Conserv Genet (2015) 16:277–287
DOI 10.1007/s10592-014-0657-1

RESEARCH ARTICLE

Transcriptional profiling of two Atlantic salmon strains: implications for reintroduction into Lake Ontario

Xiaoping He · Chris C. Wilson · Kyle W. Wellband · Aimee Lee S. Houde · Bryan D. Neff · Daniel D. Heath

Received: 27 February 2014 / Accepted: 9 September 2014 / Published online: 20 September 2014

Abstract One of the major challenges facing conservation biology is characterizing the genetic variation underlying adaptation to different environments. Gene expression is the process whereby genomic information is converted into phenotype and quantitative variation in gene expression is linked to phenotypic variation. Identifying gene transcription profiles that provide fitness benefits in specific environments would promote more effective species reintroduction and conservation practices. In this study, we developed a custom oligonucleotide microarray for Atlantic salmon (*Salmo salar*) and used this microarray to measure gene transcription in gill tissue for two Atlantic salmon strains currently being reintroduced into Lake Ontario: LaHave (anadromous) and Sebago (landlocked). We measured gene transcription in juvenile salmon from each strain that had been reared under the same conditions and identified genes differentially expressed between the two strains. We used the normalized transcription data and microsatellite genotype data to partition the variance into effects of selection versus genetic drift. We found that although there was little genetic differentiation ($F_{ST} = 0.038$) between the two strains, 21 genes were significantly differentially expressed between the two strains, and in all cases the difference was consistent with divergence by selection. We use this analysis to predict the Sebago strain will be more likely to be successfully reintroduced, highlighting how the combination of population genetics with gene expression can help to guide reintroduction efforts.

Keywords Microarray · Gene expression · Reintroduction · Conservation · F_{ST} · P_{ST} · Lake Ontario · Salmon

Introduction

The conservation genetics paradigm is that small and isolated populations are subject to loss of genetic diversity and increased levels of homozygosity that in turn lead to increased likelihood of extirpation (Frankham et al. 2002; Ouborg et al. 2010). Loss of genetic diversity is thought to reduce individual fitness and affect the ability of a population to adaptively respond to a changing environment (Frankham 2003; Spielman et al. 2004). Therefore, conserving genetic diversity is often an important component of conservation plans and efforts. However, it is not clear if these efforts actually conserve functional genetic variation.

There are three forms of genetic variation in populations: neutral, deleterious and adaptive (Hedrick 2001). Adaptive genetic variation is variation in coding or regulatory genes that have the potential to increase fitness (Hedrick 2001; Garcia de Leaniz et al. 2007). Thus, using functional genetic variation to address issues in

Electronic supplementary material The online version of this article (doi:10.1007/s10592-014-0657-1) contains supplementary material, which is available to authorized users.

X. He · K. W. Wellband · D. D. Heath (✉)
Great Lakes Institute for Environmental Research, University of Windsor, Windsor, ON, Canada
e-mail: dheath@uwindsor.ca

C. C. Wilson
Aquatic Research and Monitoring Section, Ontario Ministry of Natural Resources, Trent University, Peterborough, ON, Canada

A. L. S. Houde · B. D. Neff
Department of Biology, Western University, London, ON, Canada

conservation should be advantageous in comparison with neutral genetic variation. One form of functional genetic variation is gene expression variation, that is, the process whereby genomic variation is converted into phenotypic variation. One mechanism by which phenotypic variation can arise from a single genotype is regulation in gene expression: this can be either sensitive up- or down-regulation or, more simply, on–off control (Schlichting and Smith 2002). Many studies have suggested that variation in gene expression among populations can be adaptive (Oleksiak et al. 2002; Whitehead and Crawford 2006; Larsen et al. 2007; Luca et al. 2009; Wellband and Heath 2013) and thus could be used to address conservation issues. Compared to neutral genetic variation, measures of variation in gene expression could provide more relevant information as it reflects the activity of functional genes. For example, Giger et al. (2006) profiled gene transcription using DNA microarrays and genotyped microsatellite loci in juveniles from six brown trout populations and found that gene expression variation among populations was more affected by population life history (migratory or residential) than by their genetic distance based on neutral DNA markers.

Population differences in gene expression have been demonstrated in several species over the past decade. Some studies suggest that among-population gene expression variation is much higher than within-population variation (Townsend et al. 2003; Hutter et al. 2008), whereas others have found the opposite pattern (Oleksiak et al. 2002; Storey et al. 2007). However, those studies all indicated that variation in gene expression is an important source of variance for adaptation, and thus ultimately, evolution.

More recently, gene expression comparisons among populations have been applied to conservation. For example, gene transcription comparisons have revealed that introgression can result in changes in gene transcription profiles in both Atlantics salmon and brook charr, which may result in loss of local adaptation (Roberge et al. 2008; Lamaze et al. 2013). Also, analyses of population differences in gene expression versus neutral DNA in 12 Atlantic salmon populations showed that gene transcription can be used to identify conservation units and has many advantages over the more traditional, neutral markers (Vandersteen Tymchuk et al. 2010; Hansen 2010). Pedersen et al. (2005) compared the expression of *heat shock protein 70* (*Hsp70*) between inbred and outbred lines of *Drosophila melanogaster* to illustrate mechanisms of inbreeding depression and found that there was a significant negative correlation between transcription level of *Hsp70* and resistance to heat stress. Miller et al. (2011) collected gill tissue using nonlethal biopsy method from wild-caught Sockeye salmon (*Oncorhynchus nerka*) and identified a set of genes whose transcription can be used to predict migration and spawning success both in fresh water and in the ocean.

Atlantic salmon are broadly distributed in North America and Europe but they have declined or been extirpated in many rivers over the last 200 years (Parrish et al. 1998). Atlantic salmon was once an abundant fish in Lake Ontario (Ontario, Canada), but had disappeared by 1900, mainly because of habitat degradation (Crawford 2001). Because of its economic, ecological and cultural value, there have been increasing efforts to reintroduce Atlantic salmon into Lake Ontario over the past three decades (Dimond and Smitka 2005); however, those reintroduction attempts have been unsuccessful. Potential explanations for the reintroduction failure of Atlantic salmon into Lake Ontario include environmental changes of Lake Ontario during the past years, such as establishment of non-native prey fish and non-native competitors (Coghlan and Ringler 2004; Scott et al. 2005; Houde et al. 2014), and perhaps inappropriate Atlantic salmon strains used for the reintroduction (Van Zwol et al. 2012). Although the Lake Ontario habitat has improved, selecting appropriate source populations is a crucial step for Atlantic salmon reintroduction, as populations differ in their adaptive potential and environmental tolerances. To address possible limitations in candidate strains for reintroduction into Lake Ontario, we explored the genetic background (neutral and transcriptional) of two Atlantic salmon strains: LaHave and Sebago. LaHave is an anadromous strain which originates from the LaHave River, Nova Scotia. The LaHave strain has been used for reintroduction into Lake Ontario for many years and it was successfully reintroduced into Trout Lake, Ontario (Dimond and Smitka 2005). Sebago is a landlocked strain from Sebago Lake, Maine. This strain has a relatively large body size compared to other strains and it performed well in the Lake Champlain reintroduction where salmonid competitors (rainbow trout and brown trout) existed (Dimond and Smitka 2005; Van Zwol et al. 2012). In this study, we constructed a custom oligonucleotide microarray to compare gene transcription at selected known-function genes in gill tissue between the two source populations. We then calculated F_{ST} based on microsatellite genotypes and P_{ST} based on gene transcription levels, and used the F_{ST}-P_{ST} comparison to identify selection versus genetic drift effects on the genes differentially expressed between the two populations. The results demonstrate how populations differ in gene expression and the evolutionary forces underlying those differences. We discussed how our data can be used in conjunction with population genetics to inform Atlantic salmon reintroduction into Lake Ontario.

Materials and methods

Atlantic salmon strains

Two Atlantic salmon strains were provided by the Ontario Ministry of Natural Resources (OMNR): LaHave and Sebago. The LaHave strain was from broodstock that has been in captivity for three generations. The Sebago strain was derived from hatchery-bred fish that were released and recaptured as returning mature fish in Sebago Lake. Eggs and milt were collected from the recaptured adults and brought to Ontario in 2006 and the Sebago strain was reared in captivity to be used as broodstock. Eggs from both strains were fertilized on November 4, 2010 at OMNR Harwood Fish Culture Station, Harwood, Ontario, and then reared at the OMNR Codrington Research Facility, Codrington, Ontario. Detailed information about the families and rearing environment is provided in Houde et al. (2013). Briefly, for each strain, full factorial crosses were conducted using five males and five females to yield 25 full-sib families. Fertilized eggs were incubated in vertical stack incubators followed by rearing in tanks. The fish were transferred to semi-natural stream tanks in September 2011 as juvenile fry. Each semi-natural stream tank consisted of a riffle and a pool. More details about the juvenile salmon and semi-natural stream tank construction are described in Houde et al. (2014). In each semi-natural stream tank, there were a total of 32 Atlantic salmon from eight families of one strain with equal numbers (four fish) per family. Each stream tank was replicated once such that fish from each strain were reared in two stream tanks (for a total of four stream tanks). In July 2012, after 10 months in the semi-natural stream tanks, eight fish from each tank were euthanized by overdose of tricaine methanesulfonate solution and gill tissue was collected and preserved in RNAlater. We chose gill tissue because of its vital function in respiration, osmoregulation, nitrogen balance and disease defenses, and its fast response to environmental stressors (e.g., toxins and pathogens) relative to other organs (Campos-Perez et al. 2000; Evans et al. 2005). At the time of sampling, body mass ranged from 12.76 to 45.71 g with an average body mass (± SE) of 24.24 ± 1.68 g. There were no significant differences in mean body mass between strains or among tanks.

Oligonucleotide microarray construction

To compare transcriptional differences between the two strains, we developed a custom oligonucleotide microarray. Custom microarrays have a few advantages compared to commercial microarrays: relatively low price, higher replication and more focused set of genes. Our custom microarray consisted of probes for 380 different genes: 375 genes from Atlantic salmon and five control genes from *Arabidopsis thaliana*. Of the 375 genes, 277 genes were selected because of their functional importance and their mRNA sequences were obtained from the consortium for Genomics Research on All Salmon Project website (http://web.uvic.ca/grasp/microarray). The sequences of the other 98 genes were downloaded from Nucleotide database of NCBI (http://www.ncbi.nlm.nih.gov) and most of those genes had been reported to show transcriptional response to environmental factors. The five plant genes (*isoflavonoid reductase*, *psbP*, *psbW*, *salt-stress induced tonoplast intrinsic protein*, *ribulose 1.5-biphoshate carboxylase small subunit*) were used as negative control and their sequences were downloaded from NCBI. The probes were designed by using OligoArray 2.0 (Rouillard et al. 2003). The length of probes ranged from 45 to 55 nucleotides and T_m ranged from 82 to 88 °C. A list and detailed information for the selected genes and probe sequences are presented in Supplementary Table S1. The oligonucleotide probes were printed on poly-L-Lysine coated slides (Thermo Scientific, USA) using a SpotArray 24 Microarray Printing System (PerkinElmer, Canada). On each slide, the probes were printed in three blocks (top, middle and bottom) and each probe was printed three times adjacently within each block. Thus each probe was printed nine times on every slide. After printing, the probes were cross-linked to the slides by ultraviolet irradiation. The microarrays used in this experiment were printed in two batches and the potential batch effect caused by different printing was taken into account in the data analysis.

RNA extraction, microarray hybridization and data preparation

Gill tissue was placed in 2 mL tubes containing 1 mL TRIzol (Invitrogen, USA) and approximately 400 mg of 1.0 mm diameter glass beads (BioSpec Products, USA). The tissue samples were homogenized at speed 6 for 40 s in a Thermo Savant FastPrep homogenizer (Lab Recyclers Inc., USA). Total RNA isolation followed the manufacturer's instructions (http://tools.lifetechnologies.com/content/sfs/manuals/trizol_reagent.pdf). The concentration and purity of RNA was measured by spectrophotometry on a NanoVue spectrophotometer (GE Healthcare Bio-Science Corp, USA), and the quality of RNA was assessed by running 1 μg of total RNA on a 1 % agarose gel. Single colour microarray measurement was performed for this experiment using Array 50^{TM} Cy3 Kit (Genisphere Inc., Hatfield, America). Detailed protocols for reverse transcription, cDNA concentration, hybridization and washing are given in the Array 50^{TM} Cy3 Kit's instruction (http://genisphere.com/sites/default/files/pdf/Array50_Jan2011.pdf). Briefly, 15–20 μg total RNA was reverse transcribed using

SuperScript® II Reverse Transcriptase (Invitrogen) and RT primer (5′ - TTCTCGTGTTCCGTTTGTACTCTAAGGTG GA–T(17)- 3′). The cDNA was concentrated and hybridized to microarrays for 12 h at 43 °C. The slides were subsequently washed using 2X SSC with 0.2 % SDS, 2X SSC and 0.2X SSC, separately. The slides were centrifuged immediately for 2 min at 1,000 RPM to dry. The slides were then hybridized with Cy3-labeled fluorescent DNA dendrimer for 2.5 h at 43 °C. The slides were washed and dried again as described above, then the slides were immediately scanned using a ScanArray Express microarray scanner (PerkinElmer, Canada) with the laser at 90 % power and photo-multiplier tube (PMT) Gain at 75 %.

The scanned images were analyzed using ScanArray Express Microarray Analysis System software version 4.0 (PerkinElmer, Canada). Each spot was quantified using the adaptive circle method and the three blocks on each slide were quantified separately. After quantification, the data were background corrected and normalized using limma package of R (Smyth 2005). First, the spots which failed to meet the quality criteria were filtered out. Then, "normexp" algorithm with an offset of 50 was used for background correction. After that, "quantile" normalization method was used to conduct between-array normalization. Finally, genes that had expression data in less than 70 % of the spots across all samples were removed. The intensity of fluorescence for the remaining genes was $\log_2$ transformed for statistical analyses.

Statistical analysis

To detect semi-natural stream tank effects on gene expression, we analyzed the data strain by strain as tanks were nested in each strain in our experiment. The analysis was conducted using the lme4 package of R (Bates and Maechler 2009) with the following model:

$$Y_{ijklm} = \mu + T_i + Ba_j + I_k + Bl_{l(k)} + e_{ijklm} \quad (1)$$

where Y_{ijklm} is the $\log_2$ transformed normalized intensity value for each spot; μ is the average value; T_i is the *ith* effect of tank; Ba_j is the *jth* effect of printing batch; $Bl_{l(k)}$ is the *lth* block effect (position on the array) which is nested within the *kth* individual (fish) and e_{ijklm} is the random residual. Significance of expression differences between replicate tanks (within strain) was determined using a likelihood ratio test between two models: one with and one without the tank effect included. Due to the complicated nature of our mixed-effects model and the dependency structure of our genes, resampling based False Discovery Rate (FDR) corrections are not supported. In lieu of these, we calculated the probability of detecting a *P* value as extreme as the one we observed by randomly permuting the data 10,000 times and refitting the model for each gene to determine its significance under a completely null hypothesis. We report the probability of detecting the gene as significant as the number of times the permutated *P*-values were more extreme (less) than the observed *P*-value for that gene divided by the total number of permutations.

To test for gene transcription differences between the two Atlantic salmon strains (Sebago and LaHave), we used the following model:

$$Y_{ijklmn} = \mu + St_i + T_j + Ba_k + I_l + Bl_{m(l)} + e_{ijklmn} \quad (2)$$

where Y_{ijklmn} is the $\log_2$ transformed normalized intensity value for each spot; μ is the average value; St_i is the *ith* effect of strain; T_j is the *jth* effect of tank; Ba_k is the *kth* effect of array printing batch; $Bl_{m(l)}$ is the *mth* effect of block (position on the array) which is nested within *lth* individual (fish) and e_{ijklmn} is the random residual. The significance of the strain effect was determined using a likelihood ratio test between two models: one with and one without the strain effect included. We followed the same methodology detailed above to calculate the probability of false discovery for the strain effect analyses.

Functional analysis

The functions of those differentially expressed genes were analyzed in NCBI (http://www.ncbi.nlm.nih.gov) and the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Huang et al. 2008). The transcriptional level for the differentially expressed genes was averaged within each tank and then the data was used to construct a heat map using TM4 software (Saeed et al. 2003).

Microsatellite genotyping and F_{ST} estimation

To calculate neutral F_{ST} between the two stains, we used microsatellite genotype data for a total of 520 fish, of which 219 were collected from the Sebago strain in Sebago Lake, and 301 were collected from LaHave strain in Harwood Hatchery. Each fish was genotyped at eight microsatellite loci: *Ssa*197, *Ssa*202, *SSsp*1605, *SSsp*2201, *SSsp*2213, *SSsp*2215, *SSsp*2216 and *SSsp*G7 (O'Reilly et al. 1996; Paterson et al. 2004). Detailed information about PCR protocols are described in Bobrowski (2010). Briefly, the 8 microsatellite were amplified in 5 PCR reactions using fluorescent-labeled primers. The PCR products were run on an AB3730 DNA Analyzer (Applied Biosystems, USA) and the genotypes were analyzed using

GeneMapper version 3.1 (Applied Biosystems, USA). F_{ST} and its 99 % confidence interval was estimated using Fstat version 2.9.3.2 (Goudet 1995). The sample size for the F_{ST} estimate (520 fish) is much larger than that of our P_{ST} estimate (31 fish). To account for potential bias caused by sample size, we randomly selected microsatellite genotypes for 16 fish from each strain (n = 32 fish total) to calculate F_{ST} and replicated this analysis 1,000 times using the pegas package of R (Paradis 2010).

P_{ST} estimation

P_{ST}, the phenotypic analogue of Q_{ST}, is a measurement of phenotypic differentiation among populations. To calculate P_{ST} for each gene, we used the following model to obtain variance estimates between and within strains using restricted maximum likelihoods (REML) as priors:

$$Y_{ijklm} = \mu + St_i + T_j + Ba_k + Bl_l + e_{ijklm} \quad (3)$$

where Y_{ijklm} is the $\log_2$ transformed normalized intensity value for each spot; μ is the average value; St_i is the *ith* effect of strain; T_j is the *jth* effect of tank; Ba_k is the *kth* effect of batch; Bl_l is the *lth* effect of block and e_{ijklm} is the random residual. We then used the variance estimates to calculate highest probability density (HPD) values with Markov Chain Monte Carlo (MCMC) simulations (10,000 replications) in the languageR package of R (Baayen 2008). The median HPD values were used to calculate P_{ST} as:

$$P_{ST} = \sigma^2_{GB} / \left(\sigma^2_{GB} + 2\sigma^2_{GW}\right) \quad (4)$$

where σ^2_{GB} is median HPD value for the between-strain variance and σ^2_{GW} is median HPD value for within-strain variance.

Results

Tank effect on gene transcription

In total, 271 genes were analyzed as these genes had expression data in more than 70 % of the spots. Nineteen genes showed significant differences between the two replicate semi-natural stream tanks for the Sebago strain (Table 1, Supplemental Fig. S1). Fifteen genes showed significant differences between the two replicate stream tanks for the LaHave strain (Table 2, Supplemental Fig. S2). Among the identified genes, *proprotein convertase subtilisin/kexin type 5* (*pcsk5*) and *tissue metalloproteinase inhibitor 3 precursor* (*timp3*) showed stream tank effects in both strains.

Table 1 List of genes showing significantly different transcription between the two replicate tanks for the Sebago strain of Atlantic salmon (*Salmo salar*)

Gene symbol	Gene product	Intensity ratio (Tank1/Tank2)
cish	Cytokine-inducible SH2-containing protein	0.51
cfb	Complement factor B precursor	0.60
pcsk5	Proprotein convertase subtilisin/kexin type 5 precursor	0.67
pitpna	Phosphatidylinositol transfer protein alpha isoform	0.74
Glns	Glutamine synthetase	0.76
timp3	Metalloproteinase inhibitor 3 precursor	0.76
atl2	Atlastin-2	0.77
psmd5	26S proteasome non-ATPase regulatory subunit 5	0.77
Gem	GTP-binding protein	1.25
xaf1	XIAP-associated factor 1	1.27
psmc2	26S protease regulatory subunit 7	1.30
Mstn	Myostatin 1b	1.33
Crtam	Cytotoxic and regulatory T-cell molecule precursor	1.35
hspa14	Heat shock 70 kDa protein 14	1.37
sod3	Extracellular superoxide dismutase [Cu-Zn] precursor	1.46
hsf2	Heat shock factor protein 2	1.51
sar1a	GTP-binding protein SAR1a	1.53
cdk5	Cell division protein kinase 5	1.60
pcna	Proliferating cell nuclear antigen putative mRNA	1.61

Table 2 List of genes showing significantly different transcription between the two replicate tanks for the LaHave strain of Atlantic salmon (*Salmo salar*)

Gene symbol	Gene product	Intensity ratio (Tank3/Tank4)
sar1b	GTP-binding protein SAR1b	0.71
pcsk5	Proprotein convertase subtilisin/kexin type 5 precursor	0.79
myl6b	Myosin light chain 6B	0.88
irak3	Interleukin-1 receptor-associated kinase 3	1.15
psmd9	26S proteasome non-ATPase regulatory subunit 9	1.20
isca2	Iron-sulfur cluster assembly 2 homolog, mitochondrial precursor	1.26
atp1a1	Sodium/potassium-transporting ATPase subunit alpha-1 precursor	1.26
tmp49	Transmembrane protein 49	1.27
ptgd2	Glutathione-requiring prostaglandin D synthase	1.27
cdk9	Cell division protein kinase 9	1.30
timp3	Metalloproteinase inhibitor 3 precursor	1.31
il4	Interleukin 4/13A (il4/13a)	1.42
ctsh	Cathepsin H precursor	1.42
pgd	6-phosphogluconate dehydrogenase, decarboxylating	1.54
c7	Complement C7 precursor	1.54

Table 3 List of genes showing significantly different transcription between the Sebago and LaHave strains of Atlantic salmon (*Salmo salar*)

Gene	Gene product	Intensity ratio (Sebago/LaHave)
timp2	TIMP Metalloproteinase inhibitor 2 precursor	0.72
myl3	Myosin light chain 3	0.76
glns	Glutamine synthetase	0.76
tbl1xr1b	F-box-like/WD repeat-containing protein TBL1XR1-B	0.77
cyp3a27	Cytochrome P450 3A27	0.79
cytl1	Cytokine-like protein 1 precursor	0.79
cyp2f5	Cytochrome P450 2F5	0.79
hyal2	Hyaluronidase-2	0.81
pded	Phosphodiesterase delta-like protein	1.18
pgm2	Phosphoglucomutase-2	1.20
grn	Granulins precursor	1.21
myl6b	Myosin light chain 6B	1.22
fah	Fumarylacetoacetate hydrolase	1.23
tnr5	Tumor necrosis factor receptor superfamily member 5	1.25
hmox	Heme oxygenase	1.26
tcrb	T-cell receptor beta chain	1.28
myl1	Myosin light chain 1, skeletal muscle isoform	1.33
il1r2	Interleukin-1 receptor type II precursor	1.44
c1qc	Complement C1q subcomponent subunit C precursor	1.47
srk2tk	SRK2 tyrosine kinase	1.70
saa5	Serum amyloid A-5 protein	1.83

Differentially expressed genes between strains

Twenty-one genes showed significantly different transcription between LaHave and Sebago strains, which accounts for 7.75 % of the genes examined (Table 3, Supplemental Fig. S3). Of the genes that were differentially transcribed between strains, 13 showed higher transcription levels in Sebago than that in LaHave whereas the other eight genes showed higher transcription levels in LaHave. Of these differentially expressed genes, *glutamine synthetase* (*glns*) and *myosin light chain 6B* (*myl6b*) were also affected by tank effect. Hierarchical clustering based on the 21 differentially

expressed genes showed that the two replicate tanks within each strain clustered together (Fig. 1). The differentially expressed genes have diverse functions: nine genes (*cyp3a27*, *cyp2f5*, *fah*, *glns*, *hmox*, *hyal2*, *pded*, *pgm2* and *srk2tk*) encode enzymes; five genes (*clqc*, *il1r2*, *saa5, tcrb* and *tnr5*) are involved in the immune response; five genes (*myl1*, *myl3*, *myl6b*, *fah*, *pgm2*) are involved in ion binding; two genes (*cytl1* and *tbl1xr1b*) regulate transcription activity and one gene (*grn*) regulates cell growth.

F_{ST} and P_{ST}

The F_{ST} value based on the microsatellite genotypes of all 520 fish was 0.038, with a 99 % confidence interval of 0.020–0.057. The mean F_{ST} value (± SD) based on the randomly sub-sampled microsatellite genotypes of 32 fish was 0.037 ± 0.008. Although the F_{ST} estimates based on the randomly sub-sampled fish ranged from 0.013 to 0.082, 97.5 % (975 out of 1,000 times) of the estimates were within the 99 % confidence interval of the mean F_{ST} value based on all 520 fish. This result indicates that F_{ST}-P_{ST} comparison is not likely biased due to estimates based on different sample size.

We tested for the effects of selection versus genetic drift as contributors to the difference in gene transcription levels by comparing P_{ST} for each gene with the F_{ST} confidence interval. The P_{ST} for the 271 analyzed genes ranged from 0.034 to 0.32 (Fig. 2a). The P_{ST} for the 21 differentially expressed genes between strains ranged from 0.20 to 0.32 (Fig. 2b), all of which were substantially outside the 99 % confidence interval for the F_{ST} value reported above. We therefore conclude that the strain difference in transcription for these 21 genes is primarily driven by selection.

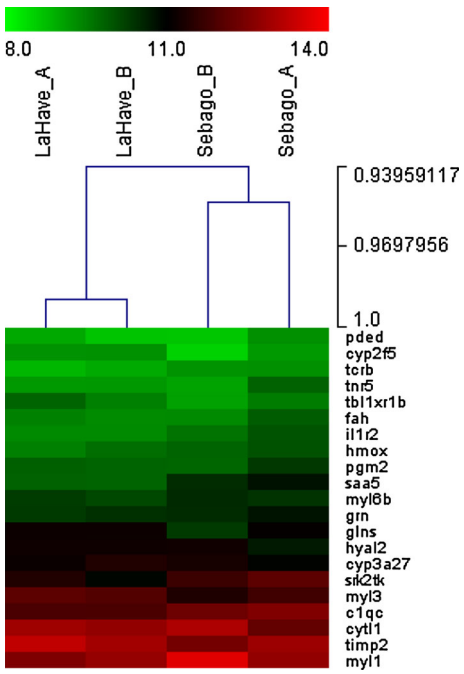

Fig. 1 Gene transcription heat map showing hierarchical clustering of the 21 differentially expressed genes between Sebago and LaHave Atlantic salmon strains. The rows represent different genes and the columns represent different tanks. The transcription level for each gene is the average $\log_2$ transformed intensity value of fish from the same tank

Discussion

Gene transcription data has only recently been used in the study of population genetics. Comparing gene transcription profiles across populations does have important applications in conservation and management (Vandersteen Tymchuk et al. 2010). In this experiment, we compared transcription between two Atlantic salmon strains (LaHave and Sebago), and found that about 8 % of analyzed genes were differentially expressed between the two strains, despite being held in identical semi-natural environments. This percentage is higher than that reported in two similar studies (1.4 and 1.7 %: Roberge et al. 2006; and 2.3 %: Debes et al. 2012) on microarray gene transcription comparisons between farmed and wild Atlantic salmon. There are three possibilities for our higher frequency of transcriptional differences. First, our custom microarray was enriched for genes that are known to be sensitive to environmental differences. Second, the two strains we compared have different evolutionary histories and marked life history differences. Third, we used an oligonucleotide microarray which may be more sensitive than the cDNA microarrays used in other studies (Yauk et al. 2004).

Like other quantitative traits, gene expression is determined by a combination of genetic and environmental effects, thus it is not surprising that we detected both tank and strain effects. Two genes, *psck5* and *timp3*, were significantly affected by tank effects in both strains, while the majority of genes differentially expressed between tanks showed difference in only one strain. Thus minor environmental differences among tanks affected the two strains differently, likely a reflection of genotype by environmental interactions (G × E) on gene transcription. This is despite our attempts to control many environmental factors, for example: the fish were crossed on the same day and reared under the same food and water source and flow regimes, plus we sampled them at the same developmental stage and used identical protocols to measure gene transcription. Nevertheless, stream tank effects contributed to differences in gene expression, and the number of genes and magnitude of differences between tanks was similar to the strain effect. As our design had the stream tanks nested within strains, we are unable to specifically partition

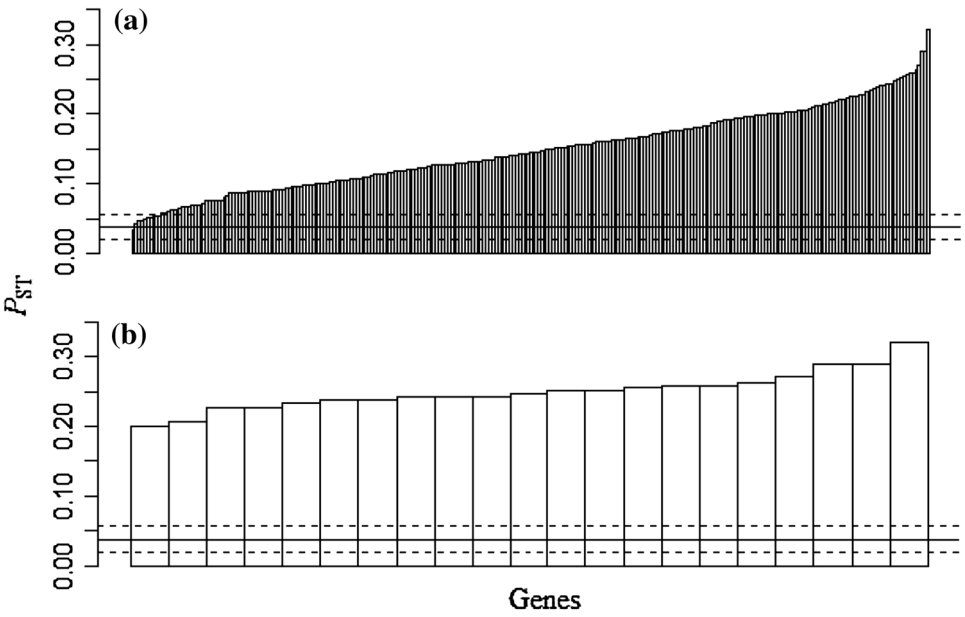

Fig. 2 Histograms of global P_{ST} for transcription of genes between two strains of juvenile Atlantic salmon (*Salmo salar*) arranged in increasing order. Panel a: P_{ST} values for all 271 analyzed genes. Panel b: P_{ST} values for genes which showed significantly different transcription levels between the two strains. The two *horizontal dashed lines* represent the *upper* and *lower* limits of the 99% confidence interval of F_{ST} based on microsatellite genotypes at eight loci. The *solid line* represents the mean F_{ST}

G × E effects, however, previous studies have shown that G × E contributes to transcriptional variation (Smith and Kruglyak 2008; Grishkevich and Yanai 2013). Although we cannot definitively conclude that the stream tank effects reflected G × E, the transcription differences indicate high environmental sensitivity in these fish, perhaps reflecting why reintroduction may succeed in one habitat but fail in another using the same donor stock.

Neutral microsatellite DNA markers have been widely used in conservation genetics over the past two decades under the assumption that the extent of neutral genetic variation is positively correlated with the genome-wide functional genetic variation-this assumption has been called into question in a number of studies (Hedrick 2001; Reed and Frankham 2001). In this study, we found that there was little neutral genetic differentiation ($F_{ST} = 0.038$) between the two strains based on microsatellite genotypes, but genes involved in known and vital functions showed significant differences between the two strains. For example, two genes (*cyp3a27* and *cyp2f5*) encoding cytochrome P450 (CYP) enzymes had higher transcription in the anadromous strain (LaHave); those enzymes play an important role in metabolism of steroids and fatty acids and detoxification of pollutants and drugs (Uno et al. 2012). Similarly, *glns* which encodes glutamine synthetase (catalyzes ammonia and glutamate to synthesize glutamine) also showed higher transcription levels in the LaHave strain. The conversion of ammonia to glutamine is a mechanism to remove ammonia and thus avoid its toxicity (Essex-Fraser et al. 2005). The higher transcription of the CYP and *glns* genes may be adaptive for the anadromous strain as part of their preparation for the novel marine environment. Similar migratory preparation was also reported in Giger et al. (2008) where 17 genes related to migratory adaptation differentially expressed between migratory and non-migratory brown trout populations. In contrast, we found 5 immune-related genes (*tnr5*, *tcrb*, *il1r2*, *clqc* and *saa5*) had higher transcription levels in the Sebago relative to the LaHave strain. The *tnr5* gene encodes a member of the tumor necrosis factor receptor superfamily and the interaction between the receptor and its ligand plays a crucial role in expression regulation of many immune molecules, such as cytokines and chemokines (Chatzigeorgiou et al. 2009). The *tcrb* gene encodes the β chain of the T cell receptor in αβ T cells which recognizes foreign antigens that are bound by major histocompatibility complex molecules (Goldrath and Bevan 1999). The *il1r2* gene encodes interleukin 1 receptor 2 which binds and inhibits interleukin 1 activity (Colotta et al. 1993). The *c1qc* gene encodes the C-chain of complement subcomponent C1q. C1q is the recognition subunit of the C1 complex and is able to recognize and bind a variety of targets to activate the complement pathway to defense pathogens (Gaboriaud et al. 2004). The *saa5* gene encodes an acute phase protein which is involved in the inflammatory response and lipid transportation (Banka et al. 1995; Goltry et al. 1998), and this gene is known to be up-regulated after bacterial and viral infection (Miwata et al. 1993; Lin et al. 2007). The different transcription of those immune genes is related to coping with pathogens in their environments, which is vital for salmonid survival in the wild (Miller et al. 2011). Previous reintroduction of Atlantic salmon into Lake Ontario focused on LaHave strain which has already been identified as a possibly inappropriate strain (Van Zwol et al. 2012), and as our data show, the LaHave strain, while showing higher expression of CYP genes, appears to have lower expression at selected immune genes. Thus our transcriptional profiling of

functionally important genes shows that not only is gene expression variation more divergent between the strains than expected based on drift (neutral DNA) but that the Sebago Atlantic salmon strain may be a better choice for reintroduction into Lake Ontario.

Identifying genes with transcription profiles that indicate selection-based differences among populations is important in conservation and management as such differences likely underlie adaptations to different environmental conditions. In our study, the P_{ST} values calculated were comparable to P_{ST} and Q_{ST} values for transcription in rainbow trout (Aykanat et al. 2011; Wellband and Heath 2013), but much higher than the Q_{ST} values estimated in two Atlantic salmon subpopulations (Roberge et al. 2007), which implies that differences in gene transcription among populations depend on the extent of divergence. Our results showed that much of the difference in gene transcription between the two strains of Atlantic salmon was consistent with divergence by selection. Moreover, the genes identified as driven by directional selection are excellent candidate markers for predicting fitness in specific environments.

Although the application of gene transcription in conservation biology is still in its infancy, transcriptional profiling of potential source populations can enhance reintroduction efforts in two ways: first, gene expression comparisons can identify functional differences that are related to important physiological processes and responses to environmental stressors, and subsequently, variation in individual gene transcription can be used to predict specific trait response upon reintroduction (Miller et al. 2011). The custom DNA microarray we developed provides a relatively inexpensive method to profile transcription for many individuals that will make it possible to choose appropriate source populations for reintroduction. Such an approach will increase the likelihood of reintroduction success and ultimately, conservation. Furthermore, as more such studies are completed, and our understanding of the role of specific gene expression responses in adaptive environmental stress responses improves, the application of transcriptional profiling will expand.

Acknowledgments This research was funded by Natural Sciences and Engineering Council of Canada and Ontario Ministry of Natural Resources. We thank W. Sloan, S. Ferguson, B. Lewis, A. Smith, C. Black, H. Dokter, J. Van Zwol, S. Garner, T. Hain, M.-C. Bellmare, M.-H. Greffard and H. Allegue for their assistance in semi-natural stream construction, fish feeding, sample collection and other support at the OMNR Codrington Research Facility. We thank R. Hepburn for microarray printing. We thank A. Kidd for microsatellite genotyping.

References

Aykanat T, Thrower FP, Heath DD (2011) Rapid evolution of osmoregulatory function by modification of gene transcription in steelhead trout. Genetica 139:233–242. doi:10.1007/s10709-010-9540-2

Baayen RH (2008) Analyzing linguistic data: a practical introduction to statistics using R. Cambridge University Press, Cambridge, UK, New York

Banka CL, Yuan T, de Beer MC et al (1995) Serum amyloid A (SAA): influence on HDL-mediated cellular cholesterol efflux. J Lipid Res 36:1058–1065

Bates D, Maechler M (2009) lme4: Linear mixed-effects models using {S4} classes.{R} package version 0.999375-32

Bobrowski R (2010) Survival, growth, and out-migration timing of reintroduced Atlantic salmon (*Salmo salar*) in Cobourg Brook, Ontario. M.Sc., Trent University (Canada)

Campos-Perez JJ, Ward M, Grabowski PS et al (2000) The gills are an important site of iNOS expression in rainbow trout *Oncorhynchus mykiss* after challenge with the gram-positive pathogen Renibacterium salmoninarum. Immunology 99:153–161

Chatzigeorgiou A, Lyberi M, Chatzilymperis G et al (2009) CD40/CD40L signaling and its implication in health and disease. BioFactors Oxf Engl 35:474–483. doi:10.1002/biof.62

Coghlan SM, Ringler NH (2004) Comparison of Atlantic salmon embryo and fry stocking in the salmon river, New York. North Am J Fish Manag 24:1385–1397. doi:10.1577/M03-253.1

Colotta F, Re F, Muzio M et al (1993) Interleukin-1 type II receptor: a decoy target for IL-1 that is regulated by IL-4. Science 261:472–475

Crawford SS (2001) Salmonine introductions to the Laurentian Great Lakes: an historical review and evaluation of ecological effects. NRC Research Press, Ottawa

Debes PV, Normandeau E, Fraser DJ et al (2012) Differences in transcription levels among wild, domesticated, and hybrid Atlantic salmon (*Salmo salar*) from two environments. Mol Ecol 21:2574–2587. doi:10.1111/j.1365-294X.2012.05567.x

Dimond P, Smitka J (2005) Evaluation of selected strains of Atlantic salmon as potential candidates for the restoration of Lake Ontario. Trout Unlimited Canada Technical Report ON-12. Guelph, Ontario: Trout Unlimited Canada

Essex-Fraser PA, Steele SL, Bernier NJ et al (2005) Expression of four Glutamine Synthetase Genes in the Early Stages of Development of Rainbow Trout (*Oncorhynchus mykiss*) in Relationship to Nitrogen Excretion. J Biol Chem 280:20268–20273. doi:10.1074/jbc.M412338200

Evans DH, Piermarini PM, Choe KP (2005) The multifunctional fish gill: dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of nitrogenous waste. Physiol Rev 85:97–177. doi:10.1152/physrev.00050.2003

Frankham R (2003) Genetics and conservation biology. C R Biol 326(Supplement 1):22–29. doi:10.1016/S1631-0691(03)00023-4

Frankham R, Briscoe DA, Ballou JD (2002) Introduction to Conservation Genetics. Cambridge University Press, Cambridge

Gaboriaud C, Thielens NM, Gregory LA et al (2004) Structure and activation of the C1 complex of complement: unraveling the puzzle. Trends Immunol 25:368–373. doi:10.1016/j.it.2004.04.008

Garcia de Leaniz C, Fleming IA, Einum S et al (2007) A critical review of adaptive genetic variation in Atlantic salmon: implications for conservation. Biol Rev 82:173–211. doi:10.1111/j.1469-185X.2006.00004.x

Giger T, Excoffier L, Day PJR et al (2006) Life history shapes gene expression in salmonids. Curr Biol CB 16:R281–282. doi:10.1016/j.cub.2006.03.053

Giger T, Excoffier L, Amstutz U et al (2008) Population transcriptomics of life-history variation in the genus *Salmo*. Mol Ecol 17:3095–3108. doi:10.1111/j.1365-294X.2008.03820.x

Goldrath AW, Bevan MJ (1999) Selecting and maintaining a diverse T-cell repertoire. Nature 402:6–13. doi:10.1038/35005508

Goltry KL, Epperly MW, Greenberger JS (1998) Induction of serum amyloid A inflammatory response genes in irradiated bone marrow cells. Radiat Res 149:570–578

Goudet J (1995) FSTAT (Version 1.2): a computer program to calculate F-statistics. J Hered 86:485–486
Grishkevich V, Yanai I (2013) The genomic determinants of genotype × environment interactions in gene expression. Trends Genet TIG 29:479–487. doi:10.1016/j.tig.2013.05.006
Hansen MM (2010) Expression of interest: transcriptomics and the designation of conservation units. Mol Ecol 19:1757–1759. doi:10.1111/j.1365-294X.2010.04597.x
Hedrick PW (2001) Conservation genetics: where are we now? Trends Ecol Evol 16:629–636. doi:10.1016/S0169-5347(01)02282-0
Houde AL, Wilson CC, Neff BD (2013) Genetic architecture of survival and fitness-related traits in two populations of Atlantic salmon. Heredity. doi:10.1038/hdy.2013.74
Houde ALS, Wilson CC, Neff BD (2014) Competitive interactions among multiple non-native salmonids and two populations of Atlantic salmon. Ecol Freshw Fish n/a–n/a. doi: 10.1111/eff.12123
Huang DW, Sherman BT, Lempicki RA (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57. doi:10.1038/nprot.2008.211
Hutter S, Saminadin-Peter SS, Stephan W, Parsch J (2008) Gene expression variation in African and European populations of *Drosophila melanogaster*. Genome Biol 9:R12. doi:10.1186/gb-2008-9-1-r12
Lamaze FC, Garant D, Bernatchez L (2013) Stocking impacts the expression of candidate genes and physiological condition in introgressed brook charr (*Salvelinus fontinalis*) populations. Evol Appl 6:393–407. doi:10.1111/eva.12022
Larsen PF, Nielsen EE, Williams TD et al (2007) Adaptive differences in gene expression in European flounder (*Platichthys flesus*). Mol Ecol 16:4674–4683. doi:10.1111/j.1365-294X.2007.03530.x
Lin B, Chen S, Cao Z et al (2007) Acute phase response in zebrafish upon *Aeromonas salmonicida* and *Staphylococcus aureus* infection: striking similarities and obvious differences with mammals. Mol Immunol 44:295–301. doi:10.1016/j.molimm.2006.03.001
Luca F, Kashyap S, Southard C et al (2009) Adaptive variation regulates the expression of the human SGK1 gene in response to stress. PLoS Genet 5:e1000489. doi:10.1371/journal.pgen.1000489
Miller KM, Li S, Kaukinen KH et al (2011) Genomic signatures predict migration and spawning failure in wild Canadian salmon. Science 331:214–217. doi:10.1126/science.1196901
Miwata H, Yamada T, Okada M et al (1993) Serum amyloid A protein in acute viral infections. Arch Dis Child 68:210–214. doi:10.1136/adc.68.2.210
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. doi:10.1139/f96-192
Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. Nat Genet 32:261–266. doi:10.1038/ng983
Ouborg NJ, Pertoldi C, Loeschcke V et al (2010) Conservation genetics in transition to conservation genomics. Trends Genet 26:177–187. doi:10.1016/j.tig.2010.01.001
Paradis E (2010) Pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics 26:419–420. doi:10.1093/bioinformatics/btp696
Parrish DL, Behnke RJ, Gephard SR et al (1998) Why aren't there more Atlantic salmon (*Salmo salar*)? Can J Fish Aquat Sci 55:281–287. doi:10.1139/d98-012
Paterson S, Piertney SB, Knox D et al (2004) Characterization and PCR multiplexing of novel highly variable tetranucleotide Atlantic salmon (*Salmo salar* L.) microsatellites. Mol Ecol Notes 4:160–162. doi:10.1111/j.1471-8286.2004.00598.x
Pedersen KS, Kristensen TN, Loeschcke V (2005) Effects of inbreeding and rate of inbreeding in *Drosophila melanogaster*—Hsp70 expression and fitness. J Evol Biol 18:756–762. doi:10.1111/j.1420-9101.2005.00884.x
Reed DH, Frankham R (2001) How closely correlated are molecular and quantitative measures of genetic variation? a meta-analysis. Evolution 55:1095–1103. doi:10.1111/j.0014-3820.2001.tb00629.x
Roberge C, Einum S, Guderley H, Bernatchez L (2006) Rapid parallel evolutionary changes of gene transcription profiles in farmed Atlantic salmon. Mol Ecol 15:9–20. doi:10.1111/j.1365-294X.2005.02807.x
Roberge C, Guderley H, Bernatchez L (2007) Genomewide identification of genes under directional selection: gene transcription Q(ST) scan in diverging Atlantic salmon subpopulations. Genetics 177:1011–1022. doi:10.1534/genetics.107.073759
Roberge C, Normandeau E, Einum S et al (2008) Genetic consequences of interbreeding between farmed and wild Atlantic salmon: insights from the transcriptome. Mol Ecol 17:314–324. doi:10.1111/j.1365-294X.2007.03438.x
Rouillard J-M, Zuker M, Gulari E (2003) OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. Nucleic Acids Res 31:3057–3062. doi:10.1093/nar/gkg426
Saeed AI, Sharov V, White J et al (2003) TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34:374–378
Schlichting CD, Smith H (2002) Phenotypic plasticity: linking molecular mechanisms with evolutionary outcomes. Evol Ecol 16:189–211. doi:10.1023/A:1019624425971
Scott RJ, Poos MS, Noakes DLG, Beamish FWH (2005) Effects of exotic salmonids on juvenile Atlantic salmon behaviour. Ecol Freshw Fish 14:283–288. doi:10.1111/j.1600-0633.2005.00099.x
Smith EN, Kruglyak L (2008) Gene-environment interaction in yeast gene expression. PLoS Biol 6:e83. doi:10.1371/journal.pbio.0060083
Smyth GK (2005) limma: linear models for microarray data. In: Gentleman R, Carey VJ, Huber W et al (eds) Bioinforma Computational Biology Solution Using R Bioconductor. Springer, New York, pp 397–420
Spielman D, Brook BW, Briscoe DA, Frankham R (2004) Does inbreeding and loss of genetic diversity decrease disease resistance? Conserv Genet 5:439–448. doi:10.1023/B:COGE.0000041030.76598.cd
Storey JD, Madeoy J, Strout JL et al (2007) Gene-expression variation within and among human populations. Am J Hum Genet 80:502–509. doi:10.1086/512017
Townsend JP, Cavalieri D, Hartl DL (2003) Population genetic variation in genome-wide gene expression. Mol Biol Evol 20:955–963. doi:10.1093/molbev/msg106
Uno T, Ishizuka M, Itakura T (2012) Cytochrome P450 (CYP) in fish. Environ Toxicol Pharmacol 34:1–13. doi:10.1016/j.etap.2012.02.004
Van Zwol JA, Neff BD, Wilson CC (2012) The influence of non-native salmonids on circulating hormone concentrations in juvenile Atlantic salmon. Anim Behav 83:119–129. doi:10.1016/j.anbehav.2011.10.015
Vandersteen Tymchuk W, O'reilly P, Bittman J et al (2010) Conservation genomics of Atlantic salmon: variation in gene expression between and within regions of the Bay of Fundy. Mol Ecol 19:1842–1859. doi:10.1111/j.1365-294X.2010.04596.x
Wellband KW, Heath DD (2013) The relative contribution of drift and selection to transcriptional divergence among Babine Lake tributary populations of juvenile rainbow trout. J Evol Biol

26:2497–2508. doi:10.1111/jeb.12247
Whitehead A, Crawford DL (2006) Neutral and adaptive variation in gene expression. Proc Natl Acad Sci USA 103:5425–5430. doi:10.1073/pnas.0507648103
Yauk CL, Berndt ML, Williams A, Douglas GR (2004) Comprehensive comparison of six microarray technologies. Nucleic Acids Res 32:e124. doi:10.1093/nar/gnh123