MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 2877–2895 doi: 10.1111/j.1365-294X.2012.05579.x

Dynamics of introgressive hybridization assessed by SNP population genomics of coding genes in stocked brook charr (*Salvelinus fontinalis*)

FABIEN C. LAMAZE,* CHRISTOPHER SAUVAGE,*† AMANDINE MARIE,‡ DANY GARANT‡ and LOUIS BERNATCHEZ*

**Institut de Biologie Intégrative et des Systèmes (IBIS), Département de Biologie, Université Laval, 1030 avenue de la Médecine, Québec, QC, G1V 0A6 Canada, †INRA, UR1052, Unité de Génétique et d'Amélioration des Fruits et Légumes, 84143 Montfavet, France, ‡Département de Biologie Faculté des Sciences, Université de Sherbrooke, 2500 Boul. de l'Université, Sherbrooke, QC, J1K 2R1 Canada*

Abstract

Salmonid fishes rank among species being most severely affected by introgressive hybridization as a result of a long tradition of stocking with hatchery-reared conspecifics. The objectives of this study were (i) to evaluate the genetic consequences of stocking and resulting introgression rates on the genetic integrity of natural populations of brook charr, (ii) to identify genomic regions potentially associated with adaptation to natural and artificial rearing environments, and (iii) to test the null hypothesis that introgression from domesticated brook charr into wild populations is homogeneous among loci. A total of 336 individuals were sampled from nine lakes, which were stocked at different intensities with domestic fish. Individuals were genotyped at 280 SNPs located in transcribed regions and developed by means of next-generation sequencing. As previously reported with microsatellites, we observed a positive relationship between stocking intensity and genetic diversity among stocking groups, and a decrease in population differentiation. Individual admixture proportions also increased with stocking intensity. Moreover, genomic cline analysis revealed 27 SNPs, seven of which were also identified as outliers in a genome scan, which showed an introgression rate either more restricted or enhanced relative to neutral expectations. This indicated that selection, mainly for growth-related biological processes, has favored or hampered the introgression of genomic blocks into the introgressed wild populations. Overall, this study highlights the usefulness of investigating the impact of stocking on the dynamics of introgression of potentially adaptive genetic variation to better understand the consequences of such practice on the genomic integrity of wild populations.

Keywords: adaptation, conservation, genome scan, genomic clines, hybridization, introgression, salmonids, stocking

Received 25 November 2011; revision received 19 February 2012; accepted 21 February 2012.

Introduction

Human-mediated hybridization between indigenous populations and stocked, escaped or cultivated exogenous conspecifics is widespread worldwide, which can potentially jeopardize the resilience of locally adapted gene pools (Allendorf *et al*. 2001; Randi 2008; Fitzpatrick *et al*. 2010; Laikre *et al*. 2010). Consequences of hybridization have thus raised complex and controversial conservation issues (Allendorf *et al*. 2001; Edmands 2007; Fraser 2008). Thus, introgressive hybridization may threaten locally adapted populations through genetic swamping, loss of intrapopulation genetic variation or genetic differences between populations (Laikre *et al*. 2010). Moreover, the introgression of selected traits of interest from exogenous strains reared in captivity may

Correspondence: Fabien C. Lamaze, Fax: +1 418 656 7176; E-mail: fabien.lamaze.1@ulaval.ca

alter fitness-related traits (e.g. growth, disease resistance) of wild populations, potentially leading to outbreeding depression (Rhymer & Simberloff 1996; Allendorf *et al.* 2001; Edmands 2007; McClelland & Naish 2007). Alternatively, exogenous alleles may potentially introgress through positive selection into native populations and thus represent a source of genetic novelty that might affect the potential for adaptation to changing environments (Anderson *et al.* 2009; Fitzpatrick *et al.* 2010; Song *et al.* 2011).

Salmonids represent one of the most socio-economically important fish families, being highly exploited by means of recreational angling, commercial fisheries and aquaculture (Davidson *et al.* 2010). Domesticated, hatchery-reared conspecifics have been propagated through intensive artificial breeding programs for enhancing population abundance and production (Aprahamian *et al.* 2003). Recent studies have shown that domestication and artificial selection can induce rapid evolutionary change in a matter of as little as four generations (Roberge *et al.* 2007; Tymchuk *et al.* 2009; Sauvage *et al.* 2010), which can have detrimental consequences for wild populations when these fish are released in the wild (McGinnity *et al.* 2003; Araki *et al.* 2007, 2009; Tymchuk *et al.* 2007). As a consequence of extensive artificial propagation of non-native populations and a long tradition of supportive stocking with hatchery-reared conspecifics, salmonids are among the most genetically altered taxa by introgression (Sloss *et al.* 2008; Cooper *et al.* 2010).

Patterns of introgressive hybridization and genetic admixture resulting from fish stocking have been typically documented using a limited set of markers (Campton & Johnston 1985; Poteaux *et al.* 1999; Hansen *et al.* 2009; Hansen & Mensberg 2009; Marie *et al.* 2010; Winkler *et al.* 2010; but see Hansen *et al.* 2010; Meier *et al.* 2011), and few of these studies have investigated the dynamics of locus-specific introgressive hybridization. Both of these limitations hamper a detailed understanding of the dynamics of introgression, for instance pertaining to the role of selection in either restricting or enhancing genetic admixture in different genomic regions. As a notable exception, in a recent study on brown trout (*Salmo trutta*), Hansen *et al.* (2010) provided some evidence of selection against hatchery-specific alleles at several microsatellite markers associated with quantitative trait loci (QTL) following stocking of domesticated fish into wild populations.

Next-generation sequencing technologies (NGS) now offer a leap towards refining our understanding of the dynamics of introgressive hybridization, by easing the development and application of Single Nucleotide Polymorphism markers (SNPs), either in coding or non-coding regions of the genome (Ellegren & Sheldon 2008; Hohenlohe *et al.* 2011; Rice *et al.* 2011). In salmonids, SNPs markers have been increasingly applied for fine-scale genetic stock identification, parentage and kinship analysis and stock monitoring of Pacific salmon (*Oncorhynchus* sp.) (Campbell & Narum 2011; Hauser *et al.* 2011; Hess *et al.* 2011; Templin *et al.* 2011). To our knowledge, however, a single study based on coding gene SNP analyses has dealt with the issue of introgressive hybridization between wild and domesticated populations in salmonids (Bourret *et al.* 2011).

Brook charr (*Salvelinus fontinalis*) is one of the most important fish species for the recreo-touristic industry in Eastern Canada. For instance, 6 million hatchery-reared domestic brook charr are stocked every year in Québec only, representing 50% of the aquaculture industry trade of the province (Ministère des Ressources Naturelles et de la Faune du Québec, 2008). Recently, a study using microsatellites showed that the intensity of stocking affects the genetic integrity of wild populations by altering the level of intrapopulation genetic diversity and homogenizing population genetic structure (Marie *et al.* 2010). In this study, we expand upon previous research efforts using a novel approach to investigate the dynamics of introgression of stocked domestic brook charr in wild populations at individual loci distributed throughout the genome. Namely, we used a novel set of 280 SNP markers developed for brook charr from NGS RNA-sequencing. Each of these SNPs was positioned on a genetic map, located within a different coding gene, and several of these are associated with QTL for phenotypic traits (Sauvage *et al.* 2012b). We first document the overall extent of divergence between domestic and wild populations and also identify the most divergent loci between them, which may reflect adaptive genetic differences. Secondly, we quantify the nature and effect of stocking intensity on introgression rates into wild populations and compare those results with those obtained previously using microsatellite markers (Marie *et al.* 2010). Finally, we test the null hypothesis of no difference in rate of introgression among individual loci to identify genes that may have introgressed differentially because of either positive or negative selective effects.

Materials and methods

RNA-seq for SNP development

A set of 280 validated SNPs were recently developed from RNA (cDNA) sequences and used to build a genetic map as well as identifying QTL for phenotypic traits of aquaculture interest in brook charr (detailed in Sauvage *et al.* 2012a). Briefly, 22 fish were used for SNP identification, including brook charr from three different

strains: eight individuals from the Rupert River population, Québec (48°44′N, 68°05′W), eight from the Laval River population Québec (51°05′N, 73°41′W) and six F_1 hybrid individuals issued from crosses between 'pure' parents of the Laval River and the main Domestic strain used for stocking for many decades in Québec. No population investigated in this study was used for SNP development, thus minimizing ascertainment bias, for instance in terms of differences in allelic diversity among populations analysed here. RNA was sequenced on a GS-FLX Titanium sequencer at the Genome Québec Innovation Centre (McGill University, Montreal, Canada). The raw pyroreads are publicly available under the accession number SRX037496 (http://www.ncbi.nlm.nih.gov/sra).

SNP validation

DNA sequences were assembled into contigs, and polymorphic positions were identified using CLC Genomic Workbench version 3.7 (CLC Bio, Denmark) (see Sauvage *et al.* 2012a for details). From many thousands of putative SNPs, a subset ($n \approx 1000$) was selected for validation using the following four steps (Sauvage *et al.* 2012a). First, a pair of PCR primer pairs was designed for each SNP of interest to generate an amplicon of 250–400 bp. Amplicons sizes of over 400 bp were then removed to avoid the amplification of intronic regions. Second, both strands of the selected amplicons were sequenced to confirm real polymorphism and discard false positives. Third, genotyping assay was designed for the remaining SNPs using the iPlex Gold® protocol for the Sequenom MassARRAY® (Sequenom, San Diego, CA, USA). Loci that did not respect the Sequenom® assay technical requirements were removed. In the final step, the 280 loci that satisfied all requirements were multiplexed in panels of 28–32 markers on the MassARRAY® platform (Sequenom) according to the manufacturer's instructions, at the Genome Québec Innovation Centre.

Sampling strategy

Sampling was conducted in the Portneuf Wildlife Reserve in Québec, Canada (47°09′N, 72°17′W). Nine lakes were chosen in this reserve for their well-documented stocking histories since 1971 and categorized by stocking intensity: nonstocked (NS, $n = 3$), moderately stocked (MS, $n = 3$) and heavily stocked (HS, $n = 3$) (see Table 1 and Marie *et al.* 2010 for details). For the NS lakes, no stocking event was documented between 1971 and 2008, whereas MS and HS lakes underwent stocking in less than 50% or more than 50% of years from 1992 to 2007, respectively. Lakes were stocked on average with 5819 ± 3427 (mean ± SD) and 14 926 ± 12 930 of domestic brook charr of various size and age per lake per year, respectively, in the MS and HS lakes (Marie *et al.* 2010). All stocked brook charr from the domestic strain originated from the same private hatchery (Jacques Cartier Hatchery). A total of 336 brook charr were sampled in June 2007 and 2008 using experimental gillnets with different mesh sizes to maximize the representation of size and age classes (Table 1). Tissues were also obtained from 48 individu-

Table 1 Summary of sample collection and descriptive genetic statistics. Included are fish from the Jacques Cartier hatchery facility (HAT), Heavily stocked (HS), Moderately stocked (MS), and nonstocked (NS) lakes, names of lakes and their coordinates, sample sizes, number of polymorphic SNPs (N_P) out of 231 genotyped, mean expected heterozygosity (H_E) and observed heterozygosity (H_O) and F_{IS} for each lake based on the analysis of the 231 SNPs

Category	Lakes	Latitude	Longitude	Sample size	N_P	H_E	H_O	F_{IS}
HAT		46°43′50″N	71°45′58″W	48	231	0.352 ± 0.135	0.362 ± 0.166	−0.029
HS	Amanites (AMA)	47°06′30″N	72°22′43″W	24	205	0.313 ± 0.166	0.273 ± 0.191	0.152*
	Belles-de-Jour (BEL)	47°05′43″N	72°17′41″W	48	228	0.352 ± 0.140	0.340 ± 0.167	0.034*
	Methot (MET)	47°10′20″N	72°19′33″W	72	229	0.349 ± 0.149	0.334 ± 0.175	0.041*
	Average			144	221	0.338 ± 0.022	0.316 ± 0.037	0.076 ± 0.066
MS	Arcand (ARC)	47°14′45″N	72°23′33″W	23	178	0.248 ± 0.196	0.257 ± 0.225	−0.039
	Rivard (RIV)	47°10′19″N	72°01′56″W	21	198	0.309 ± 0.182	0.324 ± 0.224	−0.048
	Veillette (VEI)	47°09′44″N	72°01′24″W	24	197	0.318 ± 0.185	0.321 ± 0.215	−0.009
	Average			68	191	0.292 ± 0.038	0.301 ± 0.038	−0.032 ± 0.020
NS	Caribou (CAR)	47°13′33″N	72°23′25″W	24	144	0.209 ± 0.201	0.218 ± 0.234	−0.044
	Main de fer (MAI)	47°07′39″N	72°07′30″W	24	119	0.159 ± 0.196	0.173 ± 0.237	−0.091*
	Sorbier (SOR)	47°09′25″N	72°24′47″W	24	190	0.270 ± 0.189	0.285 ± 0.282	−0.055
	Average			72	151	0.213 ± 0.056	0.225 ± 0.056	−0.063 ± 0.025

*Hardy–Weinberg disequilibrium after FDR correction.

als of the hatchery (domestic) strain (HAT, Table 1) and preserved in ethanol 95% until DNA extraction.

DNA extraction and genotyping

Total DNA was extracted from adipose fin tissue according to the Qiagen DNeasy kit manufacturer's specifications (Qiagen, Valencia, USA). DNA quantification was performed using the picogreen assay (Invitrogen, Carlsbad, USA). A total of 332 individuals were genotyped using 280 validated SNPs using the iPlex Gold® assays on the MassARRAY® platform (Sequenom) according to the manufacturer's instructions at the Genome Québec Innovation Center.

SNP annotation

To identify putative gene functions and polymorphism characteristics (e.g. transition vs. transversion, synonymous vs. nonsynonymous mutations) for each validated SNP, functional annotations were documented, when possible. Blastx similarity searches were performed with the NCBI nucleotide database (nr) using the consensus sequences of each contig that contained a validated SNP. Only *e*-values higher than 1×10^{-40} were considered significant. Then, the most probable open reading frames (ORF) for each contig that exhibited a significant Blast hit was predicted using ORFPREDICTOR (Min *et al.* 2005). SNPs were characterized as synonymous or nonsynonymous from the deduced ORFs. SNP markers were named as follows: sf (for *Salvelinus fontinalis)* and a six-digit number representing the 280 different contigs in which each SNP was identified (e.g. sf000657). We also used the AmiGO browser of gene ontology (http://www.geneontology.org), the KEGG PATHWAY database (http://www.genome.jp/kegg/pathway.html) and the UniProt database (http://www.uniprot.org/), along with corresponding literature searches to assign SNPs to putative function.

Data quality check

Loci showing a call rate inferior to 90%, monomorphic markers and markers having more than 10% missing values were removed from subsequent analyses. Loci showing significant deviations from Hardy–Weinberg Equilibrium (HWE) in NS populations were also withdrawn ($N = 12$). Correction for multiple comparisons was applied on type I error in the HWE test with a false discovery rate (FDR) method ($\alpha = 0.05$) (Benjamini & Yekutieli 2001). No frequency cut-off was applied on major, nor minor, allele frequencies. Data quality was verified using the R package SNPASSOC (Gonzalez *et al.* 2007).

Genetic variation and population differentiation

Expected and observed heterozygosity within each population was quantified using the whole-SNP data set. The average deviation in heterozygosity within populations (F_{IS}) (Weir & Cockerham 1984) was estimated with GENETIX version 4.05 (Belkhir *et al.* 2004). Conformity of HWE expectations was tested using GENEPOP version 4.0.10 (Raymond & Rousset 1995) with a global test, which use the multisample score test across loci for each population to compute *P*-value (Rousset & Raymond 1995). FDR correction was applied to the significance threshold ($\alpha = 0.05$) (Benjamini & Yekutieli 2001). Pairwise genetic differentiation between populations was estimated using the F_{ST} estimator of Weir & Cockerham (1984) as calculated with GENETIX after 1000 permutations and FDR correction ($\alpha = 0.05$) (Benjamini & Yekutieli 2001). A nonparametric Kruskal–Wallis analysis of variance was used to test for differences in gene diversity (H_E) among different lake categories (NS, MS and HS).

Bayesian clustering

The proportion of individual domestic–wild genome admixture (Q) was estimated with the whole-SNP data set using first the Bayesian procedure implemented in STRUCTURE version 2.3.3 (Pritchard *et al.* 2000; Falush *et al.* 2003). Two separated simulations were performed, the first one involving HAT with the 'pure' wild populations only (NS group), and the second one involving a pair comparison between HAT with each of the lakes from the two stocked groups (MS and HS). The first simulation aimed at quantifying the extent of genetic differentiation among 'pure' wild and hatchery populations and the second aimed at quantifying individual admixture proportions, from the domestic ancestry into wild fish, in stocked populations. In both simulations, ancestry model with admixture and allele frequencies correlated were assumed, and no population information was included as a prior for simulations. The most likely number of clusters *K* in all simulations was assumed to be in the range of $K = 1$ to $K = n + 3$ (where n is the number of populations sampled; Evanno *et al.* 2005). Ten replicates were conducted for each *K* with a burn-in period of 1×10^4, followed by 5×10^4 MCMC steps. The ad hoc statistic ΔK was used to determine the most probable *K* (Evanno *et al.* 2005). Also, the lakes included in this study are all small, and there was no evidence for the occurrence of more than one population in any of them. The three final runs consisted of a burn-in period of 1×10^5 followed by 5×10^5 MCMC steps. The population admixture proportions were calculated from the second analysis as the mean of

individual admixture for MS and HS lake groups separately. Then, a Wilcoxon test was used to test the null hypothesis of no difference in mean admixture proportion between the two stocking groups MS and HS. Secondly, the program NewHybrids version 1.1 (Anderson & Thompson 2002) was used in complement of the Structure analysis for the estimation of individual assignment. NewHybrids have been shown to increase the assignment accuracy of multiple generations of hybrids when compared with Structure (Marie *et al.* 2011). The combination of both methods is thus expected to improve the assignment of 'pure' HAT, wild and hybrids groups in MS and HS lakes needed for the genomic cline (see below). Using NewHybrids, six categories corresponding to parental ('pure' domestic and 'pure' wild), F1, F2 and backcrosses (F1 × domestic and F1 × wild) were considered. Calculations were run separately between HAT and each stocked lakes three times for accuracy using Jeffrey priors. Individual posterior probabilities of belonging to each hybrid category was estimated with a burn-in period of 1×10^5 followed by 5×10^5 MCMC steps. The z option was applied to HAT individuals to indicate that these individuals were from 'pure' origin and set to zero. The s option was also applied to HAT individuals to help the program to infer HAT allele frequencies and not inflating the estimate of admixture proportion of individuals from each MS and HS lakes.

Genome scan

To identify the most highly divergent SNPs (outliers) between wild populations (NS) and HAT, we used the simulation method of Beaumont & Nichols (1996) extended to the case of a hierarchical population structure in Arlequin version 3.5 (Excoffier *et al.* 2009; Excoffier & Lischer 2010). In this model, demes are arranged into k groups (here HAT and NS) of d demes, with varying migration rates among demes within and between groups (Excoffier *et al.* 2009). Loci are defined as outliers with respect to F_{CT} among groups as well as with respect to F_{ST} among populations after accounting for group structure (Excoffier *et al.* 2009). As we were only interested in identifying the most highly divergent SNPs between NS and HAT, we only reported F_{CT} SNP outliers. Null distributions were generated assuming a hierarchical island model with 50 groups each consisting of 100 populations and based the tests on 1×10^5 simulations across loci as a function of scaled between population heterozygosity. We applied a significance cutoff of $P < 0.05$, and a FDR correction $\alpha = 0.05$ to minimize false positives (Benjamini & Yekutieli 2001). Also to reduce the number of potential false positives, we reported only loci with scaled heterozygosity >0.2, as suggested by Excoffier *et al.* (2009).

Genomic cline and locus-specific rate of introgression

Using the r package Introgress version 1.2.3 (Gompert & Buerkle 2010), we applied the genomic cline method (Gompert & Buerkle 2009) to estimate clines for individual SNPs using genotypic frequencies as a function of genome-wide admixture and to test whether these estimated clines were consistent with a null model of neutral introgression. In a manner analogous to using a geographical transect (Gompert & Buerkle 2009; Nolte *et al.* 2009; Teeter *et al.* 2010), we interpreted the stocking intensity among lakes as a putative cline of admixture proportion. Specific patterns of introgression observed for each SNP in this first analysis were further described using an ad hoc statistic based on genetic differentiation (F_{ST}) between hatchery, 'pure' wild and hybrid individuals. This second analysis also provided global information on the rates of introgression of the three different groups of loci that were identified: neutral, positively and negatively selected. Both analyses are described in details below.

For the first analysis, two 'pure' groups were created from the Structure analysis with the confirmation of NewHybrids analysis. The first was composed of individuals from the domestic strain (HAT, $n = 48$) and hatchery individuals from MS and HS lakes. In MS and HS, individuals were selected on the basis of having the 90% credibility interval (CI) surrounding their mean posterior probability, to be within the threshold of 90% of being 'pure' domestic, when comparing HAT vs. each stocked population from MS and HS (in Structure, $K = 2$). In addition, the MS and HS individuals selected by Structure should had at least 95% average category probability of being classified as 'pure' domestic individuals using NewHybrids. This resulted in the identification of 26 additional 'pure' domestic fishes from a total of 74. The second group comprised 'pure' wild individuals from MS and HS. Given the pronounced genetic structure observed between wild populations (see Results), wild individuals from NS group were excluded as they did not represent the most probable wild genetic background that could have intermixed with the domestic background in each lake. Instead, selected wild fish in each stocked lake should have a CI at least within the threshold of 90% of being of 'pure' wild genetic background, when comparing HAT vs. each stocked population from MS and HS using Structure ($K = 2$). In addition, the MS and HS individuals selected by Structure should have at least 95% average category probability of being classified as 'pure' wild individuals using NewHybrids, for a total of 77 individuals. Then, a third group called 'hybrid' was composed of the remaining individuals of potentially admixed ancestry ($n = 98$). Individuals were considered hybrids if their CI around their mean admixture estimates either did not include the maximum posterior probability

of being of 'pure' genetic background or if CI included posterior probability over a 10% threshold (STRUCTURE analysis, $K = 2$). Eleven individuals classified as backcrossed (F1 × wild) by NEWHYBRIDS analysis were added, for a total of 109 'hybrid' individuals.

The analysis was conducted by first estimating the differential in allele frequencies between the two 'pure' groups to use in the null allele simulations and to calculate the hybrid index, which ranged between 0 ('pure' domestic) and 1 ('pure' wild). A genome-wide admixture was then estimated by maximum likelihood with 95% confidence intervals for the admixed individuals (of the third group). To identify SNPs with patterns of introgression that were inconsistent with neutral expectations, a null model of expected neutral introgression was compared with observed locus-specific genomic clines by a multinomial regression (Gompert & Buerkle 2009). Clines of genotypic frequencies for each SNP genotype (i.e. homozygotes for domestic alleles; heterozygotes, and homozygotes for wild alleles) were fitted using a multinomial regression and used to estimate the likelihood of the regression model given the observed genotypic data (i.e. homozygotes and heterozygotes at each SNP). A parametric simulation model was used for the estimation of the genomic cline under neutral expectations, as it does not assume uniformity of allelic differences across loci between parental populations (Gompert & Buerkle 2009). Significant testing was performed using 1000 parametric simulations for each SNP to obtain a distribution of expected log-likelihood ratios and variance of the neutral model because of stochastic sampling (Gompert & Buerkle 2009).

The genomic cline analysis provided a basis to interpret deviations from neutral expectations and to identify the three categories of SNP markers defined above. For homozygous clines, steep clines are expected for alleles with decreased rates of introgression relative to neutral expectations, whereas shallow clines are expected for alleles with increased rates of introgression relative to neutral expectations. For heterozygous clines, bell-shaped clines below the neutrally expected envelope (i.e. underdominance) are expected for alleles with decreased rates of introgression relative to neutral expectations, whereas bell-shaped clines above the neutrally expected envelope (i.e. overdominance) are expected for markers with increased rates of introgression relative to neutral expectations. Finally, for both types of genotypic frequency, neutrally introgressed loci are identified by clines not deviating significantly from null expectations (Gompert & Buerkle 2009). To minimize the inclusion of false positives, markers that were significant (i.e. potentially under selection) but whose genomic clines could not be visually classified as positively (shallow clines) or negatively selected (steep clines) because their clines nearly parallel those for both 'neutral' homozygous and heterozygous genotypes were discarded from subsequent analyses. A FDR correction was applied ($\alpha = 0.05$) to control for multiple testing (Benjamini & Yekutieli 2001).

The groups of SNPs identified as being under selection (either positive or negative) were further analysed as follows. In this second analysis, an ad hoc statistic based on genetic differentiation (F_{ST}, Weir & Cockerham 1984) was used to globally visualize the results of the genomic cline analyses. To do so, the same two 'pure' groups were considered but the 'hybrid' group from the genomic cline analysis was divided into subsets of individuals meant to consider post-F1 hybrids (introgressed) only. The first subset comprised the least introgressed individuals with hybrid index proportion, calculated by INTROGRESS, ranging from 0.70 to 0.95 ($n = 29$), whereas the second subset comprised the most introgressed individuals with hybrid index proportion ranging from 0.05 to 0.30 ($n = 14$). These thresholds were set to exclude F1 hybrid individuals. Next, linear regressions were applied to locus-specific F_{ST} values estimated (i) between the two post-F1 groups (dependent variable) and (ii) the two 'pure' groups defined above (independent variable). These regressions were calculated separately for the three types of markers used in the genomic cline analysis (i.e. positively and negatively selected, and neutral SNPs). Prior to linear regression, a Shapiro–Wilk normality test was performed as well as visual inspection of quantile plots. Three possible scenarios were predicted. First, if the two regressions for the negatively and positively introgressed groups of markers identified by the genomic cline analysis fall within the 95% confidence interval of neutrally evolving SNPs regression, this would indicate that overall, alleles for all markers are introgressing neutrally and at the same rate from the domestic strain into the wild populations. Alternatively, a regression falling significantly above the 95% confidence interval of the neutrally evolving SNP regression would indicate slower introgression (and thus higher overall differentiation) than neutrally expected, thus suggesting an overall negative selective effect on rate of introgression. Finally, a regression falling below the 95% confidence interval of the neutrally evolving SNP regressions would indicate faster introgression than neutrally expected (resulting in lower differentiation), thus suggesting an overall positive selective effect on rate of introgression.

Results

SNP genotyping, annotation and data quality

Out of the 280 SNPs assayed, 263 were successfully genotyped and 250 were polymorphic. Seven markers

with a call rate inferior to 90% and 12 that strongly departed from HWE in NS populations with a *P*-value corrected by FDR (with $\alpha = 0.05$) were excluded, for a total of 231 high quality, polymorphic SNPs retained for further analyses. Four individuals out of 336 were removed from subsequent analyses as they presented over 10% of missing values (three from RIV and one from ARC lakes—Table 1). Thus, the final data set comprised 332 fish (see Table 1 for details). Among the 231 validated SNPs, 36.8% (85/231) exhibited a significant blast hit (*e*-value $<1 \times 10^{-40}$). Within these 85 SNPs, ORF predictor identified 84 ORFs (98.8%—84/85), among which only 11 were nonsynonymous. The other markers (63.2% or 146/231) were considered as 'unknown'.

Genetic variation and differentiation

The mean expected heterozygosity per population ranged from 0.159 to 0.352 (Table 1). Mean gene diversity within group was proportional to stocking intensity: HS (0.338 ± 0.151) >MS (0.291 ± 0.190) >NS (0.211 ± 0.201) ($K = 157.40$, d.f. = 3 and $P < 2.20 \times 10^{-16}$). All pairwise F_{ST} estimates across all loci were significant after FDR correction ($\alpha = 0.05$; adjusted threshold $P = 0.0003$). This translated into a global F_{ST} value of 0.178 ± 0.091, reflecting an overall level of medium to high differentiation among populations. F_{ST} values ranged from 0.010 between HAT and BEL (a HS population) to 0.364 between MAI and CAR (2 NS populations) Table S1 (Supporting Information). Also, mean F_{ST} among populations within group was inversely proportional to the stocking intensity: HS (0.071 ± 0.014), MS (0.125 ± 0.084) and NS (0.296 ± 0.061) (Fig. 1). Similarly the extent of genetic differentiation between the hatchery strain (HAT) and wild populations was inversely proportional to the intensity of stocking (HAT by HS $F_{ST} = 0.068 \pm 0.033$, HAT by MS $F_{ST} = 0.150 \pm 0.068$ and HAT by NS $F_{ST} = 0.281 \pm 0.065$) (Fig. 1).

Bayesian clustering

Using STRUCTURE for the comparison between HAT and NS population, the ΔK statistic (Evanno *et al.* 2005) showed multiple peaks for $K = 2$, 4 and 6 (Table S2, Supporting Information). The two clusters with the most pronounced structure differences corresponded to HAT vs. all three NS populations. The four clusters inferred corresponded to the four populations: HAT, CAR, MAI and SOR, while the six clusters partitioned the HAT population in three clusters (Fig. 2). The mean domestic admixture in NS populations was for $K = 6$ in CAR 0.016 ± 0.043, in MAI 0.004 ± 0.007 and in SOR 0.019 ± 0.027. For the paired analysis between HAT and each of the stocked lakes, the ΔK statistic showed a single peak for $K = 2$ (Table S2, Supporting Information). The two clusters inferred corresponded to the two populations of the paired comparison. The overall mean individual admixture between wild and hatchery brook charr in the HS samples was high but highly variable among individuals (0.461 ± 0.430), with a mean admixture of 0.479 ± 0.447 for AMA, 0.688 ± 0.413 for BEL and 0.304 ± 0.367 for MET. The mean admixture was very low (0.025 ± 0.056) in the MS populations and statistically different from HS populations ($W = 8518$, d.f. = 1, $P < 2.20 \times 10^{-16}$). Admixture proportion observed for MS populations was 0.036 ± 0.084 for ARC, 0.028 ± 0.048 for RIV, and 0.012 ± 0.011 for VEI. The range of individual admixture proportion within each of the three lakes of the HS group varied from 0.001 to 0.998 in AMA, 0.008 to 0.992 in BEL, and 0.006 to 0.993 in MET, respectively. Patterns of individual admixture differed among these three lakes (Fig. 3). In BEL, 52.1% ($n = 25$) of individuals were classified as either F1 or post-F1 hybrids by having their CI estimates within the range of 10–90%, as determined by the STRUCTURE analysis, $K = 2$. In BEL, the estimated proportion of individuals from domestic ancestry was 33.3% ($n = 16$) and that of 'pure' wild individuals was 14.6% ($n = 7$). In MET, 68.0% ($n = 49$) of individuals were classified as either F1 or post-F1 hybrids. The proportion of individuals from domestic ancestry represented only 2.8% ($n = 2$), whereas wild individuals represented 29.2% ($n = 21$). In AMA, hybrids represented only 37.5% ($n = 9$), with 33.3% ($n = 8$) from 'pure' domestic and 29.2% ($n = 7$) from 'pure' wild individuals ancestry. In the MS group, the individual

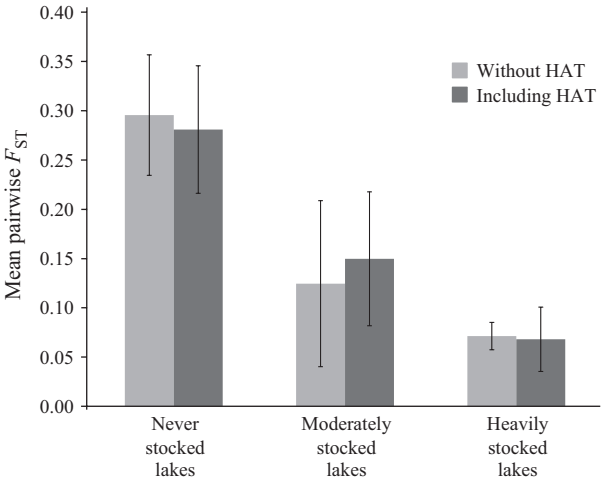

Fig. 1 Average pairwise F_{ST} (±SD) based on 231 SNPs as a function of stocking intensity. Light grey: among lakes within each group; Dark grey Hatchery fish vs. natural populations within each group.

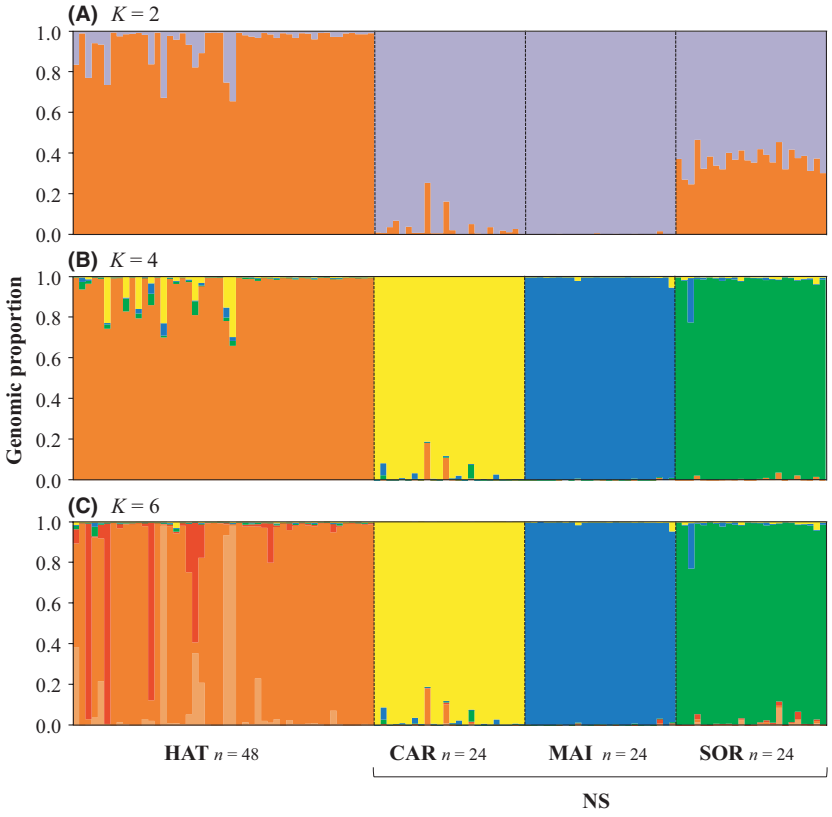

Fig. 2 Individual genomic proportion based on 231 SNPs using the STRUCTURE clustering with admixture and allele frequency correlated. (A) $K = 2$ corresponds to the uppermost level of structuring between the domestic vs. all wild fish from the NS lakes. (B) $K = 4$ corresponds to the differentiation between the four populations (HAT and the three NS lakes). (C) $K = 6$ corresponds to substructure detected within the HAT population. The *y*-axis depicts the genomic proportion belonging (coancestry) to one of the populations from either the Portneuf Wildlife Reserve or the hatchery strain (HAT). Each column corresponds to an individual, and sample locations are separated by vertical dotted bars. NS, nonstocked lakes, *n*, number of individuals.

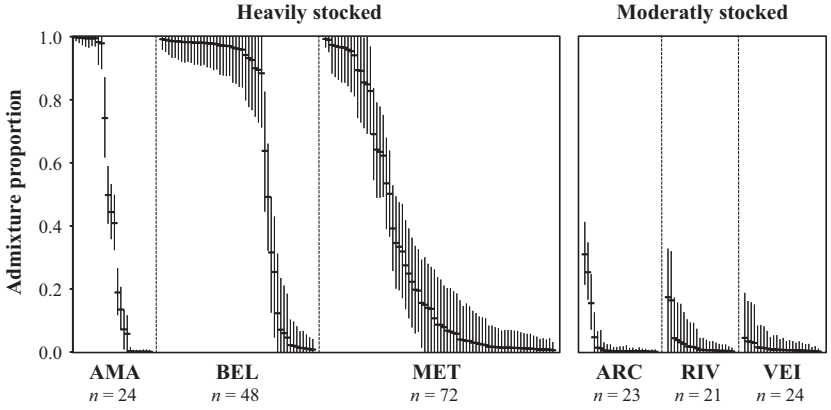

Fig. 3 Individual admixture proportions and 90% credible intervals based on 231 SNPs, estimated by comparisons between the reference hatchery (HAT) and each of the stocked lakes from MS and HS groups. STRUCTURE clustering was used with admixture, allele frequency correlated and $K = 2$ as a prior for the model. The *y*-axis depicts the proportion of the genome of HAT genetic background. *n*, number of individuals.

admixture proportion within each of the three lakes ranged from 0.001 to 0.310 in ARC, 0.002 to 0.174 in RIV, and 0.002 to 0.045 in VEI, and no fish were assigned to the 'pure' HAT group (Fig. 3). In ARC, RIV and VEI Lakes, the majority of fish were of 'pure' wild genetic background ($Q \leq 0.022$) (82.6%, $n = 19$; 66.7%, $n = 14$; 83.3%, $n = 20$; respectively) (Fig. 3).

NEWHYBRIDS analysis generally supported the results obtained by STRUCTURE, except for 11 fish. Namely, seven individuals scored by STRUCTURE as 'pure' wild fish in BEL lake were classified either as backcrosses (F1 × wild, $n = 5$) or as F2 ($n = 2$) by NEWHYBRIDS. In RIV, four of the 14 individuals classified as being 'pure' wild with STRUCTURE analysis were classified as backcrosses (F1 × wild) by

NEWHYBRIDS. The remaining individuals from AMA, MET, ARC and VEI were all assigned similarly by the two programs. Overall then, 109 individuals were considered as hybrids, 77 as 'pure' wild and 26 as 'pure' domestic.

Genome scan

The genome scan between wild (NS) and HAT populations identified 18 outliers representing the most highly divergent genes between hatchery and wild fish (Fig. 4). Table 2 presents a description and names of these markers, as well as their gene annotation and function when available. Average F_{ST} value across all loci was 0.30, and F_{CT} value across all loci was 0.07 between HAT and NS, whereas outlier F_{CT} values ranged from 0.37 up to 0.78. These loci were not linked as they mapped on separate linkage groups except for sf000966 and sf004614 (Sauvage *et al.* 2012a). Among the 18 outlier markers, one corresponded to metalloproteinase inhibitor 3 (F_{CT} = 0.713) and a second one to cysteine- and glycine-rich protein 1 (F_{CT} = 0.499), two were associated with QTL related to growth functions (hepatic glycogen and plasmatic glucose) with F_{CT} of 0.772 and 0.732, respectively, one was associated with a QTL related to stress response (plasma cortisol, F_{CT} = 0.563) and one with a QTL related to reproduction (sperm concentration, F_{CT} = 0.635) (Table 2).

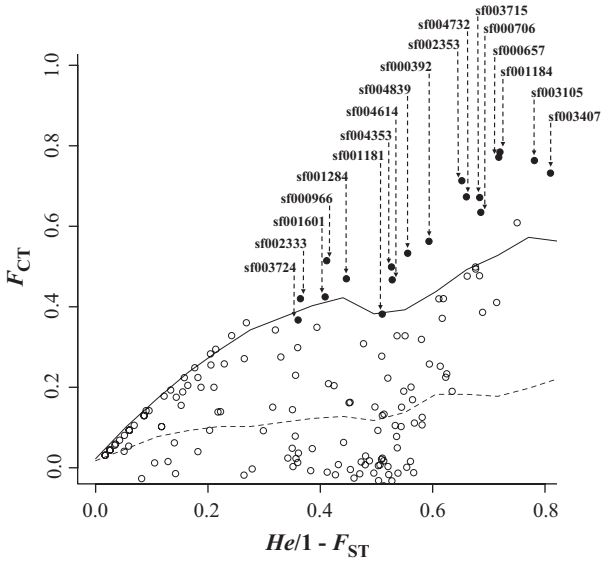

Fig. 4 Hierarchical genomic scan between the domestic strain (HAT) and the group of nonstocked (NS) lakes (CAR, MAI, SOR). The solid black line represents upper 95% confidence levels, and the dotted line indicates the median value of F_{CT} (net differentiation between HAT and NS group). Black dots represent the most divergent loci between HAT and NS lakes from the Portneuf Wildlife Reserve ($P < 0.05$).

Genomic cline

A total of 68.8% (159/231) of the entire SNP data set could be classified with the genomic cline analysis. The other 92 loci had a significant *P*-value but exhibited uncharacteristic patterns of introgression, with almost parallel clines for the homozygous and heterozygous genotypes, which hampered discriminating and interpreting their introgression patterns. Among these 159 SNPs, 13 showed a significantly slower introgression rate from domestic into the wild genetic background than neutrally expected (Table 3). Conversely, 14 SNPs were characterized by a significantly faster introgression rate (Table 3), and 133 SNPs did not significantly depart from neutral introgression. Six outliers identified by the genome scan also deviated from neutral expectations in the genomic cline analysis, while the remaining outlier loci in the genome scan did not depart from a neutral introgression model. Examples of markers with significantly steeper (reduced introgression) and shallower (accelerated introgression) clines than expected given neutral introgression are provided in Fig. 5. Loci sf001184 and sf004149 are two examples of reduced introgression rate: the homozygote cline was steeper than predicted by the neutral model (sf001184), while the heterozygote cline presented a pattern of under-dominance (sf004149). Locus sf002155 is an example of loci showing accelerated introgression rate whereby the homozygote genomic cline was less steep than neutral expectation. Finally, locus sf003105 is an example of neutral introgression with both homozygote and heterozygote clines falling within the 95% confidence interval of simulated neutral genomic clines.

Linear regressions of F_{ST} values between dependent (F_{ST} values between the least vs. the most introgressed groups of hybrid fish) and independent ('pure' wild vs. 'pure' hatchery) variables were highly significant for the three groups of markers, as follows: (i) neutral (Pearson's product-moment correlation, $R^2 = 0.75$, $t = 13.04$, d.f. = 130, $P < 2.20 \times 10^{-16}$ after rejecting H_0 of the Shapiro–Wilk test, $W = 0.85$, $P = 3.15 \times 10^{-10}$), (ii) reduced rate of introgression ($R^2 = 0.95$, $F = 228.24$, d.f. = 1, $P = 3.58 \times 10^{-9}$), and (iii) accelerated rate of introgression ($R^2 = 0.81$, $F = 56.84$, d.f. = 1, $P = 4.25 \times 10^{-6}$) (Fig. 6). Also, the confidence intervals of the three regression lines were largely nonoverlapping. Among the 13 markers from the reduced introgression group, one was annotated to proteasome subunit beta type-9, and two were associated with QTL related to growth functions (hepatic glycogen and size). Among the 14 markers from the accelerated introgression group, three were associated with QTL related to growth functions (plasmatic glucose, size and weight) and one related to reproduction (sperm concentration). Finally, one QTL related to growth was also annotated to malate dehydrogenase (Table 3).

Table 2 Summary of the most highly differentiated loci (outliers) between the hatchery strain and wild populations using a hierarchical model implemented in ARLEQUIN. Included are marker names, genetic differentiation among group (F_{CT}), *P*, probability of being a nonoutlier between hatchery and wild (NS) populations, Linkage group (LG) and QTL associated trait according to the genetic map developed by Sauvage *et al.* (2012b). Blast hits (accession number), gene annotation when available and Gene ontology

SNP markers	F_{CT}	*P*	LG	QTL	Accession number	Annotation	Biological processes
sf000392	0.563	0.023	23	Plasma cortisol			Stress
sf000657	0.772	0.003*	2	Hepatic glycogen			Growth
sf000706	0.635	0.020	6	Sperm concentration			Reproduction
sf000966	0.514	0.011	13				
sf001181	0.382	0.042	NA				
sf001184	0.784	0.002*	1		FQ310508.1	NA	NA
sf001284	0.470	0.028	35				
sf001601	0.425	0.037	8				
sf002333	0.420	0.033	NA				
sf002353	0.713	0.005	20		NM_001141843	Metalloproteinase inhibitor 3 (TIMP3)	GO:0007568: ageing GO:0007417: central nervous system development GO:0051045: negative regulation of membrane protein ectodomain proteolysis GO:0043200: response to amino acid stimulus GO:0043627: response to oestrogen stimulus GO:0051593: response to folic acid GO:0009725: response to hormone stimulus GO:0009612: response to mechanical stimulus GO:0014070: response to organic cyclic compound GO:0050896: response to stimulus GO:0042246: tissue regeneration GO:0007601: visual perception
sf003105	0.763	0.014	3				
sf003407	0.732	0.031	10	Plasmatic glucose	EZ804050.1	NA	Growth
sf003715	0.671	0.014	11		EU816603	NA	NA
sf003724	0.367	0.047	NA				
sf004353	0.499	0.021	42		NM_001124574.1	Cysteine- and glycine-rich protein 1 (CRP1)	GO:0042074: cell migration involved in gastrulation GO:0060027: convergent extension involved in gastrulation GO:0003007: heart morphogenesis GO:0007254: JNK cascade GO:0016055: Wnt receptor signalling pathway
sf004614	0.467	0.027	13				
sf004732	0.673	0.010	31		EU025707.1	NA	NA
sf004839	0.533	0.020	41				

NA, nonavailable.
*Significant after correction for multiple testing (FDR).

Discussion

Using a set of 280 coding SNP genes developed specifically in brook charr, the main objectives of this study were to (i) document the overall extent of divergence between domestic and wild populations, (ii) quantify the nature and effect of stocking intensity on introgression rates into wild populations, and (iii) test the null

Table 3 Summary of the 27 SNPs showing significantly slower or faster rate of introgression relative to neutral expectation from the HAT (hatchery) strain to the wild populations from stocked lakes (MS and HS). Included are introgression type, locus name, description of SNP polymorphism, (ORF) position of the SNP within the open reading frame when available, the amino acid substitution (aa), assignment to linkage group (LG) according to Sauvage *et al.* (2012a), Weir & Cockerham's (1984) F_{ST} estimates fish from 'pure' domestic and wild genetic backgrounds from the stocked lakes (MS and HS), the genomic cline probability for departure from neutrality and QTL association according to Sauvage *et al.* (2012b). Blast hits (accession number), gene annotation when available and Gene ontology

Introgression	Locus	SNP	ORF	aa	LG	F_{ST}	*P*	QTL	Accession numbers	Annotation	Biological processes
Slow	sf000132	A/G			13	0.166	0.038				
	sf000612	A/T			37	0.035	<0.001*				
	sf000657†	C/T			2	0.772	0.039	Hepatic glycogen			Growth
	sf001105	A/G			17	0.127	0.047				
	sf001184†	C/T			1	0.591	0.038				
	sf001601†	C/T			8	0.421	0.003				
	sf002333†	A/G			NA	0.120	0.009				
	sf003018	A/G			26	0.055	<0.001*	Size			Growth
	sf003092	C/T	135	Leu	9	0.047	<0.001*				
	sf003614	G/T			8	0.078	<0.001*				
	sf003755	A/G			7	0.129	<0.001*				
	sf004149	C/G			6	0.319	0.034		AF465752.1	Proteasome subunit beta type-9 (PSMB9)	Proteolysis
	sf004839†	C/T	124	Gln/stop	39	0.537	0.022				
Fast	sf000124	A/G	62	Asp/Gly	17	0.160	0.032				
	sf000178	A/G	162	Ala	19	0.070	0.001*				
	sf000891	A/C	79	Arg	6	0.222	0.019	Sperm concentration			Reproduction
	sf001181†	A/G	6	Thr	NA	0.070	0.001*				
	sf001831	A/C	255	Glu/Asp	40	0.121	0.041				
	sf002155	A/G			NA	0.468	0.018		EZ778883.1	NA	NA
	sf002175	C/T			34	0.162	0.026				
	sf002250	C/A			28	0.169	0.009				
	sf003334	G/A			24	0.153	0.048	Size	HQ287747.1	Malate dehydrogenase (MDH)	Growth
	sf003407†	G/A			10	0.628	0.005	Plasmatic glucose	EZ804050.1	NA	Growth
	sf003442	A/G	39	Ser	24	0.254	0.004	Weight			Growth
	sf004260	G/A	76	Leu/Phe	NA	0.313	0.007				
	sf004438	T/C	83	Met/Thr	NA	0.159	<0.001*		BT045405.1	NA	NA
	sf006073	A/G	140	Asp/Gly	39	0.337	0.007				

*Significant *P*-value after FDR correction.
†Putative outlier loci in the hierarchical genome scan analysis.

hypothesis of difference in rate of introgression among SNP markers. We found that the main domestic strain used for stocking in Québec was highly genetically differentiated from wild populations at SNP markers located in coding genes. Also, the overall rate of introgression at these markers was proportional to stocking intensity. We also observed that the intensity and patterns of individual introgression are highly variable among stocked lakes. Finally, the genomic cline analysis refuted the hypothesis that no differences in the rates of introgression among individual loci exist. It also identified two groups of loci that significantly departed either positively or negatively from expected neutral rate of introgression.

Given that most previous studies have largely used microsatellites for the investigation of the genetic impact of stocking in salmonids and that there is a current trend in the increased use of SNP markers in population and conservation genetics, it is worth first to contrast results obtained from these two genotyping approaches. The hatchery strain and the nine wild populations investigated in this study were recently charac-

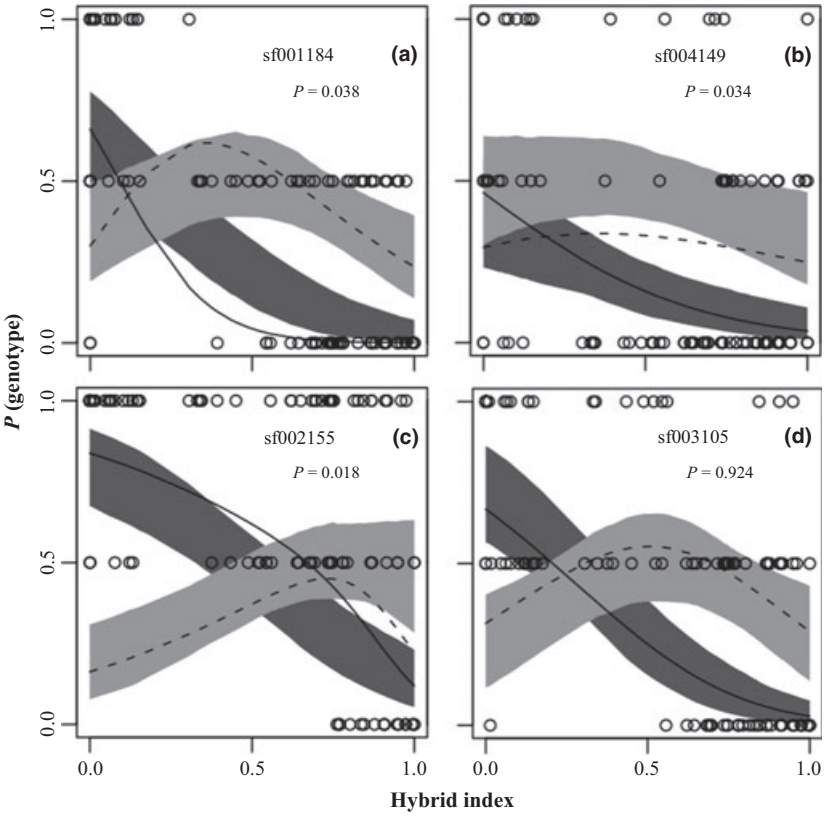

Fig. 5 Example of four multinomial regression fits of markers contrasted with a genomic cline along an admixture gradient between the hatchery and wild brook charr in six stocked lakes. Plots depict the probability of domestic genotypes [1.0 = homozygous for domestic alleles (DD); 0.5 = heterozygous (DW); 0.0 = homozygous for wild alleles (WW)] as a function of the hybrid index, which quantifies the fraction of wild alleles across all the 231 markers. (A) Genomic clines for a SNP with slow rate of introgression with a steep homozygote cline; (B) SNP with slow rate of introgression with under-dominance; (C) SNP with fast rate of introgression with a shallow homozygote cline; (D) SNP with no difference from neutral expectation. The 95% confidence intervals for genomic clines given neutral introgression are in dark grey for the frequency of homozygote (DD) and in light grey for the frequency of heterozygote (DW) genotypes. The solid line and dashed lines illustrate the estimated cline based on the observed DD and DW genotypes, respectively. Circles indicate the empirical genotypic data (DD on top line, DW in centre, and WW along the bottom). The name of each locus and their *P*-value for the test of departure from neutrality is given.

terized with microsatellites by Marie *et al.* (2010), allowing a direct comparison of results on the exact same individuals. We found that the extent of population divergence between the hatchery and the three wild unstocked populations was congruent with that obtained by microsatellites. Marie *et al.* (2010) reported a F_{ST} value of 0.23 among the same NS lakes, compared with 0.30 in this study. F_{ST} values among moderately and HS lakes were, respectively, 0.15 and 0.05 based on microsatellites, compared with 0.13 and 0.07 in this study. Both studies also revealed a significant increase in genetic diversity with increasing stocking intensity. The mean individual admixture proportions were also comparable: 0.451 vs. 0.461 for the HS lakes and 0.023 vs. 0.025 for the MS lakes (Marie *et al.* 2010 vs. this study). Thus, this study reiterates some of the main conclusions of Marie *et al.* (2010, 2011) that stocking with a domestic strain impairs the genetic integrity of wild brook charr populations by (i) homogenizing the genetic structure of wild populations, (ii) increasing of intrapopulation genetic diversity because of the introduction of new alleles from the domestic genetic background, and (iii) causing introgression of the domestic genetic background into that of wild populations. These observations also corroborate those reported for other salmonids (e.g. Eldridge *et al.* 2009; Hansen & Mensberg 2009; Consuegra *et al.* 2011; Perrier *et al.* 2011; but see Taylor *et al.* 2007).

We also observed different patterns of introgression among lakes: within the HS populations, we observed either a continuum of individual admixture value (MET lake) or a dichotomy between 'pure' individuals of domestic vs. wild genetic background with few discrete introgressed individuals (AMA lake). A previous study

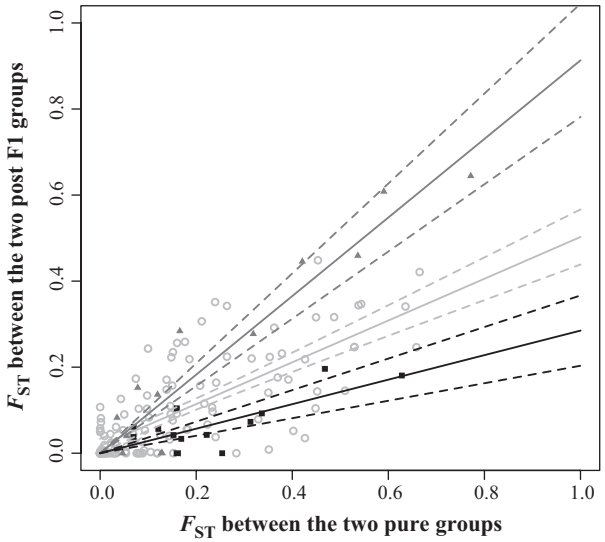

Fig. 6 Linear regressions for significant (groups of SNP showing fast or slow introgression) and neutral (nonsignificant) SNP markers from genomic clines analysis. The plot depicts the F_{ST} values between the least vs. the most introgressed groups of post-F1 hybrids (on the *y*-axis) and the F_{ST} between the two 'pure' group (domestic and wild) from moderately (MS) and heavily (HS) stocked lakes and the hatchery (HAT) (on the *x* axis). Dark grey triangles are markers with reduced introgression rate. Black squares are markers with fast introgression rate. Light grey circles are neutrally introgressed markers. Light grey, dark grey and black lines and dotted lines of the same colour are regressions and 95% confidence interval; respectively for neutral, slower and faster introgressed groups of markers. Equations of the three linear regressions were: $y = 0.53x + 0.02$; $y = 0.91x$; $y = 0.27x$, respectively, for neutral, slower and faster introgressed groups of markers.

has also reported variable patterns of introgression for a comparable level of stocking intensity in coho salmon (*Oncorhynchus kisutch*) populations (Eldridge *et al.* 2009). These results suggest that biotic or abiotic environmental factors, other than stocking intensity, play a role in shaping the extent and pattern of introgressive hybridization. A recent analysis performed on brook charr by Marie *et al.* (2012) showed that the level of hybridization among lakes significantly increased with a reduction in both surface area and maximum depth of lakes, a reduction in dissolved oxygen and an increase in temperature and pH. Thus, the extent of introgressive hybridization observed in each lake was influenced by the availability and quality of lacustrine habitats. Angling pressure could also affect levels of introgressive hybridization (e.g. Fraser 1981; Lachance & Magnan 1990; Marie *et al.* 2012).

Besides reinforcing the results previously obtained with microsatellites, this study improves our detailed understanding of the dynamics of introgressive hybridization between stocked domesticated and wild brook charr in two important ways. First, the hierarchical genome scan identified 18 outlier genes (7.8% of all genes scanned) that were highly divergent between domestic and wild fish, which could have been driven by artificial selection. Admittedly, genome scan results must be interpreted cautiously as it is prone to problems of false positives (Teshima *et al.* 2006; Beaumont 2008; Foll & Gaggiotti 2008; Excoffier *et al.* 2009). On the other hand, the pronounced population structure observed in this study could reduce the power of detecting outliers as it has been shown that the power of several genome scan methods (including Bayesian methods) is reduced when values of population differentiation (F_{ST}) > 0.2 (Pérez-Figueroa *et al.* 2010). Even though the distinction between selective and stochastic effects on detected outliers is not conclusive, our results nevertheless identify the most highly differentiated genes between wild fish and hatchery fish used for stocking and that deserve particular attention in future studies in this system.

Secondly, genomic cline analysis performed in this study identified two groups of SNP markers for which rates of introgression have either been enhanced ($n = 14$) or reduced ($n = 13$) relative to neutral expectation. This suggests a role for either positive or negative selective effects in shaping differential rates of introgression among the 231 coding SNP markers genotyped in this study. While the genomic cline analysis provides a specific statistical framework by which locus-specific patterns of introgression could be tested, this method requires an accurate estimation of parental allelic frequencies as well as a precise estimation of genome-wide admixture for hybrid individuals to better estimate the significance of the test. The 231 SNP markers used in this study revealed a pronounced difference in allele frequencies between domesticated and wild fish, such that it was possible to accurately estimate the individual admixture proportion (with a narrow confidence interval) using STRUCTURE, with results also confirmed using NEWHYBRIDS. Thus, we could confidently distinguish pure hatchery, pure wild and hybrid fish in each stocked lake. Furthermore, temporal stochasticity associated with genetic drift might lead to false positives in the parametric model. However, considering the small number of generations since admixture (Marie *et al.* 2010), this effect is unlikely to bias the results importantly (Gompert & Buerkle 2011). Also, the results of the genomic clines analysis are probably conservative given that we only retained markers that were significant and also showed a clear pattern of either reduced or enhanced rates of introgression, as described by Gompert & Buerkle (2009; 2011). Moreover, predictions of regressions based on F_{ST} are concordant with the locus-specific regressions from the genomic cline

analysis, thus confirming the general hypothesis about directionality of the selection acting on markers that departed from neutral expectation of introgression.

Seven of the loci departing from neutral expectations in the genomic cline analyses were also outliers in the hierarchical genomic scan performed between hatchery and wild populations. Moreover, patterns of introgression of five of these outliers were consistent with reduced rate of introgression relative to neutral expectations. This suggests the effect of selection acting against introgression from the hatchery into wild populations for these genes, or the genomic regions linked to them. This is consistent with previous studies on the impact of stocking in salmonids that reported selection against hatchery-introduced alleles in admixed populations (Hansen *et al.* 2010; Bourret *et al.* 2011; Meier *et al.* 2011). However, our results also identified two outlier markers showing a faster rate of introgression than expected under neutrality. This suggests that selection enhanced introgression from the hatchery into wild populations for these two genes, or the genomic regions linked to them.

Although annotation was not possible for many of the SNP markers used in this study, some of the most highly differentiated loci between domestic and wild brook charr could be annotated to functional genes and biological processes and are further corroborated by the genomic cline analysis. Thus, four outliers corresponded to QTLs that are relevant to phenotypic differences commonly observed between domestic and wild salmonid populations, in terms of growth, behaviour, and reproduction (Heath *et al.* 2003; Jonsson & Jonsson 2006; Tymchuk *et al.* 2006). Two outliers also corresponded to annotated genes. The gene metalloproteinase inhibitor 3 (TIMP3) is of particular interest as other genes of the TIMP family previously showed differential patterns of gene expression between compared domestic and wild populations of salmonids, including brook charr (Roberge *et al.* 2008; Bougas *et al.* 2010; Sauvage *et al.* 2010). TIMP3 is a physiological inhibitor of ADAM 12-S which cleaves insulin like growth factor (IGF) (Loechel *et al.* 2000), suggesting that this gene could have been affected by artificial selection for higher growth rate. The second gene, cysteine- and glycine-rich protein 1 (CSRP1), is also of interest as it is involved in early regulatory, development, and cellular differentiation processes. In salmonids, embryonic development rate has been associated with survival and physiological performance later in life (Einum & Fleming 2000; Sundstrom *et al.* 2005; Renaut & Bernatchez 2011; Xu *et al.* 2011). Thus, variation (either structural or regulatory) in TIMP3 and CSRP1 could potentially affect growth rate and differential development between domestic and wild populations.

Admittedly, the nature of factors affecting the rate of the introgression remains hypothetical, especially when considering the discrepancy associated with the rates of introgression of outlier loci linked to growth-related traits with some showing whether faster or slower rate of introgression relative to neutral expectations. It has been showed that over-winter mortality may select against faster growing introgressed rainbow trout (Vandersteen *et al.* 2012). Because domestic fish do not experience fasting periods (or carbohydrate-poor diets) in aquacultural facilities, it is possible that domestic alleles at some outliers with reduced introgression rate, such as the one associated with hepatic glycogen QTL (Sauvage *et al.* 2012b) could be disadvantageous in nature, given its potential metabolic function during fasting (Larsen *et al.* 2001). Conversely, a SNP marker linked to a plasmatic glucose QTL showed significantly enhanced introgression rate, suggesting a fitness advantage for wild introgressed fish. Besides the seven genes that were also outliers between domestic and wild fish, there were 20 other markers that significantly departed from neutral expectation in the genomic cline analyses. Among them, five were either linked to a QTL or annotated gene, and of these, three were related to growth (Sauvage *et al.* 2012b) and two of these showed a faster rate of introgression than expected under neutrality. This, in addition to knowledge on biological functions, supports the hypothesis that introgression from domestic to wild fish may provide fitness advantages associated with growth in the wild. For example, malate dehydrogenase (MDH), an enzyme involved in the gluconeogenesis, is downregulated in farmed Canadian Atlantic salmon compared with wild counterpart (Roberge *et al.* 2006) and also in normal (fast growing) lake whitefish ecotype compared with dwarf sympatric ecotype (*Coregonus clupeaformis*) (Jeukens *et al.* 2010). MDH is implicated in adaptive metabolic divergence, in whitefish, consistent with the observed trade-off in life history traits among whitefish ecotypes where regulatory or structural variations of this enzyme could be implicated in the lower metabolic rate observed in the normal ecotype (Jeukens & Bernatchez 2011). The reduction of the metabolic rate may in turn allow reallocation of energetic resources towards growth (Roberge *et al.* 2006).

We also observed that a QTL associated with sperm concentration showed an enhanced introgression rate from domestic to wild populations. It has been recently argued that hatchery induced sperm competition may lead to artificial selection for faster growing males by not equalizing milt volume in multi-male fertilization (Wedekind *et al.* 2007). Sperm concentration has been hypothesized to enhance the competitiveness for fertilization (Gage & Morrow 2003; but see Gage *et al.* 2004).

Thus, artificial selection may have favored the recent evolution of different sperm concentration between domestic and wild fish which could influence fitness of introgressed fish in the wild.

In contrast to the above examples of markers that showed enhanced rate of introgression, the gene proteasome subunit beta 9 (PSMB9) showed a reduced rate of introgression. PSMB9 is activated during viral infection, is involved in the major histocompatibility class I antigen presentation pathway (Hansen & La Patra 2002), and is found in the class I 'core' region (Lukacs *et al.* 2007). Interestingly, PSMB9 have been localized close to a recombination hot spot in mouse (Baudat & de Massy 2007). One could then hypothesize that recombination event during hybridization could cause a breakdown of coadaptive immune gene complexes, which could result in the observed slower introgression rate of domestic alleles associated with this locus.

Overall, the results of this study reveal a complex pattern of introgression with highly variable dynamics along the same lines as other recently reported fish genomic cline studies (Nolte *et al.* 2009; Gagnaire *et al.* 2011). One possible explanation for this general observation is certainly that individual markers analysed here represent large chromosomal regions potentially containing a large number of genes (Sauvage *et al.* 2012b). Larger size QTL regions may have important implications for chromosomal segments carrying genes that negatively affect fitness, especially in first backcross generations typically characterized by larger segments in linkage disequilibrium which are more likely to carry genes that could have negative fitness effects (Martinsen *et al.* 2001). Because selection will act upon the linked genes together, the possibility is increased for negative interactions between introgressing and receiver genomes (Barton & Hewitt 1985). Thus, a single negatively selected gene could determine the outcome of all the linked genes, even if there are many positively selected genes in a segment, and could at least partly explain the complex patterns of introgression we observed. Also, QTLs associated with the expression of a same phenotypic trait may be distributed over several linkage groups, with unpredictable epistatic interactions resulting from hybridization, and causing gene misregulation (Bougas *et al.* 2010). Together these results underscore the variable and mosaic nature of hybrid genomes and illustrate the potency of recombination and selection in promoting variable and often unpredictable genetic outcomes (Martinsen *et al.* 2001; Fitzpatrick *et al.* 2010; Song *et al.* 2011).

This study complements previous work based on microsatellite loci by showing that stocking with a domestic strain affects the genetic integrity of wild populations (change in diversity, homogenization of population structure, increased individual genetic admixture) not only at neutral markers, but also at coding genes or chromosomal regions containing QTLs. Although a causal link between introgressive hybridization and fitness effects on natural populations has not been explicitly supported to date at the genome level, this study suggests that current stocking practices have the potential to significantly alter the functional genetic make-up of wild populations. Thus, our results showed important genetic differences between domestic and wild brook charr. The rate of introgression of several genes associated with the highest level of divergence was reduced between hatchery and wild fish, probably resulting in negative selection and reduced fitness of introgressed wild fish. While a reduction in the rate of introgression may be seen as a positive result for conservation purposes, introgression at such genes nevertheless occurs (albeit at a slower pace) and could be important when stocking is maintained over a longer time period. On the other hand, this study also corroborates the view that introgression is not always necessarily associated with negative effects, as shown by evidence of enhanced rate of introgression at some loci which is best explained by positive selective effects. Thus, from a conservation and management perspective, the genomic heterogeneity of introgression found in this study raises fundamental questions concerning the 'goods and the bads' of introgressive hybridization on wild populations. Namely, as advocated by the recently proposed conceptual framework of conciliation biology, managers may need to take into consideration the eco-evolutionary dynamics commonly resulting from interactions between native and non-native populations (Carroll 2011). Such ways of thinking could lead to more efficient conservation practices, as opposed to considering such populations as lost 'pure' natural populations, with little relevance for conservation. One positive step would be to acknowledge the value of partially admixed individuals and populations that function like the native populations (Fitzpatrick *et al.* 2010). In this perspective, when faster domestic allele introgressions are challenging the view of absolute genetic integrity of native populations, forthcoming research should examine the potential association between positively selected alleles and traits of ecological importance.

Acknowledgements

We are grateful to Nathalie N. Brodeur, Pierre-Alexandre Gagnaire, Gregory Maes and Scott A. Pavey for their insightful comments and help during the analyses. We are also thankful to Associate editor Loren Rieseberg and three anonymous referees for their very constructive inputs. We also thank to

Bruno Mayot, Marc-André Poulin for their help with field sampling, and Marie-Hélène Perreault for laboratory assistance. We also want to thank A. Montpetit and A. Belisle from Genome Québec Innovation Center (McGill University, Montreal, Canada) for their assistance in RNA-Seq and SNP genotyping. This research was financially supported primarily by a strategic project grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to LB and DG, the Canadian Research Chair in Genomics and Conservation of Aquatic Resources to LB and The Collaborative Research and Training Experience (CREATE) program to FL. We also acknowledge the important contributions of the Ministère des Ressources Naturelles et de la Faune du Québec (MRNF), the Société des Établissement de Plein-Air du Québec (SÉPAQ) for supporting this project. This study is a contribution to the research program of Réseau Aquaculture Québec (RAQ).

References

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology and Evolution*, **16**, 613–622.

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, **160**, 1217–1229.

Anderson TM, vonHoldt BM, Candille SI *et al.* (2009) Molecular and evolutionary history of melanism in north american gray wolves. *Science*, **323**, 1339–1343.

Aprahamian MW, Martin Smith K, McGinnity P, McKelvey S, Taylor J (2003) Restocking of salmonids – opportunities and limitations. *Fisheries Research*, **62**, 211–227.

Araki H, Cooper B, Blouin MS (2007) Genetic effects of captive breeding cause a rapid, cumulative fitness decline in the wild. *Science*, **318**, 100–103.

Araki H, Cooper B, Blouin MS (2009) Carry-over effect of captive breeding reduces reproductive fitness of wild-born descendants in the wild. *Biology Letters*, **5**, 621–624.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.

Baudat F, de Massy B (2007) Cis- and trans-acting elements regulate the mouse Psmb9 meiotic recombination hotspot. *Plos Genetics*, **3**, e100.

Beaumont MA (2008) Selection and sticklebacks. *Molecular Ecology*, **17**, 3425–3427.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the National Academy of Sciences of the United States of America*, **263**, 1619–1626.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004) *GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations*. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier, France.

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**, 1165–1188.

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.

Bourret V, O'Reilly PT, Carr JW, Berg PR, Bernatchez L (2011) Temporal change in genetic integrity suggests loss of local adaptation in a wild Atlantic salmon (*Salmo salar*) population following introgression by farmed escapees. *Heredity*, **106**, 500–510.

Campbell NR, Narum SR (2011) Development of 54 novel single-nucleotide polymorphism (SNP) assays for sockeye and coho salmon and assessment of available SNPs to differentiate stocks within the Columbia River. *Molecular Ecology Resources*, **11**, 20–30.

Campton DE, Johnston JM (1985) Electrophoretic evidence for a genetic admixture of native and nonnative rainbow trout in the Yakima River, Washington. *Transactions of the American Fisheries Society*, **114**, 782–793.

Carroll SP (2011) Conciliation biology: the eco-evolutionary management of permanently invaded biotic systems. *Evolutionary Applications*, **4**, 184–199.

Consuegra S, Phillips N, Gajardo G, de Leaniz CG (2011) Winning the invasion roulette: escapes from fish farms increase admixture and facilitate establishment of non-native rainbow trout. *Evolutionary Applications*, **4**, 660–671.

Cooper AM, Miller LM, Kapuscinski AR (2010) Conservation of population structure and genetic diversity under captive breeding of remnant coaster brook trout (*Salvelinus fontinalis*) populations. *Conservation Genetics*, **11**, 1087–1093.

Davidson WS, Koop BF, Jones SJM *et al.* (2010) Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biology*, **11**, 403.

Edmands S (2007) Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Molecular Ecology*, **16**, 463–475.

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

Eldridge WH, Myers JM, Naish KA (2009) Long-term changes in the fine-scale population structure of coho salmon populations (*Oncorhynchus kisutch*) subject to extensive supportive breeding. *Heredity*, **103**, 299–309.

Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. *Nature*, **452**, 169–175.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. *Heredity*, **103**, 285–298.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Fitzpatrick BM, Johnson JR, Kump DK, Smithc JJ, Vossc SR, Shaffer HB (2010) Rapid spread of invasive genes into a threatened native species. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 3606–3610.

Foll M, Gaggiotti O (2008) A Genome-Scan method to identify selected loci appropriate for both dominant and codominant markers: a bayesian perspective. *Genetics*, **180**, 977–993.

Fraser JM (1981) Comparative survival and growth of planted wild, hybrid, and domestic strains of brook trout (*Salvelinus fontinalis*) in Ontario lakes. *Canadian Journal of Fisheries and Aquatic Sciences*, **38**, 1672–1684.

Fraser DJ (2008) How well can captive breeding programs conserve biodiversity? A review of salmonids. *Evolutionary Applications*, **1**, 535–586.

Gage MJG, Morrow EH (2003) Experimental evidence for the evolution of numerous, tiny sperm via sperm competition. *Current Biology*, **13**, 754–757.

Gage MJG, Macfarlane CP, Yeates S, Ward RG, Searle JB, Parker GA (2004) Spermatozoal traits and sperm competition in atlantic salmon: relative sperm velocity is the primary determinant of fertilization success. *Current Biology*, **14**, 44–47.

Gagnaire PA, Minegishi Y, Zenboudji S, Valade P, Aoyama J, Berrebi P (2011) Within-population structure highlighted by differential introgression across semipermeable barriers to gene flow in *anguilla marmorata*. *Evolution*, **65**, 3413–3427.

Gompert Z, Buerkle AC (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology*, **18**, 1207–1224.

Gompert Z, Buerkle AC (2010) Introgress: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, **10**, 378–384.

Gompert Z, Buerkle CA (2011) Bayesian estimation of genomic clines. *Molecular Ecology*, **20**, 2111–2127.

González JR, Armengol L, Solé X *et al.* (2007) SNPassoc: an R package to perform whole genome association studies. *Bioinformatics*, **23**, 644–645.

Hansen JD, La Patra S (2002) Induction of the rainbow trout MHC class I pathway during acute IHNV infection. *Immunogenetics*, **54**, 654–661.

Hansen MM, Mensberg KL (2009) Admixture analysis of stocked brown trout populations using mapped microsatellite DNA markers: indigenous trout persist in introgressed populations. *Biology Letters*, **5**, 656–659.

Hansen MM, Fraser DJ, Meier K, Mensberg KL (2009) Sixty years of anthropogenic pressure: a spatio-temporal genetic analysis of brown trout populations subject to stocking and population declines. *Molecular Ecology*, **18**, 2549–2562.

Hansen MM, Meier K, Mensberg KL (2010) Identifying footprints of selection in stocked brown trout populations: a spatio-temporal approach. *Molecular Ecology*, **19**, 1787–1800.

Hauser L, Baird M, Hilborn RAY, Seeb LW, Seeb JE (2011) An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Molecular Ecology Resources*, **11**, 150–161.

Heath DD, Heath JW, Bryden CA, Johnson RM, Fox CW (2003) Rapid evolution of egg size in captive salmon. *Science*, **299**, 1738–1740.

Hess JE, Matala AP, Narum SR (2011) Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of chinook salmon in the Columbia River Basin. *Molecular Ecology Resources*, **11**, 137–149.

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. *Molecular Ecology Resources*, **11**, 117–122.

Jeukens J, Bernatchez L (2011) Regulatory versus coding signatures of natural selection in a candidate gene involved in the adaptive divergence of whitefish species pairs (*Coregonus* spp.). *Ecology and Evolution*, **2**, 258–271.

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. *Molecular Ecology*, **19**, 5389–5403.

Jonsson B, Jonsson N (2006) Cultured Atlantic salmon in nature: a review of their ecology and interaction with wild fish. *Ices Journal of Marine Science*, **63**, 1162–1181.

Lachance S, Magnan P (1990) Performance of domestic, hybrid, and wild strains of brook trout, *Salvelinus fontinalis*, after stocking: the impact of intra- and interspecific competition. *Canadian Journal of Fisheries and Aquatic Sciences*, **47**, 2278–2284.

Laikre L, Schwartz MK, Waples RS, Ryman N (2010) Compromising genetic diversity in the wild: unmonitored large-scale release of plants and animals. *Trends in Ecology and Evolution*, **25**, 520–529.

Larsen DA, Beckman BR, Dickhoff WW (2001) The effect of low temperature and fasting during the winter on metabolic stores and endocrine physiology (insulin, insulin-like growth factor-I, and thyroxine) of coho salmon, *Oncorhynchus kisutch*. *General and Comparative Endocrinology*, **123**, 308–323.

Loechel F, Fox JW, Murphy G, Albrechtsen R, Wewer UM (2000) ADAM 12-S cleaves IGFBP-3 and IGFBP-5 and is inhibited by TIMP-3. *Biochemical and Biophysical Research Communications*, **278**, 511–515.

Lukacs MF, Harstad H, Grimholt U *et al.* (2007) Genomic organization of duplicated major histocompatibility complex class I regions in Atlantic salmon (*Salmo salar*). *BMC Genomics*, **8**, 251.

Marie AD, Bernatchez L, Garant D (2010) Loss of genetic integrity correlates with stocking intensity in brook charr (*Salvelinus fontinalis*). *Molecular Ecology*, **19**, 2025–2037.

Marie AD, Bernatchez L, Garant D (2011) Empirical assessment of software efficiency and accuracy to detect introgression under variable stocking scenarios in brook charr (*Salvelinus fontinalis*). *Conservation Genetics*, **12**, 1215–1227.

Marie AD, Bernatchez L, Garant D (2012) Environmental factors correlate with hybridization in stocked brook charr (*Salvelinus fontinalis*). *Canadian Journal of Fisheries and Aquatic Sciences*, In press.

Martinsen GD, Whitham TG, Turek RJ, Keim P (2001) Hybrid populations selectively filter gene introgression between species. *Evolution*, **55**, 1325–1335.

McClelland EK, Naish KA (2007) What is the fitness outcome of crossing unrelated fish populations? A meta-analysis and an evaluation of future research directions. *Conservation Genetics*, **8**, 397–416.

McGinnity P, Prodohl P, Ferguson K *et al.* (2003) Fitness reduction and potential extinction of wild populations of Atlantic salmon, *Salmo salar*, as a result of interactions with escaped farm salmon. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **270**, 2443–2450.

Meier K, Hansen MM, Bekkevold D, Skaala O, Mensberg KL (2011) An assessment of the spatial scale of local adaptation in brown trout (*Salmo trutta* L.): footprints of selection at microsatellite DNA loci. *Heredity*, **106**, 488–499.

Min XJ, Butler G, Storms R, Tsang A (2005) OrfPredictor: predicting protein-coding regions in EST-derived sequences. *Nucleic Acids Research*, **33**, W677–W680.

Ministère des Ressources Naturelles et de la Faune du Québec (2008) Lignes directrices sur les ensemencements de poissons. Secteur Faune Québec, *Direction de l'expertise sur la faune et ses habitats*. Québec. p. 41

Nolte AW, Gompert Z, Buerkle AC (2009) Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Molecular Ecology*, **18**, 2615–2627.

Pérez-Figueroa A, García-Pereira MJ, Saura M, Rolán-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, **23**, 2267–2276.

Perrier C, Guyomard R, Bagliniere J-L, Evanno G (2011) Determinants of hierarchical genetic structure in Atlantic salmon populations: environmental factors vs. anthropogenic influences. *Molecular Ecology*, **20**, 4231–4245.

Poteaux C, Bonhomme F, Berrebi P (1999) Microsatellite polymorphism and genetic impact of restocking in mediterranean brown trout (*Salmo trutta* L.). *Heredity*, **82**, 645–653.

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

Randi E (2008) Detecting hybridization between wild species and their domesticated relatives. *Molecular Ecology*, **17**, 285–293.

Raymond M, Rousset F (1995) Genepop (Version-1.2): population-genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Renaut S, Bernatchez L (2011) Transcriptome-wide signature of hybrid breakdown associated with intrinsic reproductive isolation in lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Heredity*, **106**, 1003–1011.

Rhymer JM, Simberloff D (1996) Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics*, **27**, 83–109.

Rice AM, Rudh A, Ellegren H, Qvarnstrom A (2011) A guide to the genomics of ecological speciation in natural animal populations. *Ecology Letters*, **14**, 9–18.

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

Roberge C, 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.

Roberge C, Normandeau E, Einum S, Guderley H, Bernatchez L (2008) Genetic consequences of interbreeding between farmed and wild Atlantic salmon: insights from the transcriptome. *Molecular Ecology*, **17**, 314–324.

Rousset F, Raymond M (1995) Testing heterozygote excess and deficiency. *Genetics*, **140**, 1413–1419.

Sauvage C, Derôme N, Normandeau E, St-Cyr J, Audet C, Bernatchez F (2010) Fast transcriptional responses to domestication in the brook charr *Salvelinus fontinalis*. *Genetics*, **185**, 105–112.

Sauvage C, Vagner M, Derôme N, Audet C, Bernatchez L (2012a) Coding gene single nucleotide polymorphism mapping and quantitative trait loci detection for physiological reproductive traits in brook charr, *Salvelinus fontinalis*. *G3 (Bethesda, Md.)*, **2**, 379–392.

Sauvage C, Vagner M, Derôme N, Audet Cl, Bernatchez L (2012b) Coding gene single nucleotide polymorphism mapping and quantitative trait loci linked to growth and stress response in brook charr (*Salvelinus fontinalis*). *G3 (Bethesda, Md.)*, In press.

Sloss BL, Jennings MT, Franckowiak R, Pratt DM (2008) Genetic identity of brook trout in Lake Superior south shore streams: potential for genetic monitoring of stocking and rehabilitation efforts. *Transactions of the American Fisheries Society*, **137**, 1244–1251.

Song Y, Endepols S, Klemann N *et al.* (2011) Adaptive introgression of anticoagulant rodent poison resistance by hybridization between old world mice. *Current Biology*, **21**, 1296–1301.

Sundstrom LF, Lohmus M, Devlin RH (2005) Selection on increased intrinsic growth rates in coho salmon, *Oncorhynchus kisutch*. *Evolution*, **59**, 1560–1569.

Taylor E, Tamkee P, Sterling G, Hughson W (2007) Microsatellite DNA analysis of rainbow trout (*Oncorhynchus mykiss*) from western Alberta, Canada: native status and evolutionary distinctiveness of ''Athabasca'' rainbow trout. *Conservation Genetics*, **8**, 1–15.

Teeter KC, Thibodeau LM, Gompert Z, Buerkle CA, Nachman MW, Tucker PK (2010) The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution*, **64**, 472–485.

Templin WD, Seeb JE, Jasper JR, Barclay AW, Seeb LW (2011) Genetic differentiation of Alaska chinook salmon: the missing link for migratory studies. *Molecular Ecology Resources*, **11**, 226–246.

Teshima KM, Coop G, Przeworski M (2006) How reliable are empirical genomic scans for selective sweeps? *Genome Research*, **16**, 702–712.

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

Tymchuk WE, Sundstrom LF, Devlin RH (2007) Growth and survival trade-offs and outbreeding depression in rainbow trout (*Oncorhynchus mykiss*). *Evolution*, **61**, 1225–1237.

Tymchuk WE, Sakhrani D, Devlin R (2009) Domestication causes large-scale effects on gene expression in rainbow trout: analysis of muscle, liver and brain transcriptomes. *General and Comparative Endocrinology*, **164**, 175–183.

Vandersteen W, Biro P, Harris L, Devlin R (2012) Introgression of domesticated alleles into a wild trout genotype and the impact on seasonal survival in natural lakes. *Evolutionary Applications*, **5**, 76–88.

Wedekind C, Rudolfsen G, Jacob A, Urbach D, Muller R (2007) The genetic consequences of hatchery-induced sperm competition in a salmonid. *Biological Conservation*, **137**, 180–188.

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

Winkler KA, Pamminger-Lahnsteiner B, Wanzenbock J, Weiss S (2010) Hybridization and restricted gene flow between native and introduced stocks of alpine whitefish (*Coregonus* sp.) across multiple environments. *Molecular Ecology*, **20**, 456–472.

Xu P, McIntyre LM, Scardina J, Wheeler PA, Thorgaard GH, Nichols KM (2011) Transcriptome profiling of embryonic development rate in rainbow trout advanced backcross introgression lines. *Marine Biotechnology*, **13**, 215–231.

The authors are broadly interested in the nature of genetic changes that are associated with domestication and its consequence to the conservation genetics of wild populations. This study is part of F.L.'s doctoral research in L.B.'s laboratory, which aims to document the genomic bases of adaptive divergence between domestic and wild brook charr and the resulting genetic and epigenetic consequences of hybridization. C.S. is interested in the domestication processes and understanding the genetic basis of artificial selection to develop marker-assisted selection. A.M. aims to document the genetic consequences and ecological success of stocking in brook charr. D.G. and L.B.'s research focuses on understanding the patterns and processes of molecular and organismal evolution as well as their significance to conservation.

Data accessibility

Complementary DNA sequences: NCBI SRA: SRX037496.

Description of the Sequenom panel uploaded as online supplementary material.

Individual SNP genotypes and the sequences of the contig build from the assembly of the 454 raw data uploaded as online supplementary material.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Pairwise genetic differentiation (F_{ST}) based on 231 SNPs markers. All comparisons are significant after FDR correction ($P < 0.05$).

Table S2 STRUCTURE statistic output and the ad hoc statistic of Evanno *et al.* (2005) to determine the best K for all comparisons.

Table S3 Description of the Sequenom panel used to genotype the SNP markers in the individuals.

Table S4 Sequences of the contig build from the assembly of the 454 raw data.

Table S5 individual-by-individual SNP genotype data.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.