen et al., 2012). Further investigation using an F2 intercross or association mapping with higher marker density should allow a better estimation of the genetic variability linked to survival under that locus.

It is also important to note that incomplete sampling could potentially affect the results on mortality if a given family or type is more likely to be trapped than the others. However, the Wolf trap (Millis, 1991) is a systematic sampling method that is not likely to capture a certain type of fish due to body size, behavior or swimming ability. We thus assume that if the sampling is incomplete, the captured population is a representative sample of the families and types of fish that survived and migrated down the Guddal river during the experiment.

CONCLUSIONS

This study investigates the genetic architecture of ecologically important traits that underlie fitness differences between wild and farmed salmonids in the natural environment. The results provide evidence for the consistency of QTLs across contrasting captive and wild environments, and particularly for one of the QTLs where strong selection may have occurred in aquaculture. The environmental

independence of the QTLs and thereby low genotype×environment interaction further suggest that artificial selection in the aquaculture environment leading to phenotypic changes will have similar phenotypic effects in offspring of escaped farmed salmon in the wild. Therefore, selective changes in farmed salmon are expected to have direct influence not only on the genotypes but also on phenotypes of wild salmon populations subject to spawning intrusion by farmed fish. This reinforces the general conception that interbreeding between farmed and wild salmon represents an important conservation problem, and that avoidance of escapes from aquaculture should be highly prioritized.

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001). The problems with hybrids: setting conservation guidelines. Trends Ecol Evol 16: 613–622.
Baranski M, Moen T, Vage DI (2010). Mapping of quantitative trait loci for flesh colour and growth traits in Atlantic salmon (Salmo salar). Genet Sel Evol 42: 17.
Belyaev DK, Ruvinsky AO, Trut LN (1981). Inherited activation-inactivation of the star gene in foxes: its bearing on the problem of domestication. J Hered 72: 267–274.
Bourret V, Kent MP, Primmer CR, Vasemagi A, Karlsson S, Hindar K et al. (2013). SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (Salmo salar). Mol Ecol 22: 532–551.
Carlson SM, Seamons TR (2008). A review of quantitative genetic components of fitness in salmonids: implications for adaptation to future change. Evol Appl 1: 222–238.
Churchill GA, Doerge RW (1994). Empirical threshold values for quantitative trait mapping. Genetics 138: 963–971.
Clifford SL, McGinnity P, Ferguson A (1998). Genetic changes in an Atlantic salmon population resulting from escaped juvenile farm salmon. J Fish Biol 52: 118–127.
Darwin C (1868). The Variation of Animals and Plants under Domestication. John Murray: London.
Debes PV, Normandeau E, Fraser DJ, Bernatchez L, Hutchings JA (2012). Differences in transcription levels among wild, domesticated, and hybrid Atlantic salmon (Salmo salar) from two environments. Mol Ecol 21: 2574–2587.
Einum S, Fleming IA (1997). Genetic divergence and interactions in the wild among native, farmed and hybrid Atlantic salmon. J Fish Biol 50: 634–651.
Fischer J, Lindenmayer DB (2000). An assessment of the published results of animal relocations. Biol Conserv 96: 1–11.
Fleming IA, Einum S (1997). Experimental tests of genetic divergence of farmed from wild Atlantic salmon due to domestication. ICES J Mar Sci J Cons 54: 1051–1063.
Garcia de Leaniz C, Fleming IA, Einum S, Verspoor E, Jordan WC, Consuegra S et al. (2007). A critical review of adaptive genetic variation in Atlantic salmon: implications for conservation. Biol Rev Camb Philos Soc 82: 173–211.
George AW, Visscher PM, Haley CS (2000). Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. Genetics 156: 2081–2092.
Gjedrem T, Gjøen HM, Gjerde B (1991). Genetic origin of Norwegian farmed Atlantic salmon. Aquaculture 98: 41–50.
Glover K, Pertoldi C, Besnier F, Wennevik V, Kent M, Skaala O (2013). Atlantic salmon populations invaded by farmed escapees: quantifying genetic introgression with a Bayesian approach and SNPs. BMC Genet 14: 74.
Houston RD, Bishop SC, Hamilton A, Guy DR, Tinch AE, Taggart JB et al. (2009). Detection of QTL affecting harvest traits in a commercial Atlantic salmon population. Anim Genet 40: 753–755.
Johnson DL, Thompson R (1995). Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. J Dairy Sci 78: 449–456.
Jonsson B, Jonsson N (2006). Cultured Atlantic salmon in nature: a review of their ecology and interaction with wild fish. ICES J Mar Sci J Cons 63: 1162–1181.
Karlsson S, Moen T, Lien S, Glover KA, Hindar K (2011). Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. Mol Ecol Resour 11 (Suppl 1): 247–253.

Koljonen M-L, Tähtinen J, Säisä M, Koskiniemi J (2002). Maintenance of genetic diversity of Atlantic salmon (Salmo salar) by captive breeding programmes and the geographic distribution of microsatellite variation. *Aquaculture* **212**: 69–92.

Lewin HA (2009). Genetics. It's a bull's market. *Science* **324**: 478–479.

Lien S, Gidskehaug L, Moen T, Hayes B, Berg P, Davidson W et al. (2011). A dense SNP-based linkage map for Atlantic salmon (Salmo salar) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. *BMC Genomics* **12**: 615.

Malek TB, Boughman JW, Dworkin I, Peichel CL (2012). Admixture mapping of male nuptial colour and body shape in a recently formed hybrid population of threespine stickleback. *Mol Ecol* **21**: 5265–5279.

Martinez V, Dettleff P, Lopez P, Fernandez G, Jedlicki A, Yañez JM et al. (2013). Assessing footprints of selection in commercial Atlantic salmon populations using microsatellite data. *Anim Genet* **44**: 223–226.

McGinnity P, Prodöhl P, Ferguson A, Hynes R, Maoiléidigh Nó, Baker N et al. (2003). Fitness reduction and potential extinction of wild populations of Atlantic salmon, Salmo salar, as a result of interactions with escaped farm salmon. *Proc R Soc Lond B Biol Sci* **270**: 2443–2450.

McGinnity P, Stone C, Taggart JB, Cooke D, Cotter D, Hynes R et al. (1997). Genetic impact of escaped farmed Atlantic salmon (Salmo salar L.) on native populations: use of DNA profiling to assess freshwater performance of wild, farmed, and hybrid progeny in a natural river environment. *ICES J Mar Sci J Cons* **54**: 998–1008.

Millis D (1991). *Ecology and Management of Atlantic Salmon*. illustrated Springer Science & Business Media. 1991.

O'Reilly PT, Hamilton LC, McConnell SK, Wright JM (1996). Rapid analysis of genetic variation in Atlantic salmon (Salmo salar) by PCR multiplexing of dinucleotide and tetranucleotide microsatellites. *Can J Fish Aquat Sci* **53**: 2292–2298.

Pong-Wong R, George AW, Woolliams JA, Haley CS (2001). A simple and rapid method for calculating identity-by-descent matrices using multiple markers. *Genet Sel Evol* **33**: 453–471.

Price EO (1999). Behavioral development in animals undergoing domestication. *Appl Anim Behav Sci* **65**: 245–271.

Price EO, King JA (1968). Domestication and adaptation. In: Hafez ESE (Ed) *Adaptation of Domestic Animals*. Lea and Febiger: Philadelphia. pp 34–45.

Pulcini D, Wheeler PA, Cataudella S, Russo T, Thorgaard GH (2013). Domestication shapes morphology in rainbow trout Oncorhynchus mykiss. *J Fish Biol* **82**: 390–407.

Reid DP, Szanto A, Glebe B, Danzmann RG, Ferguson MM (2005). QTL for body weight and condition factor in Atlantic salmon (Salmo salar): comparative analysis with rainbow trout (Oncorhynchus mykiss) and Arctic charr (Salvelinus alpinus). *Hered Edinb* **94**: 166–172.

Rhymer J, Simberloff D (1996). Extinction by hybridization and introgression. *Annu Rev Ecol Syst* **27**: 83–109.

Roberts LJ, Taylor J, Gough PJ, Forman DW, Garcia de Leaniz C (2014). Silver spoons in the rough: can environmental enrichment improve survival of hatchery Atlantic salmon Salmo salar in the wild? *J Fish Biol* **85**: 1972–1991.

Ronnegard L, Besnier F, Carlborg O (2008). An improved method for quantitative trait loci detection and identification of within-line segregation in F2 intercross designs. *Genetics* **178**: 2315–2326.

Rönnegård L, Carlborg O (2007). Separation of base allele and sampling term effects gives new insights in variance component QTL analysis. *BMC Genet* **8**: 1.

Ronnegard L, Shen X, Alam M (2010). hglm: A package for fitting hierarchical generalized linear models. *R J* **2**: 20–28.

Skaala Ø, Glover KA, Barlaup BT, Sv\aasand T, Besnier F, Hansen MM et al. (2012). Performance of farmed, hybrid, and wild Atlantic salmon (Salmo salar) families in a natural river environment. *Can J Fish Aquat Sci* **69**: 1994–2006.

Skaala Ø, Høyheim B, Glover K, Dahle G (2004). Microsatellite analysis in domesticated and wild Atlantic salmon (Salmo salar L.): allelic diversity and identification of individuals. *Aquaculture* **240**: 131–143.

Slate J, Santure AW, Feulner PGD, Brown EA, Ball AD, Johnston SE et al. (2010). Genome mapping in intensively studied wild vertebrate populations. *Trends Genet* **26**: 275–284.

Slate J, Visscher PM, MacGregor S, Stevens D, Tate ML, Pemberton JM (2002). A genome scan for quantitative trait loci in a wild population of red deer (Cervus elaphus). *Genetics* **162**: 1863–1873.

Slettan A, Olsaker I, Lie O (1995). Atlantic salmon, Salmo-salar, microsatellites at the SSOSL25, SSOSL85, SSOSL311, SSOSL417 loci. *Anim Genet* **26**: 281–282.

Solberg MF, Skaala Ø, Nilsen F, Glover KA (2013). Does domestication cause changes in growth reaction norms? a study of farmed, wild and hybrid atlantic salmon families exposed to environmental stress. *PLoS ONE* **8**: e54469.

Taggart JB (2007). FAP: an exclusion-based parental assignment program with enhanced predictive functions. *Mol Ecol Notes* **7**: 412–415.

Team RC (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Thurber CS, Jia MH, Jia Y, Caicedo AL (2013). Similar traits, different genes? Examining convergent evolution in related weedy rice populations. *Mol Ecol* **22**: 685–698.

Vehvilainen H, Kause A, Kuukka-Anttila H, Koskinen H, Paananen T (2012). Untangling the positive genetic correlation between rainbow trout growth and survival. *Evol Appl* **5**: 732–745.

Verspoor E, Beardmore JA, Consuegra S, García de Leániz C, Hindar K, Jordan WC et al. (2005). Population structure in the Atlantic salmon: insights from 40 years of research into genetic protein variation. *J Fish Biol* **67**: 3–54.

Webb AA, Cullen CL (2010). Coat color and coat color pattern-related neurologic and neuro-ophthalmic diseases. *Can Vet J* **51**: 653–657.

Supplementary Information accompanies this paper on Heredity website (http://www.nature.com/hdy)