Heredity

ORIGINAL ARTICLE

Linking transcriptomic and genomic variation to growth in brook charr hybrids (*Salvelinus fontinalis*, Mitchill)

B Bougas[1], E Normandeau[1], C Audet[2] and L Bernatchez[1]

Hybridization can lead to phenotypic differences arising from changes in gene expression patterns or new allele combinations. Variation in gene expression is thought to be controlled by differences in transcription regulation of parental alleles, either through *cis*- or *trans*-regulatory elements. A previous study among brook charr hybrids from different populations (Rupert, Laval, and domestic) showing distinct length at age during early life stages also revealed different patterns in transcription regulation inheritance of transcript abundance. In the present study, transcript abundance using RNA-sequencing and quantitative real-time PCR, single-nucleotide polymorphism (SNP) genotypes and allelic imbalance were assessed in order to understand the molecular mechanisms underlying the observed transcriptomic and differences in length at age among domestic × Rupert hybrids and Laval × domestic hybrids. We found 198 differentially expressed genes between the two hybrid crosses, and allelic imbalance could be analyzed for 69 of them. Among these 69 genes, 36 genes exhibited *cis*-acting regulatory effects in both of the two crosses, thus confirming the prevalent role of *cis*-acting regulatory elements in the regulation of differentially expressed genes among intraspecific hybrids. In addition, we detected a significant association between SNP genotypes of three genes and length at age. Our study is thus one of the few that have highlighted some of the molecular mechanisms potentially involved in the differential phenotypic expression in intraspecific hybrids for nonmodel species.

INTRODUCTION

Hybridization can be an important source of phenotypic variation among individuals and of novelty among species. Phenotypic differences arise from variation in gene sequence or in gene expression regulation, both of which depend on mutations. Changes in the coding sequence of a gene can affect its product's function, whereas changes in the regulatory sequence can affect the quantity of the gene product found in cells. Numerous studies have characterized genome-wide patterns of gene expression among populations exhibiting phenotypic variation. In fishes in particular, the majority of such studies have not only used microarrays (Oleksiak *et al.*, 2002; Whitehead and Crawford, 2005; Derome and Bernatchez, 2006; Derome *et al.*, 2006; Whitehead and Crawford, 2006; St-Cyr *et al.*, 2008; Whiteley *et al.*, 2008), but also next-generation sequencing to a lower extent (Elmer *et al.*, 2010; Goetz *et al.*, 2010; Jeukens *et al.*, 2010; Shen *et al.*, 2012). Hybridization can also lead to different modes of transcription regulation. For example, in hybrids, gene expression can be additive or dominant or the gene can be misexpressed (under/overdominant) relative to the expression level observed in their parents. Within species, the mode of transcription inheritance depends on the parental genetic architecture and cannot be predicted based on the information of parental phenotypes or the amount of genetic divergence between the parental populations (Bougas *et al.*, 2010). Because of these complexities, the molecular mechanisms linking transcriptomic differences to the phenotype are still poorly understood, especially so in nonmodel species.

Gene expression is under the control of regulatory polymorphism found both within or in close proximity of the gene locus (*cis*) and elsewhere in the genome (*trans*).

Cis mutations, often located in promoters or enhancers, affect only the expression level of the alleles present on the same chromosome as the mutation. *Trans* mutations, on the other hand, affect the expression of both alleles, as *trans*-regulatory elements correspond to genes encoding for transcription factors that often interact with promoters and enhancers in order to modulate gene expression. Many studies have revealed the important role of *cis*-acting regulatory mutations in generating both transcriptional and phenotypic variation (reviewed by Wray, 2007). For example, in brook charr, the MHCIIβ gene transcript abundance was negatively associated with the length of the alleles of a minisatellite present in an intron of the gene, and thus acting in *cis* (Croisetière *et al.*, 2010). However, through regulatory cascades, *trans*-acting elements have the potential to affect many genes and also play an important role in controlling gene expression (Levine and Tjian, 2003; Wittkopp 2005; Maston *et al.*, 2006). For example, dominance in gene expression inheritance has been significantly associated with *trans* regulatory elements (Lemos *et al.*, 2008). Finally, gene expression regulation can be affected by the interaction of both regulatory systems, either through *trans* by *trans* or through *cis* by *trans* interactions (Landry *et al.*, 2005).

[1]Département de biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Canada and [2]ISMER, Université du Québec à Rimouski, Rimouski, Canada
Correspondence: Dr B Bougas, Institut de Biologie Intégrative et des Systèmes (IBIS), Pavillon Charles-Eugène Marchand, Université Laval, Québec, Canada G1V 0A6.
E-mail: berenice.bougas.1@ulaval.ca

The mechanisms underlying the regulation of gene expression can be explored with expression quantitative trait loci (eQTL) that combine transcript abundance and linkage mapping data to detect the effects of *cis* or *trans* acting on gene expression (Gibson and Weir, 2005). Another method for studying regulatory variations investigates patterns of allelic imbalance, defined as the asymmetric expression of the two alleles of a locus. Allelic imbalance is caused by the direct influence of *cis*-regulatory elements. On the other hand, we may conclude to the presence of *trans*-regulatory effects when the following two conditions are met: (1) no allelic imbalance is observed in heterozygous individuals, as *trans*-regulatory elements affect both alleles in the same manner, and (2) when homozygous individuals possessing the same allele and originating from different populations show differences in allelic expression levels (Wittkopp et al., 2004; Landry et al., 2005).

Most of the methods used in previous studies required a priori knowledge of the genetic basis of observed gene expression variation, which is often not the case when studying such mechanisms in nonmodel species. For example, eQTL requires QTL mapping and genome-wide transcriptomic analyses, both of which require substantial logistical development. Similarly, in pyrosequencing, allele-specific real time-PCR or single-nucleotide polymorphism (SNP) arrays require a priori knowledge of allelic variants for the whole panel of SNPs. In nonmodel species, RNA-sequencing (RNA-seq) offers an efficient means to study allele specific expression because it requires no a priori genomic information. RNA-seq makes it possible to detect SNPs with sufficient sequencing depth to measure individual allelic expression and to detect allelic imbalance on a genome-wide scale (Fontanillas et al., 2010).

In this context, this study uses 454 sequencing to investigate genome-wide differences in allelic expression in a nonmodel vertebrate, brook charr (*Salvelinus fontinalis*, Mitchill). Brook charr is an economically important species native to eastern North America. In the Province of Québec, Canada, it represents 57% of the fresh-water aquaculture production in weight (MAPAQ, 2010). Hybrids were produced using three genetically different populations, two wild and one domestic, and used to investigate phenotypic and genomic differences relative to their parental populations (Bougas et al., 2010). In this previous study, we found that the hybrid crosses exhibited contrasting modes of gene expression inheritance, being mostly additive for domestic ♀ (D) × Rupert ♂ (R) hybrids and predominantly nonadditive for Laval ♀ (L) × domestic ♂ (D) hybrids. These crosses also presented significant growth (length at age) differences with L♀D♂ longer than D♀R♂. Here, our specific objectives were: (1) to test if differences among the transcription patterns of the D♀R♂ and L♀D♂ crosses correlate with identifiable regulatory differences; (2) to associate the mode of gene expression inheritance with the type of regulation (*cis* or *trans*) on a gene by gene basis, and (3) to test for significant association between gene expression and genotypes at SNP markers with size phenotypes in order to better understand the molecular mechanisms responsible for the observed phenotypic differences.

MATERIALS AND METHODS
Fish crosses and samples
Breeders from domestic (D), Laval (L) and Rupert (R) populations were crossed to generate 10 full-sib outbred families per pure cross and per hybrid cross. The breeders from the L population were kept in captivity for three generations at the Institut des Sciences de la Mer à Rimouski (ISMER, Québec) and the R population breeders were kept at the Laboratoire de Recherche des Sciences Aquatiques (LARSA, Laval University, Québec). Breeders from the D population were obtained from the Jacques Cartier fish farm (Cap-Santé, Quebec), and parental crosses and two hybrid crosses, D♀R♂ and L♀D♂, were generated. All families were kept separately at the LARSA under identical controlled conditions. Fertilized eggs were incubated at 6 °C. After hatching, the progeny was kept at 8 °C, with a photoperiod of 12 h of light and 12 h of darkness. At the yolk sac resorption stage, 10 sexually undifferentiated juvenile were randomly sampled from 10 families for L♀D♂ and 9 families for D♀R♂, their fork length was measured and they were then immediately frozen in a mix of dry ice and alcohol. Crosses and analysis procedures for each cross are presented in Figure 1.

Preparation of cDNA libraries and 454 sequencing
For each cross, total RNA was extracted from one individual in eight different full-sib families for each cross (in total, eight individuals per cross) with the PureLink Micro-to-Midi Total RNA Purification System Kit and then treated using DNase I, Amplification Grade (1 unit μl^{-1}; Invitrogen, Carlsbad, CA, USA) following the manufacturer's protocol. Total RNA was further purified by ultra-centrifugation using microcon (Millipore, Billerica, MA, USA) spin columns. Total RNA was stored in pure water supplemented with Superase-In RNase Inhibitor (Ambion, Foster City, CA, USA) at −80 °C. Quality and quantity of the total RNA was measured using Nanodrop (Thermo Scientific, Waltham, MA, USA). mRNA was isolated using MicroPoly(A)Purist Kit (Ambion). All of the mRNA was used for complementary DNA (cDNA) synthesis following the SMART PCR cDNA Synthesis Protocol (Clontech, Mountain View, CA, USA) and as described in Pierron et al. (2011). The cDNA fragments were tagged with adapters specific to each individual, as proposed by Roche (Basel, Switzerland) (Roche technical bulletin No. 005-2009). Two pools with equal cDNA quantity were prepared (eight D♀R♂ samples and eight L♀D♂ samples). For both libraries, fragments in the 350–900 bp size range were selected by gel migration and extraction. Approximately 4 μg of double-stranded cDNA from each library were sequenced on two separated half plates on the Roche GS-FLX DNA Sequencer at the Genome Quebec Innovation Center (McGill University, Montréal, Canada) following the original Roche protocol starting from the fragment immobilization step.
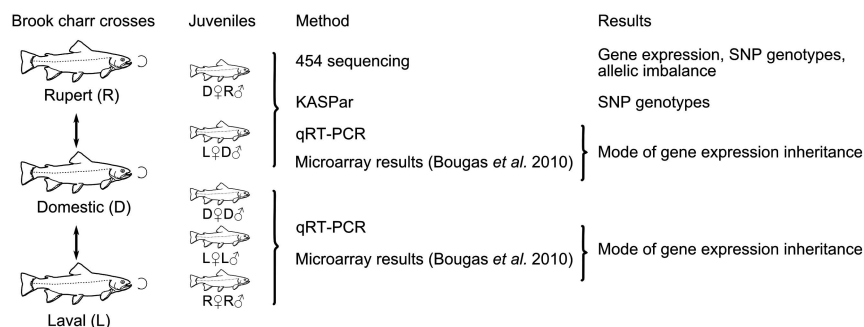

Figure 1 Crossing procedures and genomic analyses that were done for each cross.

Contig assembly

RNA-seq of the prepared cDNA libraries with 454 Life Sciences Technology (Roche) provided us with raw data files in .sff format for the D♀R♂ and L♀D♂ crosses, as well as a supplementary data set, hereafter referred to as SF (for *Salvelinus fontinalis*), which was available from a previous study and was deposited in NCBI (National Center for Biotechnology Information) short read archive (accession number SRX037496). Base quality was called on all the .sff files using PyroBayes beta v0.9 (http://bioinformatics.bc.edu/marthlab/PyroBayes) to generate fasta and fasta.qual files containing the sequence and sequence quality information, respectively. The sequences in these files were scanned in order to find and remove the individual tags and rename the sequences to include the name of their individual of origin. An average of 97.0% of the reads in L♀D♂ and D♀R♂ sequences, totaling just over a million, were individually renamed in this manner. Using the same approach, CAP amplification sequences were removed from the data. Using CLC Genomic Workbench version 3.7 (http://www.clcbio.com/), all the prepared reads were used in order to create a set of simple contigs (assembly criteria: similarity = 0.95, length fraction = 0.6, kmer length = 21). The L♀D♂ and D♀R♂ sequences were then mapped on the consensus sequences from the previous step (assembly criteria: similarity = 0.95, length fraction = 0.6).

Gene annotation

Contigs were blasted locally on both the swissprot and nr protein databases using the blastx program from BLAST 2.2.23 + (http://blast.ncbi.nlm.nih.gov). Only the single most meaningful blast hit for each contig was retained. The meaningfulness of each blast was assessed by using the *e*-value of the hit and multiplying it by 1×10^p, where p is a penalty based on the presence of certain keywords in the blast hit name. The keywords and associated penalties used are the following: (unnamed: 200, unknown: 30, uncharacterized: 30, hypothetical: 20, predicted: 10, similar to: 5, novel protein: 5). Only hits with an *e*-value of $\geq 1 \times 10^{-30}$ were retained for functional classification analysis. Functional classification and assessment of significant differential representation of functional classes were performed with the blast2GO program (http://www.blast2GO.org) using Gene Ontology annotation (Fisher test, significance threshold: $P = 0.05$).

Gene expression and correlation with length phenotype

Using the extracted read counts for each individual at each gene following the mapping phase, the DEGseq R package (http://bioconductor.org/packages/2.6/bioc/html/DEGseq.html) was used to identify genes differently expressed between the L♀D♂ and D♀R♂ groups. A threshold of 0.01 was used for the *P*-value. Only contigs containing an average of three sequences per individual, for a minimal total of 48 sequences per contig, were used in the analysis (in total: 1558 contigs). In order to find genes whose expression was correlated with length at age, individual transcript abundances were tested for correlation with length at age after accounting for mean group effects. For this analysis, the transcript abundance was normalized on individual basis using the RPKM method (reads per kilobase per million mapped). A Spearman's correlation was then fitted to the data using R function and a *P*-value of 0.01 was used to generate a list of significant correlations.

SNP discovery and allelic imbalance calculation

Following assembly, 2257 SNPs were found among the 9587 contigs using CLC Genomic workbench 3.7 (minimum SNP coverage = 128, corresponding to an average of 8 per individual, minimum variant frequency = 5%). Individual genotypes were exported and analyzed as follows. Only SNPs in regions with good-quality values and at least three nucleotides away from read ends were counted. Using the genotype counts for all the individuals, SNPs for which the most frequent variant represented a maximum of 90% of all the counted alleles and for which the two most frequent variants represented at least 98% of all counted alleles were categorized as biallelic. For each of these biallelic SNPs, we labeled individuals as either heterozygous if the reads showed that they had both alleles with at least 5% of the rare variant or otherwise as homozygous. Finally, we analyzed only loci for which there were at least three heterozygous individuals in the L♀D♂ and D♀R♂ groups, and a maximum of 12 heterozygous individuals. By applying an upper limit of 12 heterozygous

individuals on a total of 16 meant, we reduce greatly the number of SNPs from contigs derived from the assembly of two similar but different genes, as is the case for very similar paralogs. Following this procedure, the contig assemblies were visually inspected with great care in order to remove any remaining dubious SNP candidates, mostly from contigs that would have resulted from the assembly of paralogous sequences. More specifically, contigs with a high number of SNPs or with SNPs that showed clear segregation patterns (allele A at position 1 always linked with allele A' at position 2 and similarly for alleles B and B' at the same positions) were discarded as obvious case of paralogous assemblies. Moreover, any SNP found near the extremities of the contigs where the read quality was confirmed to drop were excluded. For each of the remaining loci, the presence of allelic imbalance was only assessed in heterozygous individuals with a binomial test (expected null allelic expression ratio = 1) at two levels: (1) for all of the 16 individuals pooled together; (2) for each of the two groups of eight individuals. A *P*-value of 0.01 was used as a cutoff for the binomial tests at the two levels. The Qvalue R package (http://genomics.princeton.edu/storeylab/qvalue/) was used to find the expected false discovery rate (FDR) of the list of loci that were significant at this *P*-value of 0.01. Using the list of *P*-values from the global test using the 16 pooled individuals, the proportion of expected false positives was 0.0037. For each of the L♀D♂ and D♀R♂ groups taken separately, the FDRs are 0.011 and 0.0078, respectively.

Correlation of microarrays and RNA-seq gene expression

In our previous study (Bougas *et al.*, 2010), microarrays were used to identify the mode of gene expression inheritance by comparison of transcript levels between hybrid and parental crosses. In order to associate the allelic imbalance patterns of the present study with their mode of gene expression inheritance, genes found in both this microarray and the 454 sequencing experiments D♀R♂ and L♀D♂ were used, as described in Supplementary Methods S1.

Quantitative real-time PCR

To confirm the mode of gene expression inheritance determined from our previous microarray study (Bougas *et al.*, 2010), the transcript abundance for five genes with allelic imbalance was determined in hybrids and their parental populations using quantitative real-time PCR (qRT–PCR). Gene-specific primers and probes (Supplementary Table S1) were designed from the contig consensus sequences and no SNP was contained in either the primers or the probes. RNA of eight hybrids from each hybrid and parental population cross (LD, DR, Laval, Rupert and domestic) were extracted, treated and measured as described for the preparation of the cDNA libraries. An equal amount of RNA (0.5 μg) was used in 20 μl reverse reactions for each individual using the High capacity cDNA Reverse Transcription Kit (Applied Biosystems, Life Technologies, Carlsbad, CA, USA). Relative quantification of transcript abundances was achieved by using the *peptidyl-prolyl isomerase* gene as a reference gene. The qRT–PCR was performed in triplicate using Taqman in the 7500 Fast Real-Time System following the manufacturer's protocol (ABI PRISM 7500 Fast Real-Time System, Applied Biosystems). The analysis of transcript abundances in hybrids relative to parental populations was performed with *t*-tests in R where transcript abundance in the hybrids was contrasted to transcription abundance of both parental lines combined. We concluded to a nonadditive mode of transmission when the transcript abundance of a gene in the hybrids was significantly different from the mean expression values of that gene in the parental strains ($P < 0.05$).

Association between SNP variation and length at age

Genotypes for all the SNPs with calculated allelic imbalance were tested for association with length at age using the SNPassoc package in R. This package analyzes associations and interaction between SNP genotypes and phenotype measures under five genetic patterns (additive, codominant, dominant, recessive and overdominant). Group effect on length was taken into account by incorporating 'group' as an explanatory variable in the statistical model. A minimal *P*-value of 0.1 was used to generate a first list of significant associations that represented six genes. Then, in order to verify the associations for the six genes thus selected, 100 individuals (50 from the 10 L♀D♂ families corresponding to 5 individuals per family and 50 from the 9 D♀R♂ families

corresponding to 5 or 6 individuals per family) were genotyped using KASPar SNP Genotyping System v.2.0 method following the manufacturer's protocol (KBiosciences, Hoddesdon, UK). Amplification and fluorescence scanning were performed with the PRISM 7500 Fast Real-Time System (Applied Biosystems) and the results were analyzed by the 7500 v2.0.4 software. Genotype and length data were corrected and separated by group before being analyzed with the SNPassoc package in R to verify the association and interactions (epistasis) between SNP genotypes and length phenotype.

Length at age analysis

Normality and homoscedasticity of length data were respectively tested with the Kolmogorov–Smirnov and the Bartlett tests. To detect potential cross effects, data were analyzed with the following analysis of variance mixed model, implemented in ASReml version 2 (VSN International, Hemel Hempstead, UK): $Y_{ijk} = \mu + C_i + F_{j(i)} + \varepsilon_{ijk}$, with C (cross-type) as the fixed effect and F (family) nested within C as the random effect. The test was followed by a posteriori analysis when significant. The mean differences were tested with Wilcoxon/Mann–Whitney tests in R as the data were nonhomoscedastic. Length data are presented in Supplementary File S1.

RESULTS

Length differences between L♀D♂ and D♀R♂ crosses

The length analysis showed that L♀D♂ individuals were significantly longer for a developmental stage (yolk resorption), with a mean length of 29.25 ± 1.46 mm compared with a mean length of 22.80 ± 1.08 mm for the D♀R♂ individuals (Wilcoxon/Mann–Whitney test, $P < 0.05$). In this study, the length data were only used to analyze the correlation with gene expression and SNP genotypes.

Sequencing, contig assembly, gene expression and correlation with length at age

Sequencing of the two half-plates produced 1 045 694 reads of 263 bp on average. Following the described assembly criteria, the re-assembly of L♀D♂ and D♀R♂ data produced 9587 contigs with an average length of 568 nucleotides. Among the 1558 expressed contigs represented by a minimum of 48 sequences, 333 contigs, representing 198 unique annotated genes and 72 unknown genes, were differently expressed between D♀R♂ and L♀D♂ individuals at $P < 0.01$ (corresponding to FDR <0.064). In D♀R♂ relative to L♀D♂, 119 genes were overexpressed and 79 were underexpressed. Among the 198 unique annotated genes, 178 possessed a Gene Ontology identifier. In total, four biological processes were overrepresented among the differentially expressed genes (Figure 2). Two of the overrepresented biological processes (embryonic development and anatomical structure morphogenesis) were associated with development. However, an equal number of genes involved in development processes were overexpressed in D♀R♂ relative to L♀D♂ (χ^2-test, $P > 0.05$). All differentially expressed genes, biological processes and corresponding ratios are presented in Supplementary Table S2. Among the 1558 contigs, the transcript abundance of 10 contigs, representing 5 unique genes and 5 unknown genes, were significantly correlated with length at age (Spearman's correlation, $P < 0.01$). These genes are: 60S ribosomal protein L23a (rho = −0.65), 60S ribosomal protein L26 (rho = −0.73), serum albumin 2 (rho = 0.65), troponin T (rho = −0.76) and transcription initiation factor IIA subunit 1 (rho = −0.69). However, because 16 false positives were expected from our set of 1558 contigs at this α level, we cannot rule out the possibility that these genes represent false positives.

Allelic imbalance, RNA-seq and microarray correlation

Quantification of allelic imbalance allowed the detection of cis-regulatory effects controlling the expression of some of the loci. In
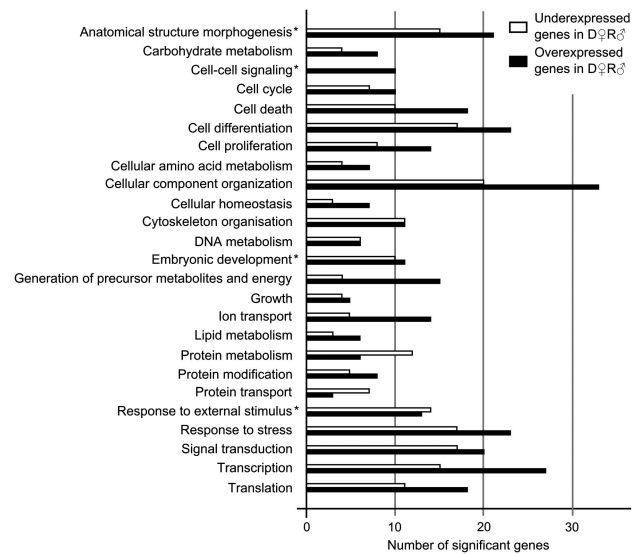

Figure 2 Distribution of the differentially expressed annotated genes in the biological processes. Only the genes annotated with at least one Gene Ontology number are represented. Underexpressed genes in D♀R♂ are in white and overexpressed genes in D♀R♂ are in black. *Overrepresented biological processes in the D♀R♂/L♀D♂ comparison (Fisher's exact test, significance threshold: $P = 0.05$; http://www.blast2GO.org). The reference list was built using the list of the analyzed contigs ($n = 1558$). The biological processes represented by minimum eight genes are presented.

total, allelic expression was measured for 144 SNPs, representing 53 unique annotated genes and 16 unknown genes (Supplementary Table S3). Among these genes, 58 genes had a SNP in both groups; 8 and 3 genes were unique to either L♀D♂ or D♀R♂ groups, respectively. For these genes, 18.7% more reads were detected in L♀D♂ individuals compared with D♀R♂ individuals (L♀D♂: 17 300 reads; D♀R♂: 11 848 reads). In total, 36 genes with a SNP detected in both groups showed significant allelic imbalance, with 31 and 25 genes in the L♀D♂ and D♀R♂ groups, respectively ($P < 0.01$; Figure 3 and Supplementary Table S3). Among these, 16 genes showed an allelic imbalance in only one group. The number of genes showing allelic imbalance was not significantly different between the two groups (Fisher's exact test, $P = 0.078$). Two other genes with a SNP specific to L♀D♂ also showed allelic imbalance. The maximum and minimum of $\log_2$ ratio for a significant allelic imbalance were 3.5 and 0.94, corresponding to ratios of 11.38 and 1.9 between the most and least abundant alleles. Among the genes with allelic imbalance, 18 genes were differentially expressed between the crosses at $P < 0.01$; 17 and 11 of these 18 genes were cis regulated in L♀D♂ and D♀R♂, respectively (Fisher's exact test: $P = 0.02$).

In addition to the detection of their cis regulation, the genes were classified according to their mode of expression regulation inheritance. To achieve this, microarray results from Bougas et al. (2010) and new qRT–PCR results were used to link the mode of expression regulation inheritance with allelic imbalance for genes that were detected in the two experiments. Genes were classified as additive when the level of expression in the hybrids was not significantly different from the average value of parental populations. When the level of transcription in the hybrids significantly differed from the average expression values of the parental populations, transcription level was considered to be nonadditive, meaning that it was either dominant or over/underdominant. Transcripts were defined as dominant if their abundance between only one of the two parental
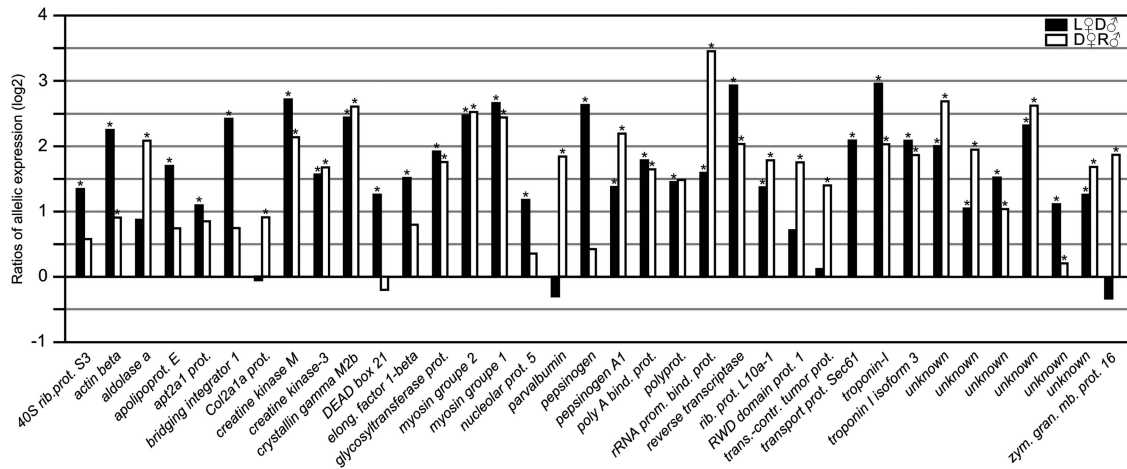
Figure 3 Ratio of allelic expression (log$_2$) between L♀D♂ and D♀R♂ hybrids for the 36 genes with a significant allelic expression difference ($P<0.01$). For each gene, only the SNP with the highest allelic difference ratio was represented (ratios for all the SNPs in all genes are presented in Supplementary Table S3). bind., binding; contr., controlled; D♀R♂, domestic female × Rupert male hybrids; elong., elongation; gran., granule; L♀D♂, Laval female × domestic male hybrids; mb., membrane; prom., promoter; prot., protein; rib., ribosomal; trans., translationally; zym., zymogen. *Ratio of allelic expression that are significant.

Table 1 Genes showing allelic imbalance according to RNA-seq analysis and their abundance of transcripts in hybrids relative to parental populations according to microarray (Bougas *et al.,* 2010) and quantitative qRT–PCR

Gene name	L♀D♂			D♀R♂		
	AI	A/NA microarray	A/NA qRT–PCR	AI	A/NA microarray	A/NA qRT–PCR
Elongation factor 1-β¤	*	A	NA OD	†	A	A
Myosin heavy chain group 1¤	*	NA D	‡NA OD	*	‡	‡
Myosin heavy chain group 2¤	*	NA D	‡NA OD	*	‡	‡
Nucleolar protein 5¤	*	A	A	†	A	A
Zymogen granule membrane protein 16 precursor¤	†	NA D	NA D	*	A	A
Apolipoprotein E	*	‡	‡	†	A	‡
Atp2a1 protein	*	NA UD	‡	†	‡	‡
Cytochrome c oxidase polypeptide VIIa-liver/heart	†	‡	‡	†	‡NA OD	‡
DEAD (Asp-Glu-Ala-Asp) box polypeptide 21	*	NA D	‡	†	‡	‡
Fast myotomal muscle troponin-I	*	A	‡	*	‡	‡
Poly A binding protein, cytoplasmic 1a	*	NA D	‡	*	‡	‡
Ribosomal protein L10a-1	*	A	‡	*	‡	‡
Troponin I	*	A	‡	*	‡	‡

Abbreviations: A, additive; AI, allelic imbalance; D, dominant; D♀R♂: domestic female × Rupert male hybrids; L♀D♂: Laval female × domestic male hybrids; NA, nonadditive; OD, overdominant; UD, underdominant.
¤The abundance of transcript in hybrids relative to their parental population was verified by qRT–PCR for the first five genes. *Significant allelic expression difference. †No significant difference.
‡Transcript did not show significant expression difference between the parental populations.

populations and the hybrid cross was nonsignificant. Finally, transcripts were defined as underdominant if their abundance was significantly lower than both parental populations and as overdominant if it was significantly higher than the abundance in both parental populations (FDR = 0.05) (Bougas *et al.,* 2010).

Blasts of the sequences and correlation of the transcription ratio from the previous microarray experiment and 454 sequencing results revealed that there were 117 genes in common to both experiments and that the transcription ratios were significantly correlated (rho = 0.46, $P = 1.879e^{-07}$). Among these 117 genes, 13 out of the 58 for which allelic expression was calculated for a SNP that was variable in both groups also have a determined mode of expression regulation inheritance in the microarray data (Table 1). In total, most of the additive and nonadditive genes showed significant allelic

imbalance. For L♀D♂, five out of five additive genes and five out of six nonadditive genes showed allelic imbalance. For D♀R♂, there was one out of four additive genes and no nonadditive gene out of one gene that showed allelic imbalance. However, for a given mode of transcription inheritance, the presence of allelic imbalance was sometimes observed in only one group. Indeed, two genes (*elongation factor 1-beta* and *nucleolar protein 5*) with an additive mode of transcription regulation in the two crosses had significant allelic imbalance in L♀D♂ but no significant allelic imbalance in D♀R♂. In addition, one gene (*zymogen granule membrane protein 16 precursor*) was nonadditive in L♀D♂ individuals and additive in D♀R♂ individuals. It showed significant allelic imbalance in D♀R♂ individuals only. The qRT–PCR experiment confirmed the patterns of gene expression mode of transmission observed in the earlier

microarray experiments, for the majority of the cases (Table 1). Only one gene, *elongation factor 1 beta* gene, displayed major differences in the mode of transmission of gene expression between qRT–PCR and microarray results, and this was only for L♀D♂ individuals. Microarrays suggested that the mode of transmission of that gene was additive and qRT–PCR suggested that it was nonadditive.

Association between SNP genotypes and length phenotypes

From the association test between the 144 SNP genotypes and length at age, the genotypes of six genes were found to vary significantly with length (at $P < 0.1$, three genes had $P < 0.01$ and three other had $P < 0.1$; Supplementary Table S4). The genotypes of these six SNPs were thus characterized on 100 individuals. Three of these genes showed an excess of homozygous individuals and could not be tested for association with length. The three other genes (*myosin group2*, *elongation factor 2* (ef2) and *solute carrier family 25* (scf 25) gave reliable genotypes. Thus, genotypic variation at the *myosin group2* gene was significantly correlated with length at age (Figure 4). Indeed, heterozygous fish were found to be significantly smaller than homozygous fish ($P = 0.024$). The SNP in myosin group 2 gene corresponds to a nonsynonymous mutation that changes the amino acid arginine to serine, which has different polarity and charge. No direct association with length was significant for the other two genes. However, epistasis between ef2 and scf 25 genes was found to be highly significantly associated to length variation ($P = 0.0008$; Figure 5). Two combinations of genotypes were associated with a high-length phenotype. Namely, hybrids being homozygous CC
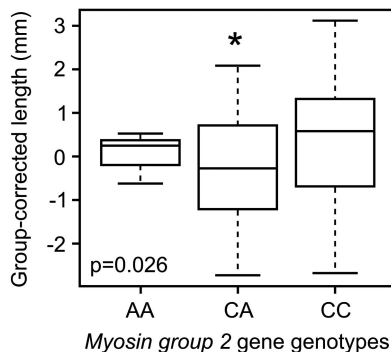

Figure 4 Association between length at age and genotype for the *myosin group 2* gene. The significant association was tested by group-corrected length for 100 L♀D♂ and D♀R♂ individuals using the SNPassoc R package. Here, *p* represents the minimal *P*-value obtained for the association test.

for scf 25 and GG for ef2 or TT for scf 25 and AA for ef2 were significantly longer relative to other hybrids with different genotype combinations. For the ef2 gene, the SNP corresponds to a synonymous mutation in the coding region, whereas for the scf25 gene, the SNP is found in the noncoding 3′untranslated region.

DISCUSSION

Gene expression among crosses

Microarrays and, more recently, RNA-seq by means of NGS technologies have been used to compare the transcriptomes of phenotypically distinct individuals, namely in salmonid fishes (Derome and Bernatchez, 2006; Derome et al., 2006; St-Cyr et al., 2008; Goetz et al., 2010; Jeukens et al., 2010; Xu et al., 2010). Typically, many transcriptomic differences involved genes with functions related to energy production, muscular activity, immunity and cell cycle regulation processes. These expression regulation differences could be related to variation for growth that was observed among these fish. Here, using RNA-seq, gene transcript abundances were measured in brook charr D♀R♂ and L♀D♂ crosses to further explore transcriptomic differences and document their potential link with size at age. This link is difficult to characterize because an equal number of genes from development and growth process are overexpressed in the two crosses. However, among the differently expressed genes of the growth process, the *insulin-like growth factor-binding protein 4* gene was found to be underexpressed in D♀R♂ relative to L♀D♂ individuals (log2 ratio of −1.15). As this gene was found to be overexpressed in fast-growing Atlantic salmon (Bower et al., 2008), its expression regulation could potentially be involved in length differences observed in L♀D♂ relative to D♀R♂ individuals at this stage. However, the individual lengths within group were not significantly correlated with the abundances of gene transcripts. This could be because of the small number of individuals (16, 8 per cross) that have been used in our experiment. In addition, length at age has been shown to be influenced by complex epistatic interaction and additive effects at many genes, each explaining a very small proportion of the size of an individual (Li et al., 2009). As our library was prepared from nonnormalized cDNA, it is possible that some key genes that were below our detection threshold were missed by our analyses (Bellin et al., 2009).

Cis-acting regulatory effects

Data from microarray analyses (Bougas et al., 2010) and from allele specific expression were combined to document the prevalence of cis- vs trans-regulatory effects present within and between hybrids, and to understand their influence on gene expression patterns
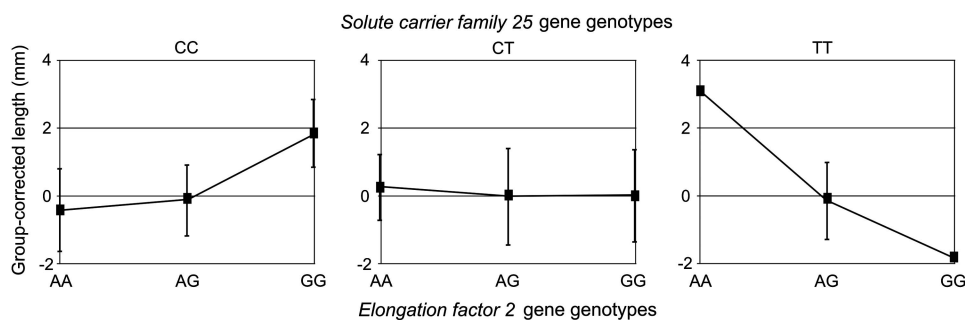

Figure 5 Association between length at age and genotype for the two epistatic genes *solute carrier family 25* and *elongation factor 2*. The significant association was tested by group-corrected length for 100 L♀D♂ and D♀R♂ individuals using the SNPassoc R package. The epistatic association was highly significant ($P = 0.0008$).

observed in hybrids. Both *cis*- and *trans*-regulatory effects determine gene expression levels and the different proportions of the possible genotypes in these *cis*- and *trans*-regulatory sequences produce the observed patterns of expression in the hybrids when species or populations are crossed. Here, because allelic expression was analyzed in hybrids, only *cis*-regulatory differences could be detected. In order to detect *trans*-regulatory effects, it would have been necessary to study allelic expression in not only hybrids but also parental populations at the same life stage, which was not possible here (Wittkopp *et al.,* 2004; Landry *et al.,* 2005; Fraser *et al.,* 2011). We therefore preferred to increase coverage for the hybrid crosses in order to at least more rigorously confirm the occurrence of *cis* regulation, as directly reflected by significant allelic imbalance in hybrids.

Nonadditive patterns of gene expression are generally caused by the action of dominant *trans* effects. For instance, Lemos *et al.* (2008) and Graze *et al.* (2009) found a significant association between dominance for gene expression and variants in *trans*-acting elements in *Drosophila.* Moreover, *trans* effects could be responsible for many phenotypic differences, as reported in yeast (Gerke *et al.,* 2009). Among the hybrids of the Rupert, Laval and domestic brook charr, we previously found important differences in the mode of gene expression inheritance relative to the parental populations. Namely, a prevalence of additivity was found in the D♀R♂ hybrids whereas a high prevalence of nonadditivity, mostly through dominance, was found in the L♀D♂ hybrids (Bougas *et al.,* 2010). In this study, however, the proportion of genes with significant allelic imbalance was not significantly different among these two groups of hybrids. This may seem surprising, as *cis*-regulated alleles should maintain their parental expression level in hybrids and result in gene expression values in hybrids close to the average observed in the parents. Consequently, more *cis*-regulated genes could be expected in crosses with a prevalence of additivity relative to crosses (for example, D♀R♂) with a prevalence of nonadditivity, which was not the case.

Here, our results suggest that variation in gene transcript abundance may have been caused largely by the presence of *cis*-regulatory element variants combined in variable genotypic proportions among the crosses studied. Indeed, almost all the genes that were differentially expressed among hybrids, including both the additive and nonadditive genes, showed variations linked to *cis* effects. For eight of the *cis*-regulated genes, allelic imbalance was detected in only one of the two crosses. For these genes, allelic imbalance could thus be a sufficient mechanism to explain the gene expression differences that were observed. For the other 10 genes, the action of *cis* elements was detected in both crosses. For these genes, the expression differences among the crosses could be explained by differences in allelic frequencies in the two crosses. Admittedly, eight individuals per group may not have provided sufficient power to detect significant differences in allelic frequencies using Fisher's exact test for proportions. Thus, only two genes with *cis*-regulatory elements showed significant differences in the frequencies of alleles between the two hybrid groups. Alternatively, gene expression differences could also have been caused by the interactions between *trans*-acting factors or between *trans*-acting factors and *cis*-acting regulatory elements.

Previous studies on regulatory elements have reported contrasting results regarding the overall proportions of *cis* vs *trans* effects, depending on the distance of the parental groups used to generate the crosses studied (interspecies or intraspecies) (Wayne *et al.,* 2004; Wittkopp *et al.,* 2004; Hughes *et al.,* 2006; Lemos *et al.,* 2008; Wittkopp *et al.,* 2008b; Graze *et al.,* 2009). However, most of these studies globally agree that *cis* regulation contributes more to gene expression variation than *trans* effects, especially in comparisons among populations relative to comparisons within populations. Allelic variants in *cis*-regulatory elements can be accumulated as a consequence of either local adaptation or neutrally (Wray, 2007; Wittkopp *et al.,* 2008b), and the amount of variation at *cis*-regulatory elements may correlate with genetic distance, both within and between species in *Drosophila* (Lemos *et al.,* 2008; Wittkopp *et al.,* 2008b). In order to explain the important contribution of variation in *cis* elements on the regulation of gene expression levels, these authors note that *cis*-regulatory element variants have a higher rate of fixation among populations than *trans* elements. Here, *cis*-regulatory divergence among the studied populations of brook charr could have accumulated in the time since the postglacial colonization of the river systems they inhabit. Moreover, artificial selection could have led to differentiation at these *cis* elements in the domestic population (Bougas *et al.,* 2010).

The prevalence of dominant and misexpressed genes that were found in the L♀D♂ cross relative to its parental populations suggests that the *trans* effects controlling gene expression in the parental populations interacted in a nonadditive manner. However, the presence of such *trans* effects will need to be investigated by further experiments. As regulatory elements are potentially co-adapted within populations or species, epistatic interactions between *cis* and *trans* effects can also lead to either dominance or over/underdominance of gene expression in hybrids (Landry *et al.,* 2005; Wittkopp *et al.,* 2008a; Graze *et al.,* 2009).

In addition, the differential allelic expression observed in the hybrids may have resulted from genomic imprinting, in which one of the parental alleles could be partially silenced by epigenetic effects (Morison *et al.,* 2005; Wolf *et al.,* 2008). Genomic imprinting may be especially important during development, and can be caused by DNA methylation, as well as histone acetylation and methylation (Angers *et al.,* 2010). Genomic imprinting could cause a dominant mode of gene expression inheritance and also modify the hybrid phenotypes. In a previous study, the expression of numerous genes in L♀D♂ hybrids was dominant relative to one of the parental populations (Bougas *et al.,* 2010). Thus, the observed patterns are consistent with what is expected in the case genomic imprinting, although the role of *cis*-regulatory effects caused by genomic imprinting remains to be investigated in more detail.

Epistasis
Epistasis may be responsible for both heterosis and outbreeding depression (Li *et al.,* 2001; Luo *et al.,* 2001; Nagel, 2005; Larièpe *et al.,* 2011). Epistasis could act by activation, repression or complementation of gene action. In rice, for example, dominance resulting from epistatic QTLs plays important role in heterosis (Luo *et al.,* 2009). The presence of epistatic interactions is implied for genes where *trans*-acting factors were detected. Out of the three genes that could be tested here by KASPar assays, two showed an epistatic influence on hybrid length at age (*elongation factor 2* and *solute carrier family 25*), whereas the genotype of the SNP in the third gene (*myosin group 2*) was directly correlated with length phenotype. *Elongation factor 2* is an essential factor for protein synthesis. *Solute carrier family 25* proteins are involved in energy transport (Walker, 1992) and may act as ATP exchangers, catalyzing the flux of adenine nucleotides into or from mitochondria and thus playing a role in cell growth (Luciakova *et al.,* 2003). The epistasis for the two genes could result from an additive effect between the modified elongation factor, containing a nonsynonymous mutation, and the noncoding mutation in the *solute carrier family 25,* that could play a role in modulating the level of expression of this protein. Although far from being a thorough test of

the presence of epistasis in the genomes of the hybrids under study, these results nevertheless represent one of the very few evidence suggesting that epistatic effect on phenotypic variation might be a common phenomenon in salmonid fishes.

In summary, using RNA-seq, we have investigated variations in gene expression and allelic imbalance between two different crosses of wild–domestic brook char hybrids. Our results show that *cis* effects play a major role in gene expression regulatory differences among hybrids. In addition, the results also suggest that interaction among variants of *trans*-acting elements and between variants of *trans*-acting elements and *cis*-acting elements are likely to lead to dominant and over/underdominant gene expression patterns. Our study also uncovered markers that could be useful for screening advantageous or disadvantageous genotypic combinations in association with length at age in salmonids. These three genes seem to be good candidates for future studies regarding association between regulatory variation and phenotypic differences in salmonids.

DATA ARCHIVING

Genotype data have been deposited at Dryad: doi:10.5061/dryad.3g19f.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Angers B, Castonguay E, Massicotte R (2010). Environmentally induced phenotypes and DNA methylation: how to deal with unpredictable conditions until the next generation and after. *Mol Ecol* **19**: 1283–1295.

Bellin D, Ferrarini A, Chimento A, Kaiser O, Levenkova N, Bouffard P (2009). Combining next-generation pyrosequencing with microarray for large scale expression analysis in non-model species. *BMC Genomics* **10**: 555.

Bougas B, Granier S, Audet C, Bernatchez L (2010). The transcriptional landscape of cross-specific hybrids and its possible link with growth in brook charr (*Salvelinus fontinalis* mitchill). *Genetics* **186**: 97–107.

Bower NI, Li X, Taylor R, Johnston IA (2008). Switching to fast growth: the insulin-like growth factor (IGF) system in skeletal muscle of Atlantic salmon. *J Exp Biol* **211**: 3859–3870.

Croisetière S, Bernatchez L, Belhumeur P (2010). Temperature and length-dependent modulation of the MH class IIβ gene expression in brook charr (*Salvelinus fontinalis*) by a *cis*-acting minisatellite. *Mol Immunol* **47**: 1817–1829.

Derome N, Bernatchez L (2006). The transcriptomics of ecological convergence between 2 limnetic coregonine fishes (Salmonidae). *Mol Biol Evol* **23**: 2370–2378.

Derome N, Duchesne P, Bernatchez L (2006). Parallelism in gene transcription among sympatric lake whitefish (*Coregonus clupeaformis* Mitchill) ecotypes. *Mol Ecol* **15**: 1239–1249.

Elmer KR, Fan S, Gunter JC, Boekhoff S, Kuraku S, Meyer A (2010). Rapid evolution and selection inferred from the transcriptomes of sympatric crater lake cichlid fishes. *Mol Ecol* **19**: 197–211.

Fontanillas P, Landry CR, Wittkopp PJ, Russ C, Gruber JD, Nusbaum C *et al.* (2010). Key considerations for measuring allelic expression on a genomic scale using high-throughput sequencing. *Mol Ecol* **19**: 212–227.

Fraser HB, Babak T, Tsang J, Zhou Y, Zhang B, Merhabian M *et al.* (2011). Systematic detection of polygenic *cis*-regulatory evolution. *PLOS Genet* **7**: e1002023.

Gerke J, Lorenz K, Cohen B (2009). Genetic interactions between transcription factors cause natural variation in yeast. *Science* **323**: 498–501.

Gibson G, Weir B (2005). The quantitative genetics of transcription. *Trends Genet* **21**: 616–623.

Goetz F, Rosauer D, Sitar S, Goetz G, Simchick C, Roberts S *et al.* (2010). A genetic basis for the phenotypic differentiation between siscowet and lean lake trout (*Salvelinus namaycush*). *Mol Ecol* **19**: 176–196.

Graze RM, McIntyre LM, Main BJ, Wayne ML, Nuzhdin SV (2009). Regulatory divergence in *Drosophila melanogaster* and *D. simulans*, a genomewide analysis of allele-specific expression. *Genetics* **183**: 547–561.

Hughes KA, Ayroles JF, Reedy MM, Drnevich JM, Rowe K, Reidi EA *et al.* (2006). Segregating variation in the transcriptome: *cis* regulation and additivity of effects. *Genetics* **173**: 1347–1364.

Jeukens J, Renaut S, St-Cyr J, Nolte AW, Bernatchez L (2010). The transcriptomics of sympatric dwarf and normal lake whitefish (*Coregonus clupeaformis* spp., Salmonidae) divergence as revealed by next-generation sequencing. *Mol Ecol* **19**: 5389–5403.

Landry CR, Wittkopp PJ, Taubes CH, Ranz JM, Clark AG, Hartl DL (2005). Compensatory *cis-trans* evolution and the dysregulation of gene expression in interspecific hybrids of drosophila. *Genetics* **171**: 1813–1822.

Lariège A, Mangin B, Jasson S, Combes V, Dumas F, Jamin P *et al.* (2011). The genetic basis of heterosis: multiparental QTL mapping reveals contrasted levels of apparent overdominance among traits of agronomical interest in maize (*Zea mays* L.). *Genetics* 10.1534/ 111.133447.

Levine M, Tjian R (2003). Transcription regulation and animal diversity. *Nature* **424**: 147–150.

Lemos B, Araripe L, Fontanillas P, Hartl DL (2008). Dominance and the evolutionary accumulation of *cis-* and *trans*-effects on gene expression. *Proc Natl Acad Sci USA* **105**: 14471–14476.

Li N, Das K, Wu R (2009). Functional mapping of human growth trajectories. *J Theor Biol* **261**: 33–42.

Li ZK, Luo LJ, Mei HW, Wang DL, Shu QY, Tabien R *et al.* (2001). Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. biomass and grain yield. *Genetics* **158**: 1737–1753.

Luciakova K, Barath P, Poliakova D, Persson A, Nelson BD (2003). Repression of the human adenine nucleotide translocase-2 gene in growth-arrested human diploid cells. *J Biol Chem* **278**: 30624–30633.

Luo L, Li Z, Mei H, Shu Q, Tabien R, Zhong D *et al.* (2001). Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice.II. Grain yield components. *Genetics* **158**: 1755–1771.

Luo X, Fu Y, Zhang P, Wu S, Tian F, Liu J *et al.* (2009). Additive and over-dominant effects resulting from epistatic loci are the primary genetic basis of heterosis in rice. *J integr Plant Biol* **51**: 393–408.

MAPAQ (2010). *Les pêches et aquaculture commerciales. Bilan 2008 et perspectives (Les)*. Bibliothèque et Archives nationales du Québec.

Maston GA, Evans SK, Green MR (2006). Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* **7**: 29–59.

Morison IM, Ramsay JP, Spencer HG (2005). A census of mammalian imprinting. *Trends Genet* **21**: 457–465.

Nagel R (2005). Epistasis and the genetics of human diseases. *C R Biol* **328**: 606–615.

Oleksiak M, Churchill G, Crawford D (2002). Variation in gene expression within and among natural populations. *Nat Genet* **32**: 261–266.

Pierron F, Normandeau E, Defo MA, Campbell PGC, Bernatchez L, Couture P (2011). Effects of chronic metal exposure on wild fish populations revealed by high-throughput cDNA sequencing. *Ecotoxicology* **20**: 1388–1399.

St-Cyr J, Derome N, Bernatchez L (2008). The transcriptomics of life-history trade-offs in whitefish species pairs (Coregonus sp.). *Mol Ecol* **17**: 1850–1870.

Shen Y, Catchen J, Garcia T, Amores A, Beldorth I, Wagner J *et al.* (2012). Identification of transcriptome SNPs between *Xiphophorus* lines and species for assessing allele specific gene expression within F1 interspecies hybrids. *Comp Biochem Physiol C Toxicol Pharmacol* **155**: 102–108.

Walker JE (1992). The mitochondrial transported family. *Curr Opin Struct Biol* **2**: 519–526.

Wayne ML, Pan YJ, Nuzhdin SV, McIntyre LM (2004). Additivity and *trans*-acting effects on gene expression in male *Drosophila simulans*. *Genetics* **168**: 1413–1420.

Whitehead A, Crawford DL (2005). Variation in tissue-specific gene expression among natural populations. *Genome Biol* **6**: R13.

Whitehead A, Crawford DL (2006). Variation within and among species in gene expression: raw material for evolution. *Mol Ecol* **15**: 1197–1211.

Whiteley AR, Derome N, Rogers SM, St-Cyr J, Laroche J, Labbe A *et al.* (2008). The phenomics and expression quantitative trait locus mapping of brain transcriptomes regulating adaptive divergence in lake whitefish species pairs (Coregonus sp.). *Genetics* **180**: 147–164.

Wittkopp PJ (2005). Genomic sources of regulatory variation in *cis* and in *trans*. *Cell Mol Life Sci* **62**: 1779–1783.

Wittkopp PJ, Haerum BK, Clark AG (2004). Evolutionary changes in *cis* and *trans* gene regulation. *Nature* **430**: 85–88.

Wittkopp PJ, Haerum BK, Clark AG (2008a). Independent effects of *cis-* and *trans-*regulatory variation on gene expression in *Drosophila melanogaster*. *Genetics* **178**: 1831–1835.

Wittkopp PJ, Haerum BK, Clark AG (2008b). Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat Genet* **40**: 346–350.

Wolf JB, Cheverud JM, Roseman C, Hager R (2008). Genome-wide analysis reveals a complex pattern of genomic imprinting in mice. *PLOS Genet* **4**: e1000091.

Wray G (2007). The evolutionary significance of *cis* regulotary mutations. *Nat Genet* **8**: 206–216.

Xu P, McIntyre LM, Scardina J, Wheeler P, Thorgaard G, Nichols KM (2010). Transcriptome profiling of embryonic development rate in rainbow trout advanced backcross introgression lines. *Mar Biotechnol* **13**: 215–231.

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