Molecular Ecology (2008) **17**, 314–324 doi: 10.1111/j.1365-294X.2007.03438.x

Genetic consequences of interbreeding between farmed and wild Atlantic salmon: insights from the transcriptome

CHRISTIAN ROBERGE,* ÉRIC NORMANDEAU,* SIGURD EINUM,†‡ HELGA GUDERLEY* and LOUIS BERNATCHEZ*
*Département de biologie, Université Laval, Québec, Canada G1K 7P4, †Norwegian University of Science & Technology, Department of Biology, NO-7491 Trondheim, Norway, ‡Norwegian Institute for Nature Research, NO-7485 Trondheim, Norway

Abstract

Large annual escapees of farmed Atlantic salmon enhance the risk of extinction of wild populations through genetic and ecological interactions. Recently, we documented evolutionary change in gene transcription between farmed and wild Atlantic salmon after only five generations of artificial selection. While differences for most quantitative traits are expected to gradually dilute through repeated backcrossing to wild populations, the genetic basis of gene transcription has been shown to be largely nonadditive and hybrid crosses may display unexpected inheritance patterns. This makes it difficult to predict to what extent interbreeding between farmed and wild individuals will change the genetic makeup of wild salmon populations. Here, we compare the genome-wide gene transcription profiles of Norwegian wild salmon to that of a second generation hybrid cross [backcross: (Farmed X Wild) X Wild]. Over 6% (298, *q*-value < 0.01) of the detected genes exhibited highly significantly different transcription levels, and the range and average magnitude of those differences was strikingly higher than previously described between pure farmed and wild strains. Most differences appear to result from nonadditive gene interactions. These results suggest that interbreeding of fugitive farmed salmon and wild individuals could substantially modify the genetic control of gene transcription in natural populations exposed to high migration from fish farms, resulting in potentially detrimental effects on the survival of these populations. This further supports the idea that measures to considerably reduce the number of escaped farmed salmon and their reproduction in the wild are urgently needed.

Keywords: conservation, farmed salmon, gene expression, hybridization, introgression, microarray, nonadditivity, rapid evolution

Received 19 March 2007; revision accepted 24 May 2007

Introduction

Since the 1970s, Atlantic salmon have responded successfully to intense artificial selection aimed at improving growth rates as well as other traits of commercial interest (Gjøen & Bentsen 1997). Genetically based phenotypic changes not specifically selected for have also resulted from such selection programmes, including increased fat content in flesh (Rye & Gjerde 1996) and poorer performance in the wild (McGinnity *et al*. 2003), as well as physiological, morphological and behavioural changes (Fleming & Einum 1997; Fleming *et al*. 2002).

Farmed salmon are considered an important problem for natural populations when they escape from sea-cages and enter streams and rivers in significant proportions. About two million escape annually, whereas the total wild adult effectives are estimated to be about four million (McGinnity *et al*. 2003). These fugitives are thought to greatly enhance the risk of extinction of wild populations through genetic and ecological interactions (McGinnity *et al*. 2003; Hindar *et al*. 2006), as well as by spreading diseases (Naylor *et al*. 2005). Monitoring in a range of Norwegian wild

Correspondance: Christian Roberge, Fax: 1-418-656-7176; E-mail: christian.roberge@giroq.ulaval.ca

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

populations indicated that escaped farmed fish constitute on average close to 20% of the adult population during the breeding season (Fiske *et al*. 2001). Experimental evidence and nature observations also suggested that most successful breeding of farmed salmon occurs through hybridization with wild fish (Fleming *et al*. 2000). A pressing question is therefore to what extent interbreeding between farmed and wild individuals will change the genetic make-up of wild Atlantic salmon populations.

It is increasingly acknowledged that evolutionary changes may strongly depend on alterations of gene transcription regulation (Wilson 1976). Using the microarray technology, which offers the possibility of a genome-wide scan for gene transcription differences, we recently compared patterns of gene expression in the progeny of farmed and wild Atlantic salmon from Norway and Canada grown in controlled conditions. Our results demonstrated that five to seven generations of artificial selection led to heritable changes in gene transcription profiles, the average magnitude of the differences being 25% and 18% for at least 1.4% and 1.7% of the expressed genes in juvenile salmon from Norway and Canada, respectively (Roberge *et al*. 2006). Moreover, genes showing significant transcription profile differences in farmed strains from both Canada and Norway (16% of the differentially expressed genes) all exhibited parallel changes, providing evidence for the role of artificial selection in driving evolutionary changes of gene transcription levels between wild and farmed salmon.

Genetic changes for most quantitative phenotypic traits in farmed salmon can be expected to gradually dilute through repeated backcrossing to wild populations (if non-native immigration ceases) due to their additive genetic basis (Anderson 1949; McGinnity *et al*. 2003; Hindar *et al*. 2006; Tymchuk *et al*. 2006). However, the genetic basis of gene transcription appears to be largely nonadditive (Gibson *et al*. 2004; Hedgecock *et al*. 2007; but see Cui *et al*. 2006; Swanson-Wagner *et al*. 2006) and hybrid crosses sometimes display nonadditive inheritance patterns for quantitative traits, such as outbreeding depression (Gharrett *et al*. 1999) and heterosis (Lippman & Zamir 2007). Moreover, in the absence of strict policies to reduce the number of escaped farmed salmon considerably, this number is expected to increase due to the rapid development of the aquaculture industry, further challenging the genetic integrity of natural populations. Sea farming is also rapidly developing for other marine fish species (e.g. Atlantic cod and halibut; see Naylor *et al*. 2005), and similar risks could threaten their natural populations as well.

Here we compare, using a 16 006-gene microarray, the genome-wide gene transcription profile of the progeny (fry stage) of Norwegian wild salmon to that of a second generation hybrid cross [backcross: (Farmed X Wild) X Wild] to test whether genetically based differences in gene transcription persisted in these hybrids. We also compare our results to those of our previous study on genome-wide transcription divergence between pure farmed and wild salmon (Roberge *et al*. 2006) in order to gain insights into the relative importance of additive vs. nonadditive genetic interactions in controlling gene transcription in salmon. In this comparison, we address four main questions: (i) how do the number of significant gene transcription differences observed; and (ii) the average magnitude and range of these differences compare between the two studies? In other words, are there fewer and smaller differences between backcross and wild salmon than between farmed and wild salmon? (iii) how does the gene transcription difference between backcross and wild salmon compare to that observed between wild and farmed salmon for genes that had shown significant transcription level differences between wild and farmed salmon; and (iv) for genes that showed significant transcription level differences between hybrid backcross and wild salmon? Finally, we present quantitative polymerase chain reaction (PCR) confirmation of some of our earlier results on gene transcription differences observed in Canadian farmed vs. wild salmon. By documenting genome-wide effects of interactions between farmed and wild salmon genomes on patterns of gene expression, our main goal was to contribute to a better understanding of the evolutionary impact of farm escapees on wild salmon populations.

Methods

Fish crosses and samples

Twenty wild genitors from river Namsen (64°27′N, 11°28′E) kept in captivity for one generation (in the absence of deliberate selection) were used to generate 10 full-sib wild families at the NINA (Norwegian Institute for Nature Research) research station in Ims (Norway) in the autumn of 2004. Ten hybrid (wild X farmed) genitors were crossed with 10 wild genitors to generate 10 full-sib backcross families (five families had wild mothers, and five had hybrid mothers). Norwegian farmed genitors were taken from the seventh generation of the Norwegian breeding programme's first brood line ('Population 1', *sensu* Gjedrem *et al*. 1991), which mainly originated from river Namsen (Gjedrem *et al*. 1991; Fleming & Einum 1997). Fertilized eggs were all kept under identical controlled conditions (24 h darkness; mean temperature (range) 3.9 °C (0.8–11.1)). The progeny of the 10 full-sib families of each type was mixed for rearing and a hundred fry of each type were randomly sampled. Because juvenile characteristics play a determinant role in the expression of life history traits at later stages (Metcalfe *et al*. 1989), and mortality and natural selection are commonly intense during early life (Einum & Fleming 2000a, b; Einum *et al*. 2002), this study focused on the fry

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

stage (young of the year). Sexually undifferentiated fry were sampled at the yolk-sac resorption stage, before exogenous feeding and immediately frozen in liquid nitrogen. Sampling stage was thus the same as in Roberge *et al*. (2006).

RNA extraction, labelling and cDNA hybridization

Frozen fry were homogenized individually in TRIZOL@Reagent (Invitrogen), and total RNA was extracted as previously described (Roberge *et al*. 2006; Roberge *et al*. 2007). For each sample, 15 μg total RNA was retro-transcribed and labelled. Transcription profiles of 23 of pure wild and 23 backcross fry were contrasted on 23 microarrays, always using both types on each array. Dye and sample cross type coupling was flipped between biological replicates. The cDNA microarrays used here were obtained through the consortium Genomic Research on All Salmon Project (cGRASP, available from Ben F. Koop, bkoop@uvic.ca), which comprises 16 006 salmonid cDNA clones (von Schalburg *et al*. 2005).

Signal detection, data preparation, statistical analysis and functional classification

Signal detection and data preparation was as previously reported (Roberge *et al*. 2006). Spots with mean signal intensities smaller than the mean intensity of control empty spots plus twice its standard deviation in both channels were removed from the analysis, leaving a total of 4618 detected spots. To assess differences in gene transcription between wild and backcross fish, data were analysed using a mixed model of ANOVA (Wolfinger *et al*. 2001) and the MAANOVA R package (Kerr *et al*. 2000, 2002). The model included the 'array' term as a random term and the 'sample type' (pure wild or backcross) and 'dye' terms as fixed terms. A permutation-based *F*-test (Fs, with 1000 sample ID permutations) was then performed, and restricted maximum likelihood was used to solve the mixed-model equations. To correct for multiple testing, *Q*-values were calculated from the permutation based *P*-values with the *Q*-value R package (Storey 2002). Since samples from a same family cannot be considered as independent, our model should ideally have included a family term (Millar & Anderson 2004). However, space limitations prevented us from rearing family groups separately and familial information on the samples was lost. We therefore had to deal with a certain level of pseudo-replication in these data, which may inflate statistical power (Hurlbert *et al*. 1984). Consequently, we used a conservative significance threshold of q-value < 0.01. Functional classification and assessment of significant differential representation of functional classes were performed in the DAVID/EASE environment (http://david.niaid.nih.gov/david/). The DAVID 2.1 (beta version) gene accession conversion tool was first used to convert gene ontology-linked identifications of various types gathered in the GRASP 16 006-gene microarray gene identification file to UNIGEN clusters. Assessment of significant differential representation of functional classes was performed with EASE 2.0.

Comparison with previous results

A comparison of the results of the present study and that of Roberge *et al*. (2006) should take into account two main potential caveats. First, the studies were performed using different cDNA microarrays, though both were prepared by the same group (cGRASP), using similar protocols (see http://web.uvic.ca/cbr/grasp). The microarray used in this study represents an 'upgraded' version which comprises substantially more features (17 328 different features) than the one used in our earlier study (8736 features with neighbour replicates, 4368 different features). Only about 25% of the cDNA clones printed on the microarray used in the present study were also printed on the earlier microarray platform and, inversely, some of the cDNA clones printed on the earlier one were not printed on the new microarray (including 18 'unknown' cDNAs for which significant expression differences were detected in Roberge *et al*. 2006). Consequently, we limited our comparisons to genes represented on both array platforms, keeping for each gene the data from the cDNA clone that showed the most significant hybridization signal difference on each array type (smallest associated *P*- and *Q*-values), in cases where a gene was represented by several different cDNA clones. Second, the number and proportion of genes showing significant transcription level differences at a given significance threshold is not readily comparable between studies because the statistical power to detect significant changes varies with sample size and experimental variance (Draghici 2003). Also, if more cDNAs are printed on the array, more statistical tests are performed and more false positives are then expected. Experimental variance is expected to be similar in both studies since all hybridizations and extractions were performed by the same person, CR, using the same material, reagents and protocols. Yet, over twice as many hybridizations were performed in the present study compared to Roberge *et al*. (2006) ($n = 46$ instead of $n = 20$ for the Norwegian system alone). To compare the average magnitude and range of the significant gene-transcription differences between the two studies, we thus focused on the 1.4% cDNA clones representing the most significantly, differentially expressed genes [i.e. the proportion of significant genes observed in Roberge *et al*. (2006)]. A common proportion was chosen instead of a given significance threshold, since, with different statistical powers, the same threshold would have represented different stringencies, which could in turn have influenced the average.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Quantitative PCR validation of microarray results

Quantitative real-time polymerase chain reaction (qRT-PCR) was used here to validate earlier microarray results (Roberge *et al*. 2006) for two candidate genes which had shown significant gene transcription differences between wild and farmed Atlantic salmon strains from Canada. The first candidate is an 'unknown' gene (GenBank accession no. CA039081), meaning the corresponding cDNA clone sequence did not generate any BLAST hits with *e*-value < 1×10^{-15} and an informative name. The second gene codes for metallothionein A. Seventy-nine frozen fry sampled at the yolk-sac resorption stage from families of farmed (39) and wild salmon (40) were used (see Roberge *et al*. 2006). These individuals had not been used in our previous microarray experiments. Total RNA was extracted from the whole individuals using a standard Trizol extraction protocol, mRNA was then retro-transcribed using the High-capacity cDNA archive kit from Applied Biosystems. TaqMan assay primers and probes were designed with PRIMER EXPRESS 1.5 and then tested in a validating experiment to ensure comparable efficiencies. The control gene (peptidyl-prolyl isomerase A) was chosen because it showed minimal variance in its transcription level, both within and between strains. Quantitative RT-PCR experiments were performed on an ABI 7000 instrument connected to a computer running version 1.2.3 of ABI Prism 7000 SDS (Applied Biosystems). During the qRT-PCR, the recommended default setups from Applied Biosystems were used. Thus, a 2-min step at 50 °C was programmed to ensure activation of the No AmpErase® UNG enzyme, which was followed by a 10-min step at 95 °C to activate the AmpliTaq® Gold enzyme. After those initial steps, 40 cycles were run, which consisted of 15 s at 95 °C followed by 1 min at 60 °C for hybridization, elongation and detection of fluorescence signal. All PCRs were scaled down from 50 μL to 25 μL and the proper detection volumes entered in the software prior each run. For the unknown gene (CA039081), the sequences and hybridization temperature of the primers used for amplification were: 5′-GATAGCTTCCAGAATACTACAGTGACAATT-3′ (59 °C) and 5′-CGTGGCCTTTTTTCGACTGA-3′ (60 °C), while the sequence and hybridization temperature of the TaqMan probe was 5′-TGCAAAACCAGACTATTAT-3′ (69 °C). For metallothionein, the sequences of the primers were 5′-GCCTCACTGACAACAGCTGGTA-3′ (59 °C) and 5′-TGTGCTTCAGGCTGTGTGTGTGTGT-3′ (59 °C) and that of the probe 5′-CACAGGTCTTGCCC-3′ (68 °C). Finally, for Peptidyl-prolyl isomerase A, the sequences of the primers were 5′-AAGAACTGGGACCCGTTGGT-3′ (59 °C) and 5′-GCATGGGCTGTCTGTCCAT-3′ (58 °C) and that of the probe 5′-TTAGGGCCAGCGTTGG-3′ (70 °C). Following the experiments, the transcription levels of the studied genes for each individual were normalized first, by subtracting the transcription level of the control gene in that individual and, second, by subtracting the difference between the transcription level of the control gene in the individual as compared to a reference individual that was maintained throughout the experiment. All statistic tests were run in R (version 2.3.1) using the normalized transcription levels described above. Normality of the dataset was rejected by the Shapiro-Wilk test while homogeneity of variances was accepted by the Ansari-Bradley test. Wilcoxon's rank sum test was therefore used to determine the *P*-values associated with the differences in normalized transcription levels between the two strains for the two studied genes. Since this experiment was aimed at validating differences already observed independently, a one-sided test was performed.

Results

Under assumptions of an additive genetic basis, more phenotypic variance is expected in second-generation hybrids than in the parental populations or species (Anderson 1949). However, the *F*-test for variance homoscedasticity did not reveal more significant differences in the variance of gene transcription levels between the progeny of hybrid and wild salmon than expected by chance. 6.4% (298) of all detected cDNA clones represented genes which showed significantly different transcription levels (*Q*-value < 0.01) between the progeny of wild and hybrid (backcross) salmon. DAVID/EASE analysis for over-representation of gene ontology categories among those genes, when compared to all genes represented on the microarray, revealed several significantly over-represented categories (Fig. 1, EASE score < 0.025). Each of these over-represented categories was directly related to one of four main functional axes: cytoskeleton/cellular organization, muscle development, oxidative phosphorylation and extracellular component. Strikingly, the vast majority of the genes falling into the first category (which largely overlapped with the second) appeared over-transcribed in hybrid vs. wild salmon, including for instance cytoskeletal keratin (10 significant cDNA clones, average: 44% over-expression), cofilin (51% over-expression) and gelsolin (47% over-expression). Inversely, all genes coding for ATP synthase subunits (main genes belonging to the third functional axis) were under-transcribed in hybrid vs. wild salmon (three subunits, 12 cDNA clones, average = 31% under-expression). Data for genes from the fourth functional axis (extra-cellular component) showed conflicting trends which reflected the wider array of functions this axis comprised. For instance, while all collagen-coding genes were under-transcribed in hybrid relative to wild salmon (12 cDNA clones, average = 28% under-expression), two different genes whose products are involved in antigen presentation showed over-transcription in hybrids (MHC class II antigen IE-α,

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

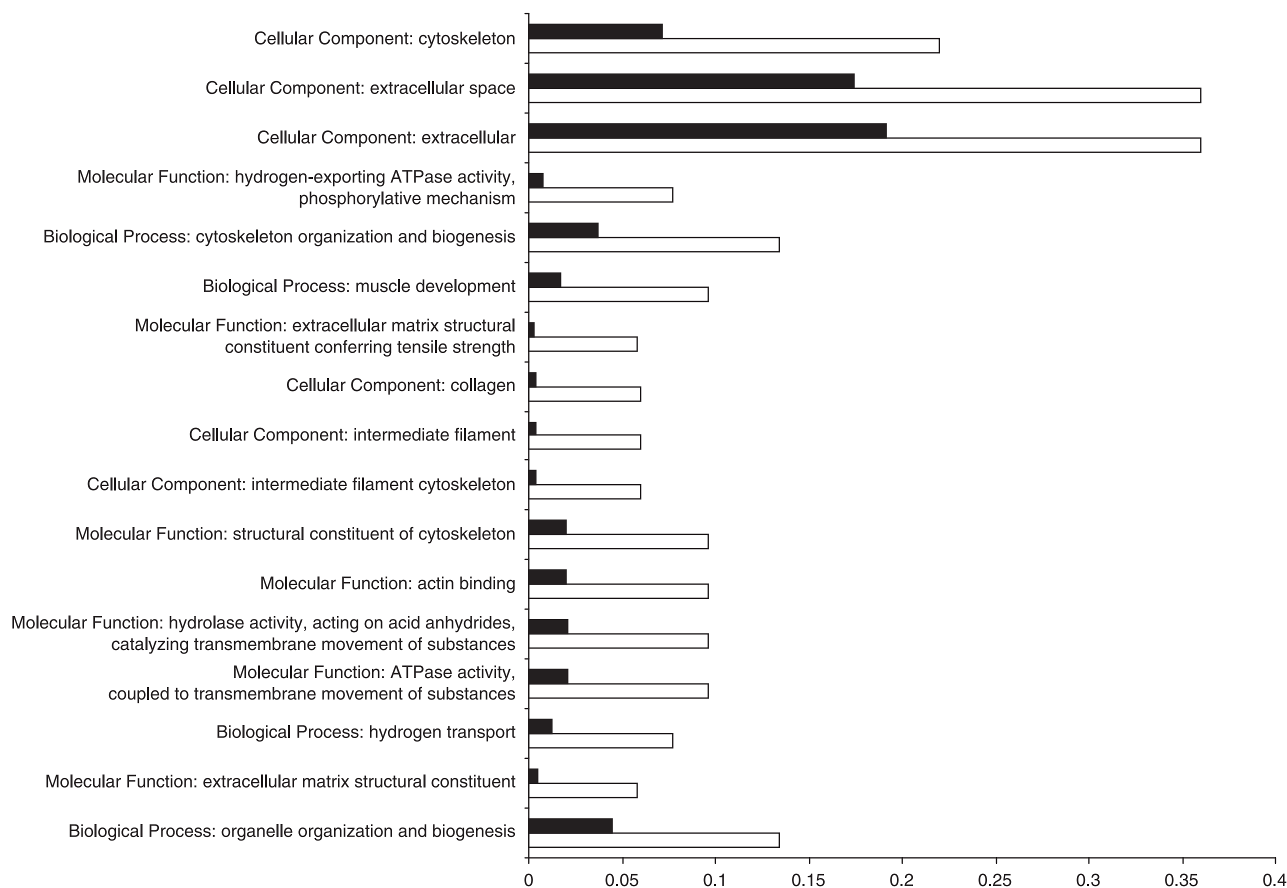

Fig. 1 Significant over-representation of gene-ontology categories among genes which showed significantly different transcription levels between the progeny of wild and hybrid backcross salmon when compared to all genes represented on the microarray (EASE score < 0.025). For each gene-ontology category listed, the black bar indicates the proportion of genes in this category among all known genes on the microarray while the white bar indicates the proportion of genes in this category among genes which showed significantly different transcription levels (Q-value < 0.01) between the progeny of wild and hybrid backcross salmon.

four cDNA clones, average = 52% over-transcription; α-2 microglobuline, five cDNA clones, 61% over-transcription). However, since functional annotation was possible for only a portion of the genes (40% of the significant cDNA clones and 26% of all clones on the microarray), results from this functional over-representation analysis might not be representative of the results in their entirety.

Known genes for which the transcription level differences between hybrid and wild salmon fry were significant even when using a Bonferroni-corrected threshold are presented in Table 1. Data for 62 cDNA clones satisfied this very conservative criterion; these represented 24 genes with functional annotation and 18 'unknown' genes. Among these, the magnitude of the gene transcription differences averaged almost two fold (1.98, either over- or under-transcription in hybrid relative to wild salmon) and ranged from a 7.15 fold under-transcription to a 7.18 fold over-transcription in hybrid vs. wild salmon.

At the $P < 0.01$ threshold on permutation-corrected P-values, 74 significant cDNA clones were observed in Roberge *et al*. (2006) compared to 656 in the present study. This represents, respectively, 1.4% and 13.2% of all detected clones after subtracting the expected number of false positives in each case. To evaluate how much of this difference is explained by increased statistical power due to the greater number of replicates in the present study, we assessed the number of significant clones at $P < 0.01$ in a random sub-sample of 20 comparisons from the original sample of 46. The proportion and number of cDNAs representing differentially expressed genes (7.5%, 392, $P < 0.01$) was still substantially higher than those observed between farmed and wild Norwegian salmon in Roberge *et al*. (2006).

The average magnitude (for the 1.4% cDNA clones representing the most significantly differentially expressed genes in both studies) of the differences between backcross and wild salmon was 76% (in either direction), which is

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Table 1 Gene products corresponding to known genes that showed significant gene transcription differences between the progeny of hybrid backcross and wild salmon when using a Bonferroni-corrected significance threshold ($P < 1.1 \times 10^{-5}$, all in bold in the 'BC/wild' column). The first column presents the number of different cDNA clones among the 62, showing significant differences at this threshold corresponding to each gene. Transcription-level fold changes for those genes when compared between the progeny of wild and farmed salmon (Roberge *et al.* 2006) or wild and hybrid backcross (BC) salmon (this study) are presented for comparison, as well as our interpretation of the transmission pattern observed. A, additive transmission, meaning that the BC value is intermediate between farmed and wild values; D, dominant transmission, meaning that the BC value is near to that observed in farmed salmon; TMUT, transgressive mean under-transcription, meaning that the value observed in BC is lower than in both parental strains; TMOT, transgressive mean over-transcription, meaning that the value observed in backcross is substantially higher than in both parental strains. Fold changes in bold in the 'farmed/wild' column were significant in Roberge *et al.* (2006; $P < 0.01$). NA indicates cases where clones representing a gene could not be identified on the microarray used in Roberge *et al.* (2006)

Gene product	No. of clones	BC/wild	Farmed/wild	Transmission
ADP, ATP carrier protein 2	2	**0.75**	0.91	TMUT
C1q-like adipose specific protein	2	**1.10**	NA	NA
Fc receptor beta subunit family	2	**2.24**	NA	NA
Low affinity immunoglobulin epsilon Fc receptor	2	**0.14**	NA	NA
Anterior gradient-like protein	5	**2.81**	0.93	TMOT
Parvalbumin	4	**1.14**	0.82	TMOT
Beta-actin	2	**1.22**	**1.36**	A
Lysozyme g	2	**0.85**	**0.85**	D
Thioredoxin-like protein p19 precursor	2	**3.49**	0.93	TMOT
Collagen alpha 1(X) chain	3	**0.50**	1.01	TMUT
Troponin I, fast skeletal muscle	2	**1.37**	1.13	TMOT
Invariant chain S25-7	1	**1.55**	NA	NA
Vitellogenin	1	**1.26**	1.13	TMOT
Ependymin	2	**0.50**	0.88	TMUT
Deoxyribonuclease I-like 3	2	**1.49**	NA	NA
Phosphoglycerate kinase 1	1	**0.31**	NA	NA
Gastrotropin	1	**0.57**	NA	NA
Claudin	1	**2.09**	1.03	TMOT
alpha-globin and beta-globin	1	**0.68**	**1.29**	TMUT
Ribosomal protein S3a	1	**1.80**	1.07	TMOT
ATP synthase lipid-binding protein	2	**0.66**	1.11	TMUT
Matrix metalloproteinase-2	1	**1.07**	**1.24**	A
Myosin regulatory light chain 2	1	**0.62**	1.08	TMUT
Hemopexin	1	**0.66**	0.85	TMUT

over three times that observed between wild and farmed strains (25%). Significant gene-expression differences in backcross vs. wild individuals ranged from 7.15-fold under-transcription to 7.18-fold over-transcription, while the highest difference observed between pure wild and pure farmed Norwegian salmon strains was of 1.67 fold.

Various candidate genes whose transcription levels appear to have evolved as a result of artificial selection and domestication were identified in Roberge *et al.* (2006). What happened to these differences after two generations of interbreeding between farmed and wild salmon? Table 2 contrasts transcription-level fold changes for these genes when compared between the progeny of wild and farmed salmon on the one hand, and that of wild and backcross salmon on the other. The difference in gene transcription observed between pure wild and pure farmed strains (Roberge *et al.* 2006) was either significant but smaller, or not significant between wild and backcross hybrids for 67% of the 33 different known genes that could be compared (Table 2). This suggests that the genetic control of the expression of those genes was additive (Gibson *et al.* 2004). The remaining genes showed genetic control of gene transcription more akin to low- or high-parent dominance (9%), or to other patterns (24%) which were previously described as gene expression over- or under-dominance (e.g. Gibson *et al.* 2004). Here, we did not adopt this terminology since over- or under-dominance refers to one locus' inter-allelic interactions, while gene transcription is expected to be generally controlled by several loci (Brem *et al.* 2002). Nor do we consider the heterosis/outbreeding depression terminology (e.g. Vuylsteke *et al.* 2005) as appropriate here since we do not know how the transgressive mean fold-changes observed affect growth or other performances. Instead we used the terminology 'transgressive mean under-' and 'transgressive mean over-transcription' (TMUT and TMOT).

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Table 2 Gene products corresponding to known genes that showed significant gene-transcription differences between the progeny of Norwegian farmed and wild salmon in Roberge *et al.* (2006) ($P < 0.01$; all in bold in the 'farmed/wild' column). Transcription-level fold changes for those genes when compared between the progeny of wild and farmed salmon (Roberge *et al.* 2006) or wild and hybrid backcross (BC) salmon (this study) are presented for comparison. The *q*-value indicating the significance of the fold change observed here between wild and hybrid BC salmon is also included, as well as our interpretation of the transmission pattern observed. A: additive transmission, meaning that the BC value is intermediate between farmed and wild values or that a significant difference in Roberge *et al.* (2006) is not significant anymore in the present study (A*). D, dominant transmission, meaning that the BC value is similar to that observed in farmed salmon; TMUT, transgressive mean under-transcription, meaning that the value observed in backcross is substantially lower than in both parental strains; TMOT, transgressive mean over-transcription, meaning that the value observed in backcross is substantially higher than in both parental strains. Fold changes in bold in the 'BC/wild' column were significant at the Q-value < 0.05 threshold. Q-values represent FDR-corrected P-values

Gene product	Farmed/wild	BC/wild	*q*-value	Transmission
Creatine kinase	**1.23**	1.05	1.20×10^{-1}	A
Glyceraldehyde phosphate dehydrogenase	**0.83**	**0.87**	1.34×10^{-2}	A or D
NADH dehydrogenase subunit 5	**1.31**	0.95	2.07×10^{-1}	A*
NADH dehydrogenase subunit 4	**1.30**	0.98	4.99×10^{-1}	A*
ATP synthase beta-subunit	**0.84**	**0.78**	2.93×10^{-2}	D or TMUT
Ferritin H	**1.23**	**1.30**	7.15×10^{-2}	A*
Chaperonin containing T-complex polypeptide 1, epsilon subunit	**1.29**	**1.35**	1.14×10^{-2}	D or TMOT
Ribosomal protein L10	**1.22**	**1.08**	4.08×10^{-2}	A
Ran protein	**1.24**	**1.14**	4.20×10^{-2}	A
Elongation factor 1 alpha	**1.24**	0.93	3.80×10^{-1}	A*
Protein synthesis initiation factor 4	**1.24**	**1.04**	3.66×10^{-3}	A
Alpha 3 type I collagen	**1.59**	**0.82**	2.38×10^{-2}	TMUT
Alpha 2 type I collagen	**1.67**	**0.79**	3.04×10^{-3}	TMUT
Matrix metalloproteinase-2	**1.34**	**1.07**	7.17×10^{-4}	A
Secreted protein, acidic, rich in cysteine (SPARC)	**1.30**	**0.81**	1.81×10^{-3}	TMUT
Mannose binding-like lectin	**0.84**	1.00	1.80×10^{-1}	A*
Pentraxin	**0.86**	1.16	6.10×10^{-1}	A*
Lysozyme G	**0.85**	0.85	0	D
BA1 beta-2 microglobulin	**0.84**	**1.72**	1.02×10^{-3}	TMOT
alpha-globin	**1.28**	**0.68**	3.52×10^{-4}	TMUT
Apolipoprotein A-I	**1.89**	**0.71**	1.06×10^{-2}	TMUT
Plasma retinol-binding protein 1	**0.83**	1.32	1.58×10^{-1}	A*
Calmodulin	**1.29**	1.13	2.15×10^{-1}	A
Transducer of ERBB-2	**0.81**	0.97	2.58×10^{-1}	A
Beta actin 1	**1.64**	1.22	0	A
Chitinase	**1.30**	**1.26**	2.02×10^{-2}	A ou D
Cathepsin D	**0.84**	0.92	1.56×10^{-1}	A
Cathepsin L	**1.22**	**0.73**	4.57×10^{-3}	TMUT
Similar to Ornithine decarboxylase antizyme	**1.24**	1.04	6.10×10^{-2}	A
High mobility group-T protein	**1.27**	1.08	3.93×10^{-1}	A
Nogo-A	**1.31**	1.08	3.79×10^{-1}	A
Dihydropyrimidine Dehydrogenase, Chain A	**1.19**	0.92	5.98×10^{-1}	A*
Cofilin	**1.21**	**1.51**	6.15×10^{-3}	TMOT

If most differences in gene transcription previously observed between wild and farmed salmon were reduced when comparing hybrid and pure wild salmon, where do the differences observed in the present study come from? Table 1 presents data for the 24 different known genes which showed significant gene transcription difference between hybrid and wild salmon under a Bonferroni-corrected threshold. For comparison, we included fold change data from Roberge *et al.* (2006) in the 17 cases for which the genes were also represented on the earlier array. Strikingly, control of gene transcription by additive gene interactions was compatible with only 12% of the cases while most (82%) of the gene transcription differences observed between hybrids and wild salmon was transgressive (either TMUT or TMOT).

Gene-transcription-level differences obtained with qRT-PCR assays were highly significant for the two candidate genes tested and concordant the microarray results (Fig. 2), although with different average fold changes (metallothionein: 21% under-expression from the microarray assays and

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

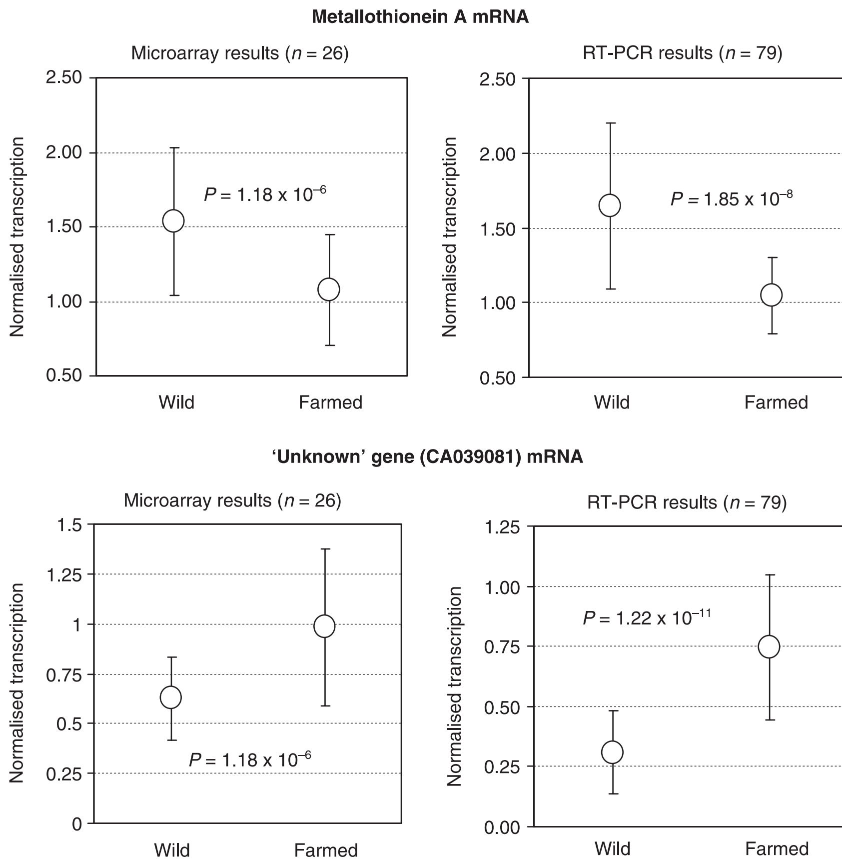

Fig. 2 Quantitative real-time polymerase chain reaction (qRT-PCR) verification of previous results for two differentially expressed genes (metallothionein A-coding gene and the 'unknown' gene cDNA clone CA039081) between the progeny of wild and farmed Canadian salmon. Microarray results ($n = 26$) for those two candidates are presented on the left and qRT-PCR results ($n = 79$) on the right. Dots indicate the normalized average transcription level. Bars indicate the standard deviation of the normalized gene transcription levels. For metallothionein microarray results, results from the most significantly differentially expressed cDNA clone are presented (six metallothionein-A clones showed significant gene transcription difference in the same direction between farmed and wild salmon).

36% under-expression with the qRT-PCR; CA039081: 58% over-expression from the microarray assays and 141% over-expression expression with the qRT-PCR). Within-group standard deviation was comparable between the microarray and qRT-PCR results (Fig. 2), even though sample size was smaller for the microarray experiment ($n = 26$ vs. $n = 79$). This suggests that the microarray platform and protocols we used produced very reproducible results. Yet, on a larger number of biological samples, qRT-PCR provides more statistical power to detect differences. It is therefore useful as a validation technique, in particular in cases where the significance of the microarray observed differences is not conservatively high.

Discussion

Many genes ($n = 298$) with highly significant transcription-level differences between the progeny of Norwegian wild salmon and that of a second generation hybrid backcross of farmed to wild salmon were identified here. Over-representation of genes implicated in cytoskeleton/cellular organization, muscle development, oxidative phosphorylation and extracellular components (collagen matrice, antigen presentation) among those genes suggests that important physiological and morphological differences may well be persisting between wild and hybrid farmed salmon, even though genetic differences for most phenotypic traits tested were shown to be milder between backcross hybrids and wild than between farmed and wild salmonids (McGinnity *et al*. 2003; Hindar *et al*. 2006; Tymchuk *et al*. 2006).

The proportion and number of differentially expressed genes was substantially higher here than observed between farmed and wild Norwegian salmon in Roberge *et al*. (2006), even when correcting for the increased statistical power due to larger sampling. This observation is somewhat analogous to that of Gibson *et al*. (2004), where 33% of the genes showed significant transcription differences between a hybrid *Drosophila* strain and its parental isogenic lines, while only 25% of the genes showed significant gene transcription differences between the two parental strains. However, we cannot rule out the possibility that other factors could have increased statistical power in the present study [e.g. pseudo-replication (as discussed above) or an increased constancy in the cDNA printing process, reducing experimental variance]. We thus conservatively conclude that the number and proportion of genes showing significant gene-expression differences between hybrid backcross and wild salmon are at least as important as observed in the comparison of pure wild and farmed salmon of the same strain.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

The range and average magnitude of the significant differences observed here between the hybrid and wild salmon were strikingly higher than previously observed between pure farmed and wild individuals. Differences of this magnitude are comparable to those observed between wild salmon from Norway and Canada (Roberge *et al.* 2006), except that in this case, a large proportion of the gene expression profiles showed nonhomogenous variances, which was surprisingly not the case in the present study. We interpret these results as evidence for the prevalence of nonadditive genetic control of gene transcription (see below). However, the fact that more genes were represented on the microarray used in this study could have increased the odds of observing extreme differences and hence be responsible for part of the observed results. Yet, the observation of high fold differences for genes which were also represented on the earlier microarray platform (see Table 1) does not support this possibility. For instance, thioredoxin-like protein-19-coding gene was represented on both microarrays and showed highly significant transcriptional change of high magnitude (3.49 fold) in the present study but no significant difference in Roberge *et al.* (2006).

A majority (67%) of the genes which had shown significant transcription differences between farmed and wild salmon (Roberge *et al.* 2006) showed smaller or nonsignificant differences here, which concords with additivity of their gene-transcription levels. However, the vast majority of the genes which showed significant differences in the present study (82%) had average normalized gene-transcription levels either higher or lower than those of both pure farmed and wild salmon, which suggests a nonadditive genetic control of gene transcription. This observation indicates that the genetic basis of transcriptional control is likely to have been modified through artificial selection and domestication for substantially more genes than we previously detected when comparing pure strains only (Roberge *et al.* 2006). A hypothetical explanation for the high prevalence of new gene-transcription changes with a nonadditive genetic basis in hybrids is that pleiotropic changes may have had balancing effects on the gene-transcription levels of several genes in farmed salmon. This balance would have been lost in segregating hybrids, revealing additional gene-transcription differences. Overall, these results support the view that the genetic architecture of gene transcription in farmed salmon has rapidly evolved under artificial selection, such that genetic interactions between farmed and wild salmon may generate unpredicted phenotypes by inducing novel patterns of gene expression.

The result that only a minority of the gene-transcription differences observed between hybrid backcross and wild salmon showed transmission patterns compatible with additive gene interaction has also been reported in studies performed on other organisms (*Drosophila*: Gibson *et al.* 2004, < 2% additivity of gene transcription; *Arabidopsis*: Vuylsteke *et al.* 2005, 27–37.5% additivity; maize: Auger *et al.* 2005, 37% additivity; oyster: Hedgecock *et al.* 2007, 2% additivity). In contrast, other studies have observed a majority of transcription profiles under additive genetic control (maize: Swanson-Wagner *et al.* 2006, 78% additivity; mice: Cui *et al.* 2006). At the level of morphological or physiological traits, additivity is generally admitted as being the norm, yet nonadditive interactions resulting in either heterosis or outbreeding depression are nonetheless widespread (Lippman & Zamir 2007). In natural conditions, Atlantic salmon farmed X wild hybrids (F1 and backcross) were shown to have phenotypes intermediate to their wild and farmed parental lineages for several life history traits (growth, survival and parr maturity rates), suggesting that the genetic basis for these traits is mainly additive (McGinnity *et al.* 2003). These results were also confirmed for other salmonids in laboratory studies where farmed and wild lineages of rainbow trout (*Oncorhynchus mykiss*; Tymchuk & Devlin 2005) and Coho salmon (*O. kisutch;* Tymchuk *et al.* 2006) were crossed and the various hybrids compared to their parental lineages for several traits. On the other hand, nonadditive genetic interactions resulting in heterosis and outbreeding depression has also been reported for salmonids (Leary *et al.* 1983; Gharrett *et al.* 1999; Gilk *et al.* 2004). As pointed out by Gibson *et al.* (2004), most gene-transcription profiles appear as nonadditive traits while most traditional phenotypic traits (morphological, behavioural, physiological) appear as additive traits. This suggests that the apparent additivity at the phenotypic level might result from complex interactions between transcripts whose expression is controlled in a nonadditive manner. To clarify this important issue, further studies should aim at establishing a causal link between transgressive phenotypes (including heterosis and heterozygote depression) and gene transcription, as exemplified by Hedgecock *et al.* (2007) and Meyer *et al.* (2007). Such studies would also allow a better understanding of the potential fitness consequences of nonadditive control of gene transcription in post-F1-farmed X wild salmon hybrids.

The interbreeding of escaped farmed salmon with wild individuals is considered a serious threat to natural populations (McGinnity *et al.* 2003). A pressing question is therefore to what extent this interbreeding can be expected to modify the genetic make-up of wild Atlantic salmon populations. Here, we compared the genome-wide gene-transcription profiles of wild salmon and second generation hybrids [backcross: (farmed X wild) X wild]. Strikingly, we observed more and larger differences than we had previously observed between the progeny of pure farmed and wild genitors (Roberge *et al.* 2006). This suggests that interbreeding of fugitive farmed salmon and wild individuals could substantially modify the genetic control of gene

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

transcription in natural populations exposed to high migration from fish farms, resulting in unpredictable and potentially detrimental effects on the survival of these populations. Moreover, in the absence of strict policies to reduce the number of escaped farmed salmon considerably, this number can be expected to increase due to the rapid development of the aquaculture industry, further modifying the genetic makeup of natural populations. Finally, sea farming is developing rapidly for other marine fish species (e.g. Atlantic cod and halibut, see Naylor *et al*. 2005) and similar risks could eventually threaten natural populations of those species as well. Results from the present study thus further support the idea that measures to considerably reduce the number of escaped farmed salmon and their reproduction in the wild are urgently needed (e.g. McGinnity *et al*. 2003; Naylor *et al*. 2005).

Acknowledgements

This work was supported by a research grant from Science and Engineering Research Canada (NSERC strategic programme) to L.B. and H.G., a postgraduate scholarship from the *Fonds Québécois de la Recherche sur la Nature et les Technologies* (FQRNT) to C.R. and funding from the Research Council of Norway to S.E. We also thank Tom Smith, Sébastien Renaut and two anonymous referees for their constructive comments. This study is a contribution to the research programmes of CIRSA (*Centre Inter-universitaire de Recherche sur le Saumon Atlantique*). We thank the staff at the NINA Research Station, Ims, Norway, for their assistance in rearing fish.

References

Anderson E (1949) *Introgressive Hybridization*. J. Wiley, New York.

Auger DL, Gray AD, Ream TS *et al*. (2005) Nonadditive gene expression in diploid and triploid hybrids of maize. *Genetics*, **169**, 389–397.

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science*, **296**, 752–755.

Cui XQ, Affourtit J, Shockley KR, Woo Y, Churchill GA (2006) Inheritance patterns of transcript levels in F-1 hybrid mice. *Genetics*, **174**, 627–637.

Draghici S (2003) *Data Analysis Tools for DNA Microarrays*. CRC Press, London.

Einum S, Fleming IA (2000a) Highly fecund mothers sacrifice offspring survival to maximise fitness. *Nature*, **405**, 565–567.

Einum S, Fleming IA (2000b) Selection against late emergence and small offspring in Atlantic salmon (*Salmo salar*). *Evolution*, **54**, 628–639.

Einum S, Hendry AP, Fleming IA (2002) Egg-size evolution in aquatic environments: does oxygen availability constrain size? *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **269**, 2325–2330.

Fiske P, Lund RA, Østborg GM, Fløystad L (2001) Escapees of reared salmon in coastal and riverine fisheries in the period 1989–2000. *NINA Oppdragsmelding*, **704**, 1–26 (in Norwegian, English summary).

Fleming IA, Einum S (1997) Experimental tests of genetic divergence of farmed from wild Atlantic salmon due to domestication. *ICES Journal of Marine Science*, **54**, 1051–1063.

Fleming IA, Hindar K, Mjølnerød IB, Jonsson B, Balstad T, Lamberg A (2000) Lifetime success and interactions of farm salmon invading a native population. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **267**, 1517–1524.

Fleming IA, Agustsson T, Finstad B, Johnsson JI, Bjornsson BT (2002) Effects of domestication on growth physiology and endocrinology of Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences*, **59**, 1323–1330.

Gharrett AJ, Smoker WW, Reisenbichler RR, Taylor SG (1999) Outbreeding depression in hybrids between odd- and even-broodyear pink salmon. *Aquaculture*, **173**, 117–129.

Gibson G, Riley-Berger R, Harshman L *et al*. (2004) Extensive sex-specific nonadditivity of gene expression in Drosophila melanogaster. *Genetics*, **167**, 1791–1799.

Gilk SE, Wang IA, Hoover CL *et al*. (2004) Outbreeding depression in hybrids between spatially separated pink salmon, Oncorhynchus gorbuscha, populations: marine survival, homing ability, and variability in family size. *Environmental Biology of Fishes*, **69**, 287–297.

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

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

Hedgecock D, Lin JZ, Decola S *et al*. (2007) Transcriptomic analysis of growth heterosis in larval Pacific oysters (Crassostrea gigas). *Proceedings of the National Academy of Sciences, USA*, **104**, 2313–2318.

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

Hurlbert SH (1984) Pseudoreplication and the design of ecological field experiments. *Ecological Monographs*, **54**, 187–211.

Kerr MK, Martin M, Churchill GA (2000) Analysis of variance for gene expression microarray data. *Journal of Computational Biology*, **7**, 819–837.

Kerr MK, Afshari CA, Bennett L *et al*. (2002) Statistical analysis of a gene expression microarray experiment with replication. *Statistica Sinica*, **12**, 203–217.

Leary RF, Allendorf FW, Knudsen KL (1983) Consistently high meristic counts in natural hybrids between brook trout and bull trout. *Systematic Zoology*, **32**, 369–376.

Lippman ZB, Zamir D (2007) Heterosis: revisiting the magic. *Trends in Genetics*, **23**, 60–66.

McGinnity P, Prodohl P, Ferguson K *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. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **270**, 2443–2450.

Metcalfe NB, Huntingford FA, Graham WD, Thorpe JE (1989) Early Social-Status and the Development of Life-History Strategies in Atlantic Salmon. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, **236**, 7–19.

Meyer S, Pospisil H, Scholten S (2007) Heterosis associated gene expression in maize embryos 6 days after fertilization exhibits additive, dominant and overdominant pattern. *Plant Molecular Biology*, **63**, 381–391.

Millar RB, Anderson MJ (2004) Remedies for pseudoreplication. *Fisheries Research*, **70**, 397–407.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd

Naylor R, Hindar K, Fleming IA *et al*. (2005) Fugitive salmon: Assessing the risks of escaped fish from net-pen aquaculture. *Bioscience*, **55**, 427–437.

Roberge C, Einum S, Guderley H, Bernatchez L (2006) Rapid parallel evolutionary changes of gene transcription profiles in farmed Atlantic salmon. *Molecular Ecology*, **15**, 9–20.

Roberge C, Paez DJ, Rossignol O *et al*. (2007) Genome-wide survey of the gene expression response to saprolegniasis in Atlantic salmon. *Molecular Immunology*, **44**, 1374–1383.

Rye M, Gjerde B (1996) Phenotypic and genetic parameters of body composition traits and flesh colour in Atlantic salmon, *Salmo salar* L. *Aquaculture Research*, **27**, 121–133.

von Schalburg KR, Rise ML, Cooper GA *et al*. (2005) Fish and chips: Various methodologies demonstrate utility of a 16 006-gene salmonid microarray. *BMC Genomics*, **6**, 126–133.

Storey JD (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, **64**, 479–498.

Swanson-Wagner RA, Jia Y, DeCook R *et al*. (2006) All possible modes of gene action are observed in a global comparison of gene expression in a maize F-1 hybrid and its inbred parents. *Proceedings of the National Academy of Sciences, USA*, **103**, 6805–6810.

Tymchuk WE, Devlin RH (2005) Growth differences among first and second generation hybrids of domesticated and wild rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*, **245**, 295–300.

Tymchuk WE, Biagi C, Withler R, Devlin RH (2006) Growth and behavioral consequences of introgression of a domesticated aquaculture genotype into a native strain of Coho salmon. *Transactions of the American Fisheries Society*, **135**, 442–455.

Vuylsteke M, van Eeuwijk F, Van Hummelen P, Kuiper M, Zabeau M (2005) Genetic analysis of variation in gene expression in Arabidopsis thaliana. *Genetics*, **171**, 1267–1275.

Wilson AC (1976) Gene regulation in evolution. In: *Molecular Evolution* (ed. Ayala FJ), pp. 225–234. Sinauer Associates, Sunderland, MA.

Wolfinger RD, Gibson G, Wolfinger ED *et al*. (2001) Assessing gene significance from cDNA microarray expression data via mixed models. *Journal of Computational Biology*, **8**, 625–637.

This study is part of C. Roberge's doctoral research and É. Normandeau studies in L. Bernatchez's laboratory, which aim to study the genetic basis of recent evolution in Atlantic salmon, both in nature and through artificial selection. S. Einum is interested in various aspects of salmon evolution and conservation, including assessing the effects of interactions between wild and farmed salmon. This study also overlaps H. Guderley's interest in biochemical adaptation to changes in abiotic and biotic factors. The major interests of L. Bernatchez are in the understanding of the patterns and processes of molecular and organismal evolution, as well as their significance to conservation.

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd