

Parallel and nonparallel genome-wide divergence among replicate population pairs of freshwater and anadromous Atlantic salmon

CHARLES PERRIER,* VINCENT BOURRET,* MATTHEW P. KENT† and LOUIS BERNATCHEZ*

**Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec City, Québec, Canada, G1V 0A6*, †*Department of Animal and Aquaculture Sciences, Centre for Integrative Genetics (CIGENE), Norwegian University of Life Sciences, PO Box 5003, Aas 1432, Norway*

Abstract

Little is known about the genetic basis differentiating resident and anadromous forms found in many salmonid species. Using a medium-density SNP array, we documented genomic diversity and divergence at 2336 genetically mapped loci among three pairs of North American anadromous and freshwater Atlantic salmon populations. Our results show that across the genome, freshwater populations have lower diversity and a smaller proportion of private polymorphism relative to anadromous populations. Moreover, differentiation was more pronounced among freshwater than among anadromous populations at multiple spatial scales, suggesting a large effect of genetic drift in these isolated freshwater populations. Using nonhierarchical and hierarchical genome scans, we identified hundreds of markers spread across the genome that are potentially under divergent selection between anadromous and freshwater populations, but few outlier loci were repeatedly found in all three freshwater–anadromous comparisons. Similarly, a sliding window analysis revealed numerous regions of high divergence that were nonparallel among the three comparisons. These last results show little evidence for the parallel evolution of alleles selected for in freshwater populations, but suggest nonparallel adaptive divergence at many loci of small effects distributed through the genome. Overall, this study emphasizes the important role of genetic drift in driving genome-wide reduction in diversity and divergence in freshwater Atlantic salmon populations and suggests a complex multigenic basis of adaptation to resident and anadromous strategies with little parallelism.

Keywords: anadromous, divergence, freshwater, outliers, *salmo salar*, SNP

Received 5 March 2013; revision received 19 August 2013; accepted 26 August 2013

Introduction

Delineating the genetic basis of neutral and adaptive divergence of populations is a central objective of evolutionary biology. On the one hand, when populations experience different environments, local selection regimes can drive phenotypic divergence and modulate the underlying genomic architecture promoting local adaptation and ultimately initiating evolutionary diversification and speciation (Schluter 2000; Wu 2001; Kawecki & Ebert 2004; Nosil *et al.* 2009). In addition, facultative

long-distance migrations among breeding and feeding sites may also lead to adaptive genomic divergence and evolutionary diversification due to selection on the ‘migration gene package’ (Colbert *et al.* 2001; Jonsson & Jonsson 2011; Liedvogel *et al.* 2011). On the other hand, neutral processes (e.g. effects of random dispersal and drift) also shape the genomic composition of populations. Therefore, a critical assessment of the mechanisms driving neutral and adaptive genomic divergence is required to delineate the relative importance of such processes in wild, isolated populations. Such knowledge of the historical and contemporary processes that shaped the genomic architecture of populations is also increasingly considered for the management and conservation of wild

Correspondence: Charles Perrier, Fax: +1 418 656 7176; E-mail: charles.perrier.1@ulaval.ca

populations (Fraser & Bernatchez 2001; Nielsen *et al.* 2009; Seeb *et al.* 2011).

Local adaptation in salmonids and particularly in Atlantic salmon has been widely documented (Taylor 1991; Garcia de Leaniz *et al.* 2007; Dionne *et al.* 2008; Fraser *et al.* 2011; Bourret *et al.* 2013a,b). Environmental conditions can differ among populations and impose differential selection on life-history characteristics. In fact, numerous studies documented phenotypic adaptations associated with migration in salmonids (Schaffer & Elson 1975; Stabell 1984; Hansen *et al.* 1993; Quinn 1993; Dittman & Quinn 1996; Jonsson & Jonsson 2011). The capability to migrate from freshwater to saltwater and *vice versa* involves morphological, behavioural and physiological adaptations (Folmar & Dickhoff 1980), including differential expression of genes involved in osmoregulatory functions (Hubert *et al.* 2007; Seear *et al.* 2010; Boulet *et al.* 2012). Some of these traits have been shown to be highly heritable (Hara *et al.* 2007; Nichols *et al.* 2008; Duston *et al.* 2011), which support their potential for selection across populations. The avenue of high-resolution genome-wide studies now allows improving our knowledge on the underlying genetic bases of differentiation between resident and anadromous (migratory) salmonids. For instance, Hecht *et al.* (2013) and Hale *et al.* (2013) identified many loci associated with migratory traits distributed throughout the genome, suggesting a complex, genome-wide multigenic basis of migration in the rainbow trout (*Oncorhynchus mykiss*). Several studies also documented that SNPs tagging alleles for Na/K ATPase, a candidate gene for differences in salinity tolerance, were repeatedly found to be highly differentiated between marine and freshwater stickleback populations (Hohenlohe *et al.* 2010; DeFaveri *et al.* 2011; Jones *et al.* 2012; Deagle *et al.* 2013).

Salmonids are also recognized for the near-ubiquitous occurrence of freshwater (resident) and anadromous (migratory) populations within the same species (Fleming & Reynolds 2004). While most Atlantic salmon populations are anadromous, the isostatic rebound following the last glaciation led to the independent emergence of landlocked populations, which complete their entire life cycle exclusively in freshwater (Ward 1932; Berg 1985; King *et al.* 2007). Previous studies largely based on a few number of microsatellite loci have reported strong genetic divergence between these populations and their anadromous counterparts, while also showing a reduced genetic diversity in freshwater populations (Tessier & Bernatchez 1999, 2000; Saisa *et al.* 2005; Tonteri *et al.* 2007; Ozerov *et al.* 2010; Bourret *et al.* 2013b). Besides, several studies have shown that freshwater individuals differ in a similar manner from their anadromous counterparts for multiple traits including morphology (smaller size), development rates and

osmoregulation-related traits (Taylor 1991; Kazakov & Veselov 1998; Hendry *et al.* 2003). Remarkably, several studies found that some individuals from freshwater populations have lost the ability to prepare for saltwater by upregulating osmoregulatory functions and do not survive saltwater transfer (Burton & Idler 1984; Nilsen *et al.* 2003, 2007). Therefore, it is relevant to document the patterns of genomic divergence between anadromous and freshwater Atlantic salmon populations to look for outlier markers or genomic regions that may be implicated in the repeated phenotypic divergence and the adaptation of these populations to a resident strategy.

In this study, we used a combination of recently developed genomic tools for North American Atlantic salmon including a 6-k SNP array (Bourret *et al.* 2013b) and a high-resolution linkage map (Brenna-Hansen *et al.* 2012) to study neutral and adaptive divergence underlying genetic differentiation and adaptations between populations with divergent life histories. We genotyped three pairs of freshwater and anadromous Atlantic salmon populations from Québec, Canada, to assess their respective genome-wide diversity and divergence among populations. We specifically aimed to (i) compare the levels of genetic diversity among freshwater and anadromous populations and determine the extent of polymorphism exclusive to each group; (ii) compare levels of divergence between freshwater and between anadromous populations at multiple spatial scales; (iii) test for evidence of parallel and nonparallel adaptive divergence among anadromous and freshwater populations through the identification of divergent outlier markers and genomic regions; and (iv) identify biological functions potentially under selection.

Material and methods

Study sites and samples

A total of 203 sexually mature (>40 cm total length) Atlantic salmon sampled from 10 sites (19 to 25 individuals per site) were included in this study (Fig. 1, Table 1). We sampled two population from the Musquaro River basin and eight populations from the Saguenay River basin (Fig. 1). In particular, we sampled an anadromous population from the Musquaro River (population n°1; ANA1), a freshwater population from Musquaro Lake (n°2; FRE1), four proximate anadromous populations from four rivers flowing into the Saguenay Fjord (n°3, 4, 5, 6; ANA2) and four proximate freshwater populations from tributaries of Lake St Jean draining into the Saguenay River (n°7 corresponding to group FRE2 and n°8, 9, 10 corresponding to group FRE3). The Saguenay River basin and Musquaro River basin are located

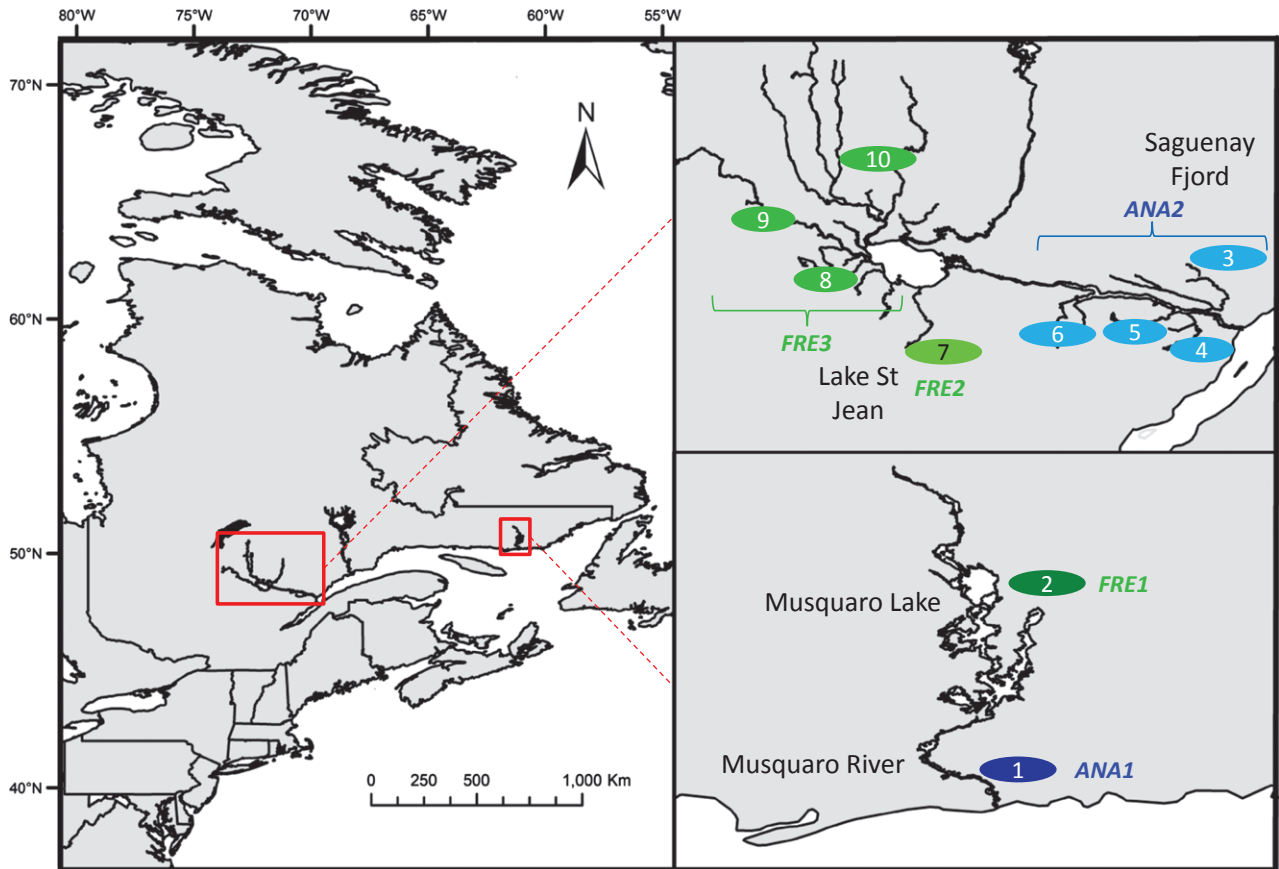


Fig. 1 Map of study locations. Population 1: Musquaro River, 2: Musquaro Lake, 3: Sainte Marguerite R., 4: Petit Saguenay R., 5: Saint Jean R., 6: Mars R., 7: Métabetchouane, 8: Aux saumons R., 9: Ashuapmushuan R., 10: Ouasiemsca R. ANA1, ANA2, FRE1, FRE2 and FRE3 correspond to group of populations being used in hierarchical analyses.

Table 1 Number of individuals genotyped (*N*) at 2336 polymorphic SNPs (MAF ≥ 0.05 per population or overall), number and proportion of polymorphic SNPs and expected heterozygosity (H_E)

Group	Population	<i>n</i>	Polymorphic SNPs		H_E over the 2336 SNPs
			Number	Proportion	
ANA1	1	25	1837	0.79	0.216
FRE1	2	19	910	0.39	0.128
ANA2	3	20	2005	0.86	0.225
	4	20	1995	0.85	0.226
	5	20	2042	0.87	0.223
	6	20	1998	0.86	0.230
FRE2	7	20	673	0.29	0.090
FRE3	8	19	870	0.37	0.115
	9	20	850	0.36	0.113
	10	20	856	0.37	0.114
Overall	10	203	2336	1.00	0.168

approximately 850 km apart on the north shore of the St. Lawrence River estuary. The Saguenay Fjord was the most important colonization route used by anadromous

fish to colonize Lake Saint Jean from 10 250 BP to 7350 BP (Bernatchez 1997). Waterfalls that rose during postglacial continental rebound represent impassable barriers for upstream migration of anadromous fish to lakes in both systems. On the basis of seven microsatellite loci, the Métabetchouane population (FRE2) has been previously found to be highly differentiated from the three other populations from the Lake St Jean (FRE3), possibly due to independent colonization events (Tessier & Bernatchez 2000). FRE2 and FRE3 were thus used as a local replicates for testing divergence among freshwater and anadromous populations for the Saguenay–Lake St Jean system.

Genotyping

DNA samples from populations 1 and 2 were extracted as reported by Dionne *et al.* (2008), and DNA samples 3 to 10 were extracted using a phenol–chloroform protocol as reported in Tessier & Bernatchez (2000). Individuals were screened for 5568 SNP markers according to the manufacturer’s instructions using the Illumina infinium assay (Illumina, San Diego, CA, USA)

and version 2 of the Atlantic salmon SNP array, following the procedure described by Bourret *et al.* (2013b). Briefly, 55% of the markers were derived from expressed sequence tag (EST); 43%, from genome complexity reduction (GCR); and the remaining 2%, from other SNP sources. The markers were discovered using both anadromous and freshwater individuals originating from European and North American populations.

Quality control filters and marker positioning

Markers identified as multisite variants, paralogous sequence variants and failed assays were removed following the quality control procedure described by Bourret *et al.* (2013b). POWERMAKER (Liu & Muse 2005) and R (R core development team) were used to filter data as follows: (i) heterozygote frequencies >50% were excluded; (ii) loci with call rates <95% were excluded (i.e. 95% of the individuals were successfully genotyped for each locus); (iii) individuals with call rates <90% were excluded (i.e. 90% of the SNPs successfully amplified in individual sample); (iv) SNPs were additionally filtered for a minor allele frequency of 0.05 over all populations or per population. For subsequent analyses on subsets of data (e.g. pairwise F_{ST} , genome scans and sliding window analyses), we filtered by a global minor allele frequency of 0.05. The extent of linkage disequilibrium between markers was estimated using POWERMAKER through the measure of SNPs' pairwise D' . The polymorphic SNPs were positioned on the Atlantic salmon linkage map (Brenna-Hansen *et al.* 2012). The map comprised 27 linkage groups with length ranging from 41.6 to 124.0 cM, for a total map length of 2179 cM.

Overall, 2336 SNPs were retained after filtering to estimate the genetic diversity among samples. SNPs were widely distributed across the linkage map, with 32 to 167 markers per linkage group, 86 in average (Fig. 2), resulting in an average resolution of 0.90 cM between loci. SNPs therefore exhibited low linkage disequilibrium (Fig. S1, Supporting information).

Measure of genetic diversity

We measured expected heterozygosity within each population using ARLEQUIN 3.5 (Excoffier & Lischer 2010). Data normality was confirmed using Shapiro–Wilk normality tests, and Student's *t*-test was used to compare expected heterozygosity among freshwater and anadromous populations. Average expected heterozygosity was reported for each linkage group and population. For the three pairs of anadromous and freshwater populations, we also described the distribution of genetic diversity as the proportions of shared polymorphism (both alleles found within both

environments), parallel fixation (a single identical allele found within both environments), antagonistic fixation (a single and different alleles found in each environment) and polymorphism exclusive to one of the two environments (SNPs with two alleles in freshwater populations, but a single allele in anadromous ones and *vice versa*).

Measures of divergence among populations

To provide an overview of population structure, we measured pairwise D_A genetic distance (Nei *et al.* 1983) using SNPs that were found to be neutral among freshwater and anadromous populations or among distant freshwater and anadromous populations (procedure to detect outliers detailed in the next section). The resulting genetic distance matrix was used to construct a neighbour-joining (NJ) phylogram using MEGA 4.0.1 (Tamura *et al.* 2007). Confidence of tree topology was obtained by resampling over loci with 1000 bootstrap replicates.

The program Structure (Pritchard *et al.* 2000) was used to delineate genetic clustering of individuals and to identify putative migrants. This analysis was performed using an admixture model and a number of genetic clusters (k) from one to 10 and 10 replicates for each k . The Structure analysis was restricted to SNPs that were found to be neutral based on genome scan analysis. Each run started with a burn-in period of 50 000 steps followed by 300 000 Markov Chain Monte Carlo (MCMC) replicates. To determine the best number of clusters, we inspected likelihood values computed by Structure and used the delta k statistic (Evanno *et al.* 2005). We ran Structure hierarchically, first on all the populations and then only on the four populations from the Lake St Jean (FRE2 and FRE3). For this second analysis, we only included SNPs that did not show evidence of divergent selection among FRE2 and FRE3 or within FRE3.

To estimate the extent of genetic differentiation (F_{ST} , Weir & Cockerham 1984), we used ARLEQUIN 3.5 (Excoffier *et al.* 2009).

Identification of outliers SNPs and genomic regions

To identify SNPs potentially under divergent selection, we applied both the Fdist and hierarchical Fdist (Excoffier *et al.* 2009) genome scan methods implemented in ARLEQUIN 3.5. In hierarchical Fdist, migration rates among groups are different than migration rates among populations within groups, allowing for the detection of outlier loci in systems where a hierarchical population structure is expected. To compare the extent of divergence among populations at multiple spatial

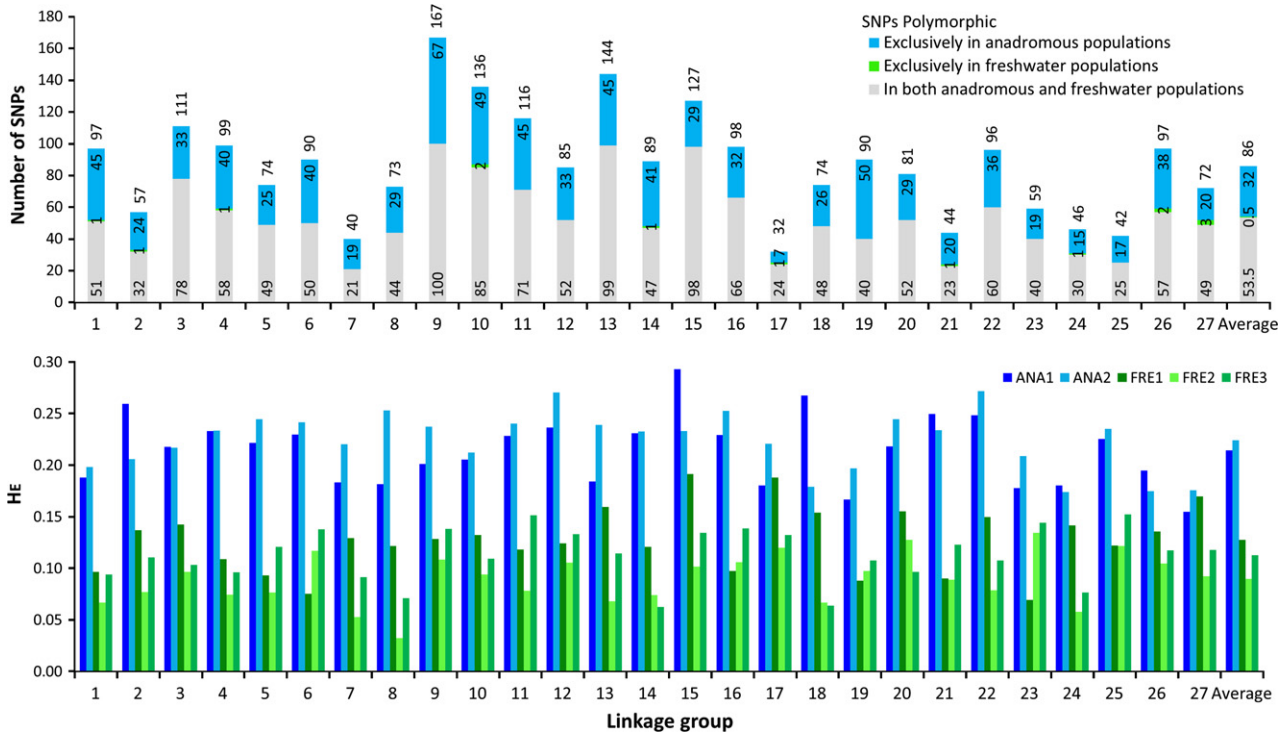


Fig. 2 Upper panel: number of SNPs polymorphic exclusively in anadromous populations, exclusively in freshwater populations or in both types of populations. Lower panel: average expected heterozygosity at each linkage group for the five groups of anadromous (blue bars) and freshwater (green bars) Atlantic salmon populations considered.

scales, we computed the following comparisons: among proximate anadromous populations (F_{ST} within ANA2), among proximate freshwater populations (F_{ST} within FRE3 and F_{CT} between FRE2 and FRE3), among distant anadromous populations (F_{CT} between ANA2 and ANA1), among distant freshwater populations (F_{CT} between FRE1, FRE2 and FRE3). Then, we estimated divergence among anadromous and freshwater populations using either the hierarchical method or the nonhierarchical method. To perform hierarchical tests, we first estimated F_{ST} between ANA1 and FRE1, F_{CT} between ANA2 and FRE2 and F_{CT} between ANA2 and FRE3. Nonhierarchical comparisons were achieved by estimating F_{STs} between all anadromous and freshwater populations for the following pairs: ANA1–FRE1, each of the ANA2–FRE2 pairs and each of the ANA2–FRE3 pairs. We used a 0.05 significance level to identify outlier SNP markers. SNPs potentially under divergent selection in the several freshwater/anadromous comparisons with a p-value less than 0.05 were considered as parallel outliers potentially under divergent selection in freshwater and anadromous environments. F_{ST} and F_{CT} values as well as outlier loci were positioned on the Atlantic salmon linkage map to observe the spatial distribution of divergence and to investigate any potential clustering of outliers.

As an alternative to single-locus outlier tests, we used a kernel-smoothing moving average approach (similarly to Hohenlohe *et al.* 2010; Gagnaire *et al.* 2013) to document relatively large genomic regions potentially influenced by divergent selection and to generate genome-wide distributions of the divergence estimates (measured as F_{ST} and F_{CT}) across LGs. The window length was set to 1 cM. Markers included within 3 cM regions covering both sides of the window were also considered in the average. We performed 10 000 permutations to estimate local P -values. This analysis allows for the identification of genomic regions with a higher proportion of SNPs showing elevated or decreased divergence, which are suggestive of divergent or balancing selection, and is complementary to the single-locus genome scan previously described.

Gene ontology and SNP annotation

Blast2go (Götz *et al.* 2008) was used to associate gene ontology (GO) annotation terms with SNPs. A homology search was first performed through a BLAST (Altschul *et al.* 1990) search of the available flanking sequences for each SNP on the NCBI public database with the e -value threshold set to 1×10^{-10} . Blast2go

then retrieved GO terms associated with the obtained BLAST hits. The output GO annotations were then classified by biological processes, molecular functions and cellular components for the most general (level two) terms of each category. To identify signatures of divergent selection on key biological processes or functions, we determined whether any biological processes, molecular functions or cellular components were over-, equally or under-represented among the outliers herein identified when compared to all retained SNPs. This was done by means of the Fisher's exact test corrected for multiple tests by applying a false discovery rate (FDR) of 0.05 (Benjamini & Yekutieli 2001).

Results

Genome-wide diversity within populations

Overall, freshwater populations exhibited lower proportions of private polymorphism compared with anadromous populations and also had a lower expected heterozygosity. On average, the number of polymorphic SNPs was reduced by more than twofold (from 1.68 to 2.56 depending on the population pairs) in freshwater populations compared with anadromous populations (Table 1, Fisher's exact test, P -values <0.0001). Over the entire data set, a much larger number of SNPs were exclusively polymorphic within anadromous populations ($n = 873$; 37.4% of the polymorphic SNPs) than in freshwater ones ($n = 14$; 0.6%) (Table 2, Fisher's exact test, P -values <0.0001). This pattern was consistent across all LGs, with 22–56% of the SNPs per LG being polymorphic in anadromous populations only, compared with 0–3% of the SNPs being exclusively polymorphic within freshwater populations (Fig. 2; Fisher's exact tests: P -values <0.05). Moreover, across the three pairs of anadromous and freshwater populations, no SNPs were found to be exclusively polymorphic in all freshwater populations, but 526 (51.3%) were repeatedly

found to be exclusively polymorphic in anadromous populations (Table 2). Finally, no SNPs were differentially fixed among anadromous and freshwater populations.

Accordingly, freshwater populations showed on average, and for each comparison, a much lower expected heterozygosity (H_E) than anadromous populations (Table 1; t -tests: all P -values <0.0001). Moreover, while H_E was not different among anadromous populations (t -tests: all P -values >0.1), H_E was lower in FRE3 than in FRE1 and lower in FRE2 than in FRE3 (t -test, P -val <0.05). At almost of the LGs and for almost all population pairs, H_E was lower in freshwater than in anadromous populations. Depending on the LG, we found 1.39–3.25 times lower H_E in freshwater than in anadromous groups of populations (Fig. 2; 75 of 81 comparisons were significant, t -tests: P -values <0.05). While we found a high correlation between SNPs H_E reduction in freshwater compared with anadromous populations in geographically close population pairs (Fig. 3, $r^2 = 0.543$, $P < 0.001$), no correlation was observed for distant pairs ($r^2 = 0.002$, $P > 0.10$ for both comparisons). A similar pattern was found when comparing average H_E differences per LG ($r^2 = 0.62$, $P < 0.001$ for close populations and $r^2 = 0.030$ and 0.07 , $P > 0.10$ for distant ones).

Genome-wide differentiation among populations

The neighbour-joining tree and clustering analysis (Fig. 4) illustrate the large divergence among freshwater populations, but a relatively small divergence among anadromous populations. Furthermore, structure analysis identified two putative migrants (q values of, respectively, 1.00 and 0.98 with confidence intervals <0.05 around estimates) from a freshwater (FRE1) to an anadromous population (ANA1), and no migrants from anadromous into freshwater populations were detected (Fig. 4b). Finally, pronounced differentiation was

Table 2 Distribution of polymorphism between anadromous and freshwater populations: Number and proportion of SNPs for which the polymorphism was found (i) in both anadromous and freshwater populations (shared polymorphism); (ii) in anadromous populations only (exclusive to anadromous); and (iii) in freshwater populations only (exclusive to freshwater). The numbers and proportions are given for the overall populations, for ANA1–FRE1, for ANA2–(FRE2–FRE3) and in parallel for these two groups of populations

	ANA–FRE		ANA1–FRE1		ANA2–(FRE2–FRE3)		Parallel	
	<i>N</i>	Proportion*	<i>N</i>	Proportion*	<i>N</i>	Proportion*	<i>N</i>	Proportion*
(i) Shared polymorphism	1449	0.620	873	0.466	1022	0.452	499	0.487
(ii) Exclusive to anadromous	873	0.374	964	0.514	1222	0.540	526	0.513
(iii) Exclusive to freshwater	14	0.006	37	0.020	19	0.008	0	0.000

*Proportion relative to the number of polymorphic SNPs in the considered comparison.

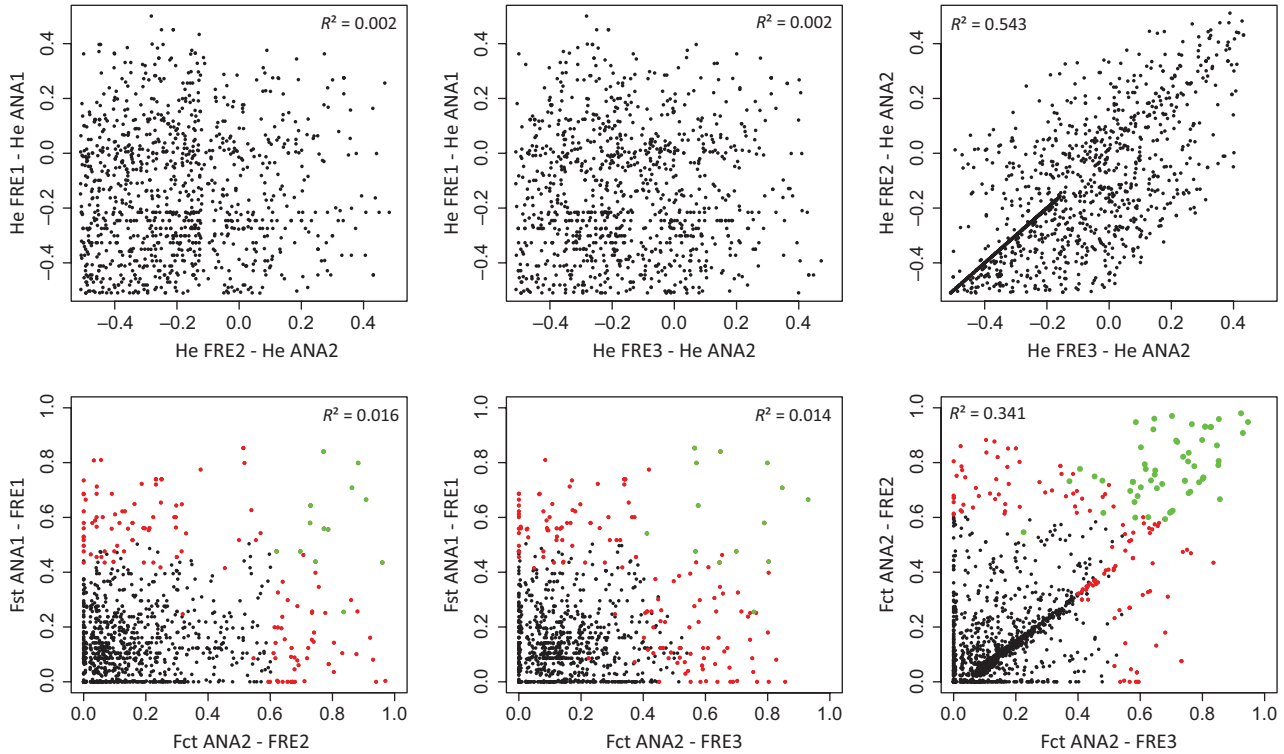


Fig. 3 Upper panel: correlation among the several pairs of populations in the reduction of expected heterozygosity in freshwater compared with anadromous populations. Lower panel: correlation between SNPs F_{ST} and F_{CT} among freshwater and anadromous populations for the three comparisons. Red dots represent SNPs found outliers for one of the two comparisons, and green dots correspond to SNPs found outliers for both comparisons.

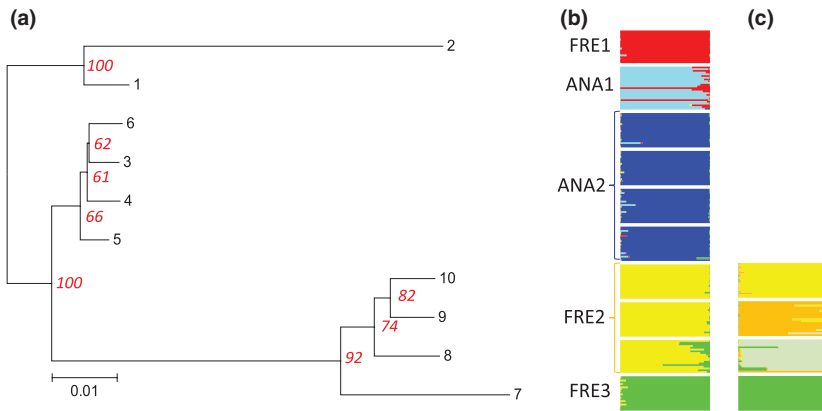


Fig. 4 (a) Neighbour-joining tree on Nei's genetic distances among the 10 Atlantic salmon populations based on 1817 putatively neutral markers (see Methods) with bootstrap values based on 1000 replicates, (b) Bayesian individual assignment using Structure for the 10 populations and 1817 putatively neutral markers and (c) for the 4 populations from the Lake St Jean and 832 putatively neutral markers.

observed between FRE2 and FRE3 although they are geographically close (Fig. 4).

There was a gradual increase in genetic differentiation among the several pairs of populations compared (Fig. 5), with a mean F_{ST} of 0.013 among proximate anadromous populations (among ANA2 populations), a F_{ST} of 0.090 among freshwater populations comprised in FRE3, a F_{CT} of 0.097 among distant anadromous populations (ANA1-ANA2), a F_{CT} of 0.144 among neighbouring freshwater populations (FRE2-FRE3) and

a F_{CT} of 0.340 between distant freshwater populations (FRE1-FRE2-FRE3). Genome-wide genetic differentiation among anadromous and freshwater populations was globally similar for the three comparisons, with mean F_{ST} of 0.174 (ANA1-FRE1), 0.173 (ANA2-FRE2) and 0.186 (ANA2-FRE3). No SNPs showed the maximal F_{ST} value of 1 among anadromous and freshwater populations, the largest values being 0.85, 0.98 and 0.95 in comparisons ANA1-FRE1, ANA2-FRE2 and ANA2-FRE3, respectively. SNP differentiation among

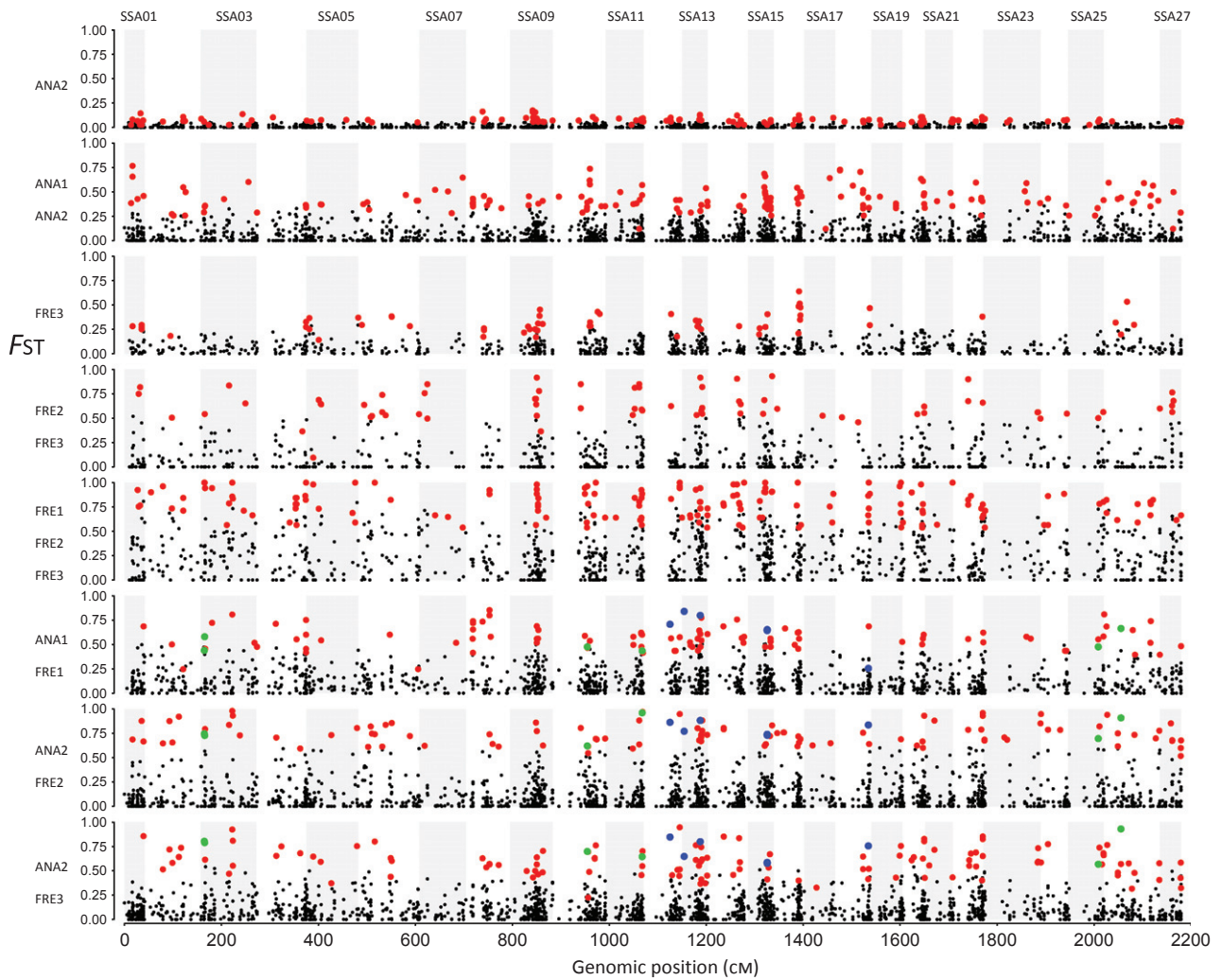


Fig. 5 Genome-wide pattern of divergence among freshwater and anadromous populations and for the three pairs of anadromous and freshwater populations, and red circles indicate outlier SNPs potentially under divergent selection ($P < 0.05$) for each comparison. Green and blue circles indicate, respectively, parallel divergent outliers and true parallel divergent outliers ($P > 0.05$) among the three anadromous and freshwater populations ($P > 0.05$).

freshwater and anadromous population pairs was correlated among geographically close populations, but not between distant ones (Fig. 3). Similarly, while we found a significant correlation in the average divergence per LG between close anadromous and freshwater populations pairs ($r^2 = 0.26$, $P < 0.01$), the correlation was not significant among distant pairs of populations ($P > 0.10$).

Detection of outlier SNPs and outlier genomic regions

The several genome scans we performed revealed weak parallelism in outlier SNPs from different population pairs. Among the five comparisons involving either anadromous or freshwater populations at small or large spatial scales (F_{ST} ANA2, F_{CT} ANA1–ANA2, F_{ST} FRE3,

F_{CT} FRE2–FRE3, F_{CT} FRE1–FRE2–FRE3), genome scans detected 69 to 178 outliers potentially under divergent selection (Table 3), but none were found across all comparisons. There was a trend for a higher proportion of outliers in comparisons involving freshwater compared with anadromous populations (Table 3). There was also a trend for a higher proportion of outliers in comparisons involving distant pairs compared with proximate populations.

Hierarchical genome scans performed among pairs of freshwater and anadromous populations detected a total of 303 SNPs potentially under divergent selection, that is, 117, 116 and 145 per genome scan between ANA1–FRE1, ANA2–FRE2 and ANA2–FRE3, respectively (Table 3). Among these SNPs potentially under divergent selection, 60, 37 and 39 SNPs were monomorphic in

Table 3 Genome scans summary. Number, proportion and divergence among populations (F_{ST}) are reported for the overall SNPs and for the SNPs classified as potentially under divergent selection for the several comparisons

			Polymorphic SNPs (MAF ≥ 0.05)		Potentially under divergent selection $P < 0.05$			
			N	F_{ST}	N	Proportion*	F_{ST}	
Similar migratory strategy	Regional grouping (AMOVAs)	F_{ST} ANA2	1730	0.013	104	0.060	0.082	
		F_{CT} ANA1–ANA2	1791	0.097	153	0.085	0.445	
		F_{ST} FRE3	811	0.090	69	0.085	0.317	
		F_{CT} FRE2–FRE3	824	0.144	75	0.091	0.638	
		F_{CT} FRE1–FRE2–FRE3	1103	0.340	178	0.161	0.782	
Anadromous vs. freshwater	Regional grouping (AMOVAs)	F_{ST} ANA1–FRE1	1378	0.174	117	0.085	0.580	
		F_{CT} ANA2–FRE2	1695	0.173	116	0.068	0.747	
		F_{CT} ANA2–FRE3	1529	0.186	145	0.095	0.604	
		Parallel ANA1–FRE1/ANA2–FRE2	1094	0.182	14	0.013	0.682	
		Parallel ANA1–FRE1/ANA2–FRE3	1032	0.180	16	0.016	0.649	
		Parallel ANA2–FRE2/ANA2–FRE3	1489	0.187	58	0.039	0.736	
		Total	2002	0.178	303	0.151	0.622	
		Parallel	1015	0.187	12 (5**)	0.012	0.696	
		Pairwise tests	F_{ST} 1–2	1378	0.174	117	0.085	0.580
			F_{ST} 3–7	1474	0.238	251	0.170	0.612
	F_{ST} 4–7		1493	0.227	230	0.154	0.617	
	F_{ST} 5–7		1493	0.226	204	0.137	0.637	
	F_{ST} 6–7		1497	0.223	223	0.149	0.608	
	F_{ST} 3–8		1509	0.210	180	0.119	0.608	
	F_{ST} 4–8		1537	0.204	186	0.121	0.586	
	F_{ST} 5–8		1533	0.198	180	0.117	0.581	
	F_{ST} 6–8		1529	0.200	159	0.104	0.607	
	F_{ST} 3–9		1494	0.211	224	0.150	0.584	
	F_{ST} 4–9	1521	0.207	206	0.135	0.575		
	F_{ST} 5–9	1526	0.228	186	0.122	0.608		
F_{ST} 6–9	1522	0.231	185	0.122	0.620			
F_{ST} 3–10	1504	0.210	189	0.126	0.608			
F_{ST} 4–10	1539	0.202	199	0.129	0.579			
F_{ST} 5–10	1526	0.225	188	0.123	0.587			
F_{ST} 6–10	1525	0.227	208	0.136	0.565			
Total	1990	0.175	688	0.346	0.602			
Parallel	730	0.241	4 (1**)	0.005	0.796			

*Proportion relative to polymorphic SNPs retained for each comparison.

**True parallel (not divergent between distant freshwater and anadromous populations).

freshwater, compared with 4, 4 and 5 SNPs monomorphic in anadromous populations, respectively. Average genetic differentiation ranged from 0.580 to 0.747 among these divergent outliers for the three comparisons. For most of the outlier SNPs, the allele found in large frequency in freshwater was also normally found in anadromous populations at relatively low frequencies (Fig. S2, Supporting information). These divergent outliers were widely distributed throughout the genome without clear clustering on specific linkage groups or obvious parallelism among population comparisons (Fig. 5 & Table 4). The number of SNPs repeatedly found among comparisons was lower for distant pairs (from 14 to 16 SNPs) than for proximate pairs of populations (58

SNPs). Only 12 SNPs were found to be potentially under divergent selection among the three comparisons with an average F_{ST} of 0.696 (allele frequencies depicted in Fig. S3, Supporting information). These 12 outliers were monomorphic in at least one of the three comparisons between anadromous and freshwater populations. These 12 SNPs were located on nine different LGs (Fig. 5 & Table 4). However, seven of these 12 SNPs were also detected as outliers among anadromous populations or among freshwater populations and were consequently not considered as parallel outliers. The remaining five most likely targets of parallel divergent selection were distributed on four LGs: SSA10, SSA12, SSA13 and SSA18 (Fig. 5 & Table 4).

Table 4 Distribution on linkage groups of the outlier loci in freshwater–anadromous genome scans and of the significantly divergent genomic regions

LG	Outlier SNPs										Outlier genomic regions													
	ANA1-FRE1		ANA2-FRE2		ANA2-FRE3		Total		Repeated		True parallel		ANA1-FRE1		ANA2-FRE2		ANA2-FRE3		Total		Repeated		True parallel	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
LG1	1	2	3	3	1	1	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LG2	2	4	4	8	5	10	8	14	0	0	0	0	0	2	2	2	2	0	0	2	2	0	0	0
LG3	7	9	9	9	9	9	16	15	2	0	0	0	2	2	2	4	5	0	0	4	5	0	0	0
LG4	7	11	2	2	3	3	10	10	0	0	0	0	0	0	1	1	1	0	0	1	1	0	0	0
LG5	1	2	2	3	4	5	5	7	0	0	0	0	0	1	2	3	0	0	2	3	0	0	0	0
LG6	4	7	7	9	4	5	14	16	0	0	0	0	1	4	2	4	0	0	2	4	0	0	0	0
LG7	1	5	1	3	0	0	2	5	0	0	0	0	1	1	0	0	2	0	0	2	0	0	0	0
LG8	15	29	3	4	6	8	22	29	0	0	0	2	1	1	3	3	1	0	3	3	1	0	0	0
LG9	5	4	4	3	10	6	16	10	0	0	0	1	0	0	3	3	0	0	3	3	0	0	0	0
LG10	4	4	7	6	7	6	13	10	1	1	1	0	0	0	2	2	0	0	2	2	0	0	0	0
LG11	8	10	5	5	4	4	14	12	1	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0
LG12	5	8	2	2	7	9	11	13	1	1	1	1	0	0	3	3	0	0	3	3	0	0	0	0
LG13	14	14	12	9	15	11	30	21	2	2	2	2	2	1	3	4	0	0	3	4	0	0	0	0
LG14	6	11	2	2	7	8	13	15	0	0	0	0	2	1	2	2	3	1	2	3	1	0	0	0
LG15	6	6	6	5	7	6	14	11	2	0	0	0	0	0	0	2	2	0	2	2	0	0	0	0
LG16	6	9	7	8	1	1	14	15	0	0	0	0	1	1	1	1	3	0	1	3	0	0	0	0
LG17	0	0	2	7	1	3	3	9	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0
LG18	2	3	3	5	5	9	7	9	1	1	1	1	0	0	2	2	0	0	2	2	0	0	0	0
LG19	1	2	1	1	3	4	4	4	0	0	0	0	0	0	0	1	1	0	1	1	0	0	0	0
LG20	3	5	4	5	6	8	11	14	0	0	0	0	1	1	3	3	0	0	3	3	0	0	0	0
LG21	0	0	1	2	3	7	3	7	0	0	0	0	0	1	1	1	0	0	1	1	0	0	0	0
LG22	1	1	5	5	9	10	12	13	0	0	0	0	0	1	3	3	0	0	3	3	0	0	0	0
LG23	4	12	5	10	4	8	13	22	0	0	0	0	2	2	5	5	0	0	2	5	0	0	0	0
LG24	2	7	3	7	2	5	5	11	0	0	0	0	0	2	1	1	0	0	2	1	0	0	0	0
LG25	3	10	2	5	7	18	9	21	1	0	0	0	1	1	2	2	0	0	2	2	0	0	0	0
LG26	7	11	7	8	12	13	19	20	1	0	0	0	2	2	4	5	1	0	4	5	1	0	0	0
LG27	2	4	7	11	3	5	11	15	0	0	0	0	0	1	1	2	0	0	1	2	0	0	0	0
Total	117		116		145		303		12	5	17	25	51	67	3	67	3	0	67	3	0	0	0	0
Average	4.33	0.07	4.30	0.05	5.37	0.07	11.22	0.13	0.44	0.19	0.63	0.93	1.89	2.48	0.11	2.48	0.11	2.48	0.11	2.48	0.11	0.00	0.00	0.00

Nonhierarchical genome scans among anadromous and freshwater populations yielded a comparable number of outlier SNPs and F_{ST} values and similarly indicated limited parallelism among comparisons with four of the 688 outliers being repeatedly found among the 17 comparisons (Table 3). A total of 75% of the 688 divergent outlier SNPs were detected from one to seven times as outliers over the 17 comparisons (Table S1, Supporting information). The four SNPs repeatedly identified as outliers across the 17 genome scans were also identified using the hierarchical genome scans.

The sliding window approach identified 17, 25 and 51 islands of divergence among ANA1–FRE1, ANA2–FRE2 and ANA2–FRE3, respectively (Fig. 6 & Table 4). These regions ranged from 1 to 12 cM per comparison. Only three of these regions located on SSA08 (Fig. S4, Supporting information), SSA14 and SSA26 were repeatedly found among the three pairs of anadromous and freshwater populations. However, these three regions were also found as divergent among distant freshwater populations. Among all the single-locus outliers found among anadromous and freshwater pairs of populations, 44%, 32% and 79% were found within islands of divergence in ANA1–FRE1, ANA2–FRE2 and ANA2–FRE3 comparisons. Among the five divergent outliers found to be parallel among the three pairs of anadromous and freshwater populations in genome scans, three, one and four outliers were localized within islands of divergence between ANA1–FRE1, ANA2–FRE2 and ANA2–FRE3 comparisons, respectively, remaining outliers being found outside the genomic islands of divergence identified through the sliding window analysis.

SNPs annotation

Overall, the BLAST yielded 838 SNPs (36% of all SNPs) with significant hits (e -value $< 1 \times 10^{-10}$), which were associated with a total of 9380 GO terms. An enrichment analysis did not indicate significant over- or under-representation of any biological process or function among outliers identified. Among the 12 divergent outliers common to all three freshwater-anadromous

comparisons, three were associated with retinol metabolism, oligosaccharide biosynthesis and histone ubiquitination. Retinol metabolism was the only annotation found among parallel divergent outliers.

Discussion

This study provides evidence for a large effect of genetic drift, but limited evidence for parallel divergence from standing genetic variation in freshwater Atlantic salmon populations by combining linkage map information and genotypes for 2336 SNP markers in several population pairs of anadromous and freshwater Atlantic salmon. A salient result of this study was the huge decrease in standing genetic variation and increased divergence observed among freshwater populations when compared to anadromous populations. The fact that freshwater population sizes were low and fluctuated (Fortin *et al.* 2009) may have played a central role in the decay of genetic diversity in these populations. An absence of gene flow into freshwater populations was supported by the nonexistence of putative migrants from anadromous to freshwater population, consistent with the presence of impassable waterfalls, and may also explain a large part of the divergence and diversity reduction in freshwater populations. Our results are congruent with Tessier & Bernatchez (2000) who reported significantly reduced diversity among freshwater populations of Lake St Jean using six microsatellite loci. Tonteri *et al.* (2007) also reported a similar pattern of reduced diversity at microsatellites in freshwater Atlantic salmon populations compared with anadromous populations in Northern Europe. Moreover, the extent of reduction in diversity in freshwater populations reported in this study is comparable to values reported in European landlocked populations using a similar SNP array (Bourret *et al.* 2013b). In the same way, lower levels of genetic variation have also been observed in freshwater stickleback compared with marine populations, consistent with potential founder effects and increased genetic drift during colonization of freshwater habitats (Jones *et al.* 2012; Deagle *et al.* 2013). This general pattern of loss of diversity in fresh-

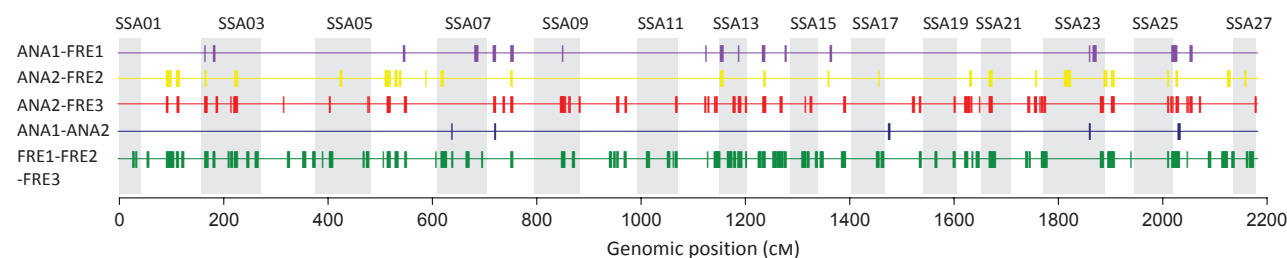


Fig. 6 Summary of the islands of divergence found using the sliding window analysis for the three anadromous vs. freshwater comparisons and for the comparisons of distant anadromous and freshwater groups of populations.

water populations was observed in interspecific comparisons of marine, anadromous and freshwater fish species (DeWoody & Avise 2000), highlighting the importance of the connectivity among populations for maintaining their genetic diversity. Finally, few alleles exclusive from freshwater were found. In addition, these polymorphisms were not found repeatedly across the three comparisons. This is congruent with the relatively recent colonization of freshwater lakes that may have left little chance for adaptive mutations to occur. From a management perspective, due to this important divergence of freshwater populations, freshwater populations should be considered as independent conservation units (Ozerov *et al.* 2010). Moreover, the relatively low diversity observed in these populations may increase the risks associated with inbreeding, cautioning for stringent angling regulations and for maximizing genetic diversity of the local population while stocking (Fraser 2008; Araki & Schmid 2010).

Interestingly, the genetic structure found among the populations inhabiting the tributaries of Lake St Jean was much higher than among the anadromous populations at similar and larger spatial scales. This result is consistent with the significant genetic structure reported for this lake by Tessier & Bernatchez (2000) as well as in North European lakes (Ozerov *et al.* 2010). In addition to the effect of genetic drift, it is likely that the divergence among proximate freshwater populations within a single lake may be magnified by low gene flow among these populations. Obviously, the nearly absence of gene flow from anadromous populations within these freshwater populations may have enhanced this divergence. Moreover, relatively high philopatry of freshwater salmon among adjacent lacustrine populations could have been enhanced by the preferential use of different lacustrine feeding areas that salmon use in these lakes (Potvin & Bernatchez 2001). Such powerful discrimination of adjacent populations from the Lake St Jean using SNPs may allow us to better investigate several aspects of the biology of these populations (e.g. mixed stocks analysis of the fish recreationally caught to document their migrations and adjust fisheries regulation) (Potvin & Bernatchez 2001; Smith & Campana 2010; Ackerman *et al.* 2011; Seeb *et al.* 2011).

Given the high genetic drift expected in freshwater populations, notably due to founding effects and fixation events, identified outliers may include numerous false positives (Narum & Hess 2011). Indeed, identifying true outlier loci and genomic regions is extremely challenging in such isolated populations with high level of neutral divergence (Le Corre & Kremer 2012; Bierne *et al.* 2013; Fourcade *et al.* 2013). Moreover, populations with low effective population size harbour fewer mutations and have a lower probability adaptive allele

fixation compared with populations with larger effective sizes (Olson-Manning *et al.* 2012). Importantly, the reduced diversity and very low occurrence of private polymorphisms in freshwater populations may imply that adaptive divergence may have occurred principally through directional selection on standing variation carried by founding individuals (Barrett & Schluter 2008). Besides, if instead of a single colonization event, two colonization events led to the settling of FRE2 and FRE3 groups of populations in the Lake St Jean (Tessier & Bernatchez 2000), one may expect relatively low parallelism in adaptive divergence and the among outliers identified between ANA2–FRE2 and ANA2–FRE3. Indeed, in the case of such double colonization, despite comparable selection regimes, the standing variation transported by founders and the subsequent genetic drift may both have been contrasted. Potential effect of drift and of adaptive divergence among anadromous populations may also be considered in this exploration of repeated adaptive divergence among freshwater and anadromous populations. Indeed, we documented a relatively high divergence among distant anadromous populations, although it was much lower than among distant freshwater populations. Furthermore, potential local adaptation has been recently documented by means of population genomics among North American populations, suggesting associations between potentially adaptive divergence and climate (Bourret *et al.* 2013a). Overall, despite hundreds of markers and dozens of genomic regions showing significant signature of selection among freshwater and anadromous populations in the genome scan and sliding window analyses, we found very limited evidence for parallel signatures of selection between the three freshwater and anadromous comparisons.

Nevertheless, the proportions of total outliers and of parallel outliers identified here were consistent with existing genome scans among pairs of European freshwater and anadromous Atlantic salmon populations (Bourret *et al.* 2013b). Similarly, searching for adaptive divergence among recently landlocked *O. mykiss* and their migratory ancestors, Hecht *et al.* (2013) detected hundreds of loci that were associated with migratory traits. From a spatial perspective across the genome, while several studies have found clear restricted genomic regions showing repeated divergence between groups of populations under different ecological pressures (Hohenlohe *et al.* 2010; Deagle *et al.* 2012; Bradbury *et al.* 2013), outliers were found here to be spread across the genome, and we did not find evidence for large parallel genomic regions showing higher divergence than expected. However, using replicates and hierarchical tests helped to filter a few parallel candidate markers implicated in parallel adaptive divergence

in freshwater. These SNPs repeatedly found to be under potentially divergent selection among anadromous and freshwater populations are the most likely to be directly or indirectly associated with targets of selection between life histories. Indeed, the demonstration of parallel evolution for a trait in repeated selective environments is often taken as strong evidence for local adaptation (Deagle *et al.* 2012, 2013; Hohenlohe *et al.* 2012; Prunier *et al.* 2012). For example, Deagle *et al.* (2013) repeatedly found several markers (e.g. SNPs in Na/K ATPase) potentially under divergent selection between numerous marine and freshwater populations, but found typically marine alleles present in a few freshwater lakes. However, the SNPs repeatedly identified here as outliers were not associated with any obvious relevant biological function in the context of ecological divergence among freshwater and anadromous Atlantic salmon populations. We suggest that the overall targets of selection we were able to highlight in this study should be further investigated when complete genomic information is made available (Lien *et al.* 2011; Brenna-Hansen *et al.* 2012; Gutierrez *et al.* 2012).

Even though many of the nonparallel outliers detected in this study may have been caused by genetic drift either in freshwater or in anadromous populations, some of them may also be indicative of a nonparallel genetic basis underlying parallel or nonparallel phenotypic divergence across multiple population pairs of freshwater and anadromous Atlantic salmon. First, numerous phenotypic traits differ between anadromous and resident Atlantic salmon, not only migration behaviour and salinity tolerance, but also growth, maturation, iteroparity rate, temperature tolerance, etc. The presence of multiple nonparallel outlier loci and regions may thus reflect different extent of directional selection among freshwater and anadromous populations for these various traits. Second, many phenotypic traits are controlled by a large numbers of loci with small effects (Atwell *et al.* 2010; Davies *et al.* 2011). Therefore, the large proportion of nonparallel outliers spread across the genome may be consistent with the fact that adaptation to anadromous and resident life cycles involves different complex traits influenced by numerous loci. Accordingly, there are an increasing number of studies finding similar nonparallelism in outlier genes among replicate population pairs showing parallel phenotypic divergence (DeFaveri *et al.* 2011; Kaeuffer *et al.* 2012; Roesti *et al.* 2012; Gagnaire *et al.* 2013). More specifically, our results are in line with those of Hecht *et al.* (2013) who found numerous nonparallel outlier loci between migratory steelhead and recently isolated resident rainbow trout, suggesting a complex genetic basis for migration. This is also consistent with the existence of multiple chromosome-wide

QTL for Na/K ATPase activity and blood plasma osmolality found over nine linkage groups in Atlantic salmon (Norman *et al.* 2012). These results are overall in line with the notion that different evolutionary pathways may often cause parallel physiological adaptation to heterogeneous habitats within the global distribution of a species.

While several markers showed high level of divergence among freshwater and anadromous populations, no differential fixation was observed for these outliers. Within freshwater populations, the presence of anadromous alleles at low frequencies may indicate that evolutionary processes initiated during the colonization of the lakes are still going on. Besides, the presence of freshwater alleles at low frequencies in anadromous populations may originate from asymmetric gene flow from freshwater into anadromous populations. Indeed, two putative migrants from Musquaro Lake to Musquaro River were have been detected in this study. Such asymmetric dispersal from freshwater to anadromous Atlantic salmon populations has also been documented in other anadromous populations from Quebec (L. Bernatchez, unpublished data) and is expected given the low probability for the fish to jump upstream over high waterfalls, but the real possibility to drift down stream. Overall, the presence of freshwater alleles at low frequencies within anadromous populations is consistent with the 'transporter' hypothesis (Schluter & Conte 2009) proposing that freshwater alleles return to the ocean where they persist at low frequency, before being selected to high frequency in newly colonized freshwater habitats. Therefore, such continuous asymmetric gene flow from freshwater to anadromous populations may result in the maintenance of putatively adaptive freshwater alleles in anadromous populations, which may become available for rapid adaptation during the colonization of new freshwater habitats. This mechanism could be reinforced by the relatively high gene flow among anadromous populations (Tonteri *et al.* 2007; Bourret *et al.* 2013a,b).

Overall, we detected large and genome-wide effects of genetic drift in freshwater populations resulting in high and mostly nonparallel genetic divergence among freshwater and anadromous populations as well as among freshwater populations of Atlantic salmon. This study also illustrates that delineating the influence of drift, parallel and nonparallel evolution remains a major challenge in the case of small isolated populations and may notably require the analysis of more independent replicates (Deagle *et al.* 2013) and confirmation of results by means of elaborated simulations (Le Corre & Kremer 2012; Bierne *et al.* 2013; Fourcade *et al.* 2013). Besides, the role of plasticity in the establishment of phenotypically differentiated resident and anadromous

populations must not be neglected and could be assessed using, for example, reciprocal transplants (McCairns & Bernatchez 2012). Lastly, the study of Atlantic salmon sympatric freshwater and anadromous populations that coexist in some rivers (Power *et al.* 1987; Verspoor & Cole 1989; Fleming 1996) may also allow performing powerful genotype–phenotype association tests to detect polymorphisms implicated in the divergence and maintenance of these resident and anadromous forms. In conclusion, this study emphasizes the important role of genetic drift in driving genome-wide reduction in diversity and divergence in freshwater Atlantic salmon populations. Nevertheless, the few SNPs and genomic regions repeatedly identified as potential divergent outliers may represent key targets of divergent selection. Furthermore, while nonparallel outlier SNPs and genomic regions may be often associated with the effect of drift, they may also be implicated in divergent evolution between freshwater and anadromous populations, in line with the fact that adaptation to such divergent life cycles might be influenced by numerous loci. This study might therefore suggest a complex polygenic basis of adaptation to resident and anadromous strategies, as would be predicted by quantitative genetics theory (Hohenlohe *et al.* 2012; Le Corre & Kremer 2012).

Acknowledgements

We are grateful to E. Normandeau, A.M. Dion-Coté, P.A. Gagnaire, S. Pavey, F. Lamaze and A.C. Dalziel for their constructive inputs in analysing and interpreting the results. We also thank Associate Editor A. Bonin and six anonymous referees for helpful comments and suggestions. The SNP genotyping was performed by CIGENE at the national technology platform, supported by the functional genomics programme (FUGE) in the Research Council of Norway. This research was supported by a strategic grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) led by LB. CP was supported by a postdoctoral fellowship from Réseau Aquaculture Québec (RAQ), VB was supported by an Alexander Graham Bell scholarship from NSERC, and LB. is the Canadian Research Chair in genomics and conservation of aquatic resources. This work is a contribution to the research programme of CIRSA (Centre Interuniversitaire de Recherche sur le Saumon Atlantique).

References

- Ackerman MW, Habicht C, Seeb LW (2011) Single-nucleotide polymorphisms (SNPs) under diversifying selection provide increased accuracy and precision in mixed-stock analyses of sockeye salmon from the Copper River, Alaska. *Transactions of the American Fisheries Society*, **140**, 865–881.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.
- Araki H, Schmid C (2010) Is hatchery stocking a help or harm? Evidence, limitations and future directions in ecological and genetic surveys. *Aquaculture*, **308**, S2–S11.
- Atwell S, Huang YS, Vilhjalmsson BJ *et al.* (2010) Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature*, **465**, 627–631.
- Barrett RDH, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution*, **23**, 38–44.
- Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of Statistics*, **29**, 1165–1188.
- Berg OK (1985) The formation of non-anadromous populations of Atlantic salmon, *Salmo salar* L, in Europe. *Journal of Fish Biology*, **27**, 805–815.
- Bernatchez L (1997) Mitochondrial DNA analysis confirms the existence of two glacial races of rainbow smelt *Osmerus mordax* and their reproductive isolation in the St Lawrence River estuary (Quebec, Canada). *Molecular Ecology*, **6**, 73–83.
- Bierne N, Roze D, Welch JJ (2013) Pervasive selection or is it...? Why are Fst outliers sometimes so frequent? *Molecular Ecology*, **22**, 2061–2064.
- Boulet M, Normandeau E, Bougas B, Audet C, Bernatchez L (2012) Comparative Transcriptomics of Anadromous and Resident Brook Charr (*Salvelinus fontinalis*) before their First Salt Water Transition. *Current Zoology*, **58**, 155–167.
- Bourret V, Dionne M, Kent MP, Lien S, Bernatchez L (2013a) Landscape genomics in Atlantic salmon (*Salmo salar*): searching for gene-environment interactions driving local adaptation. *Evolution*, **67**. doi:10.1111/evo.12139 (Early view).
- Bourret V, Kent MP, Primmer CR *et al.* (2013b) SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Molecular Ecology*, **22**, 532–551.
- Bradbury IR, Hubert S, Higgins B *et al.* (2013) Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. *Evolutionary Applications*, **6**, 450–461.
- Brenna-Hansen S, Li J, Kent MP *et al.* (2012) Chromosomal differences between European and North American Atlantic salmon discovered by linkage mapping and supported by fluorescence *in situ* hybridization analysis. *BMC Genomics*, **13**, 432.
- Burton MP, Idler DR (1984) Can Newfoundland landlocked salmon, *Salmo-salar* L, adapt to sea-water. *Journal of Fish Biology*, **24**, 59–64.
- Colbert J, Danchin E, Dhondt AA, Nichols JD (2001) *Dispersal*. Oxford University Press, New York.
- Davies G, Tenesa A, Payton A *et al.* (2011) Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry*, **16**, 996–1005.
- Deagle BE, Jones FC, Chan YF, Absher DM, Kingsley DM, Reimchen TE (2012) Population genomics of parallel phenotypic evolution in stickleback across stream-lake ecological transitions. *Proceedings. Biological Sciences/The Royal Society*, **279**, 1277–1286.
- Deagle BE, Jones FC, Absher DM, Kingsley DM, Reimchen TE (2013) Phylogeography and adaptation genetics of stickleback from the Haida Gwaii archipelago revealed using genome-wide single nucleotide polymorphism genotyping. *Molecular Ecology*, **22**, 1917–1932.

- DeFaveri J, Shikano T, Shimada Y, Goto A, Merilä J (2011) Global analysis of genes involved in freshwater adaptation in threespine sticklebacks (*Gasterosteus aculeatus*). *Evolution*, **65**, 1800–1807.
- DeWoody JA, Avise JC (2000) Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *Journal of Fish Biology*, **56**, 461–473.
- Dionne M, Caron F, Dodson JJ, Bernatchez L (2008) Landscape genetics and hierarchical genetic structure in Atlantic salmon: the interaction of gene flow and local adaptation. *Molecular Ecology*, **17**, 2382–2396.
- Dittman AH, Quinn TP (1996) Homing in pacific salmon: mechanisms and ecological basis. *Journal of Experimental Biology*, **199**, 83–91.
- Duston J, Astatkie T, Zhang C (2011) Hypo-osmoregulatory capacity during smolting of endangered inner bay of Fundy Atlantic salmon and other eastern Canadian stocks. *Aquaculture*, **319**, 221–225.
- 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.
- Fleming IA (1996) Reproductive strategies of atlantic salmon: ecology and evolution. *Reviews in Fish Biology and Fisheries*, **6**, 379–416.
- Fleming IA, Reynolds JD (2004) Salmonid breeding systems. In: *Evolution Illuminated: Salmon and their Relatives* (eds Hendry AP & Stearns SC), pp. 264–294. Oxford University Press, New York.
- Folmar LC, Dickhoff WW (1980) The parr-smolt transformation (smoltification) and seawater adaptation in salmonids - a review of selected literature. *Aquaculture*, **21**, 1–37.
- Fortin AL, Sirois P, Legault M (2009) Synthèse et analyse des connaissances sur la ouananiche et l'éperlan arc-en-ciel du lac Saint-Jean, p. 137. Université du Québec à Chicoutimi, Laboratoire des sciences aquatiques et Ministère des Ressources naturelles et de la Faune du Québec, Direction de l'expertise sur la faune et ses habitats.
- Fourcade Y, Chaput-Bardy A, Secondi J, Fleurant C, Lemaire C (2013) Is local selection so widespread in river organisms? Fractal geometry of river networks leads to high bias in outlier detection. *Molecular Ecology*, **22**, 2065–2073.
- Fraser DJ (2008) How well can captive breeding programs conserve biodiversity? A review of salmonids. *Evolutionary Applications*, **1**, 535–586.
- Fraser DJ, Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology*, **10**, 2741–2752.
- Fraser DJ, Weir LK, Bernatchez L, Hansen MM, Taylor EB (2011) Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. *Heredity*, **106**, 404–420.
- Gagnaire P-A, Pavey SA, Normandeau E, Bernatchez L (2013) The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by rad sequencing. *Evolution*, **67**, 2483–2497.
- Garcia de Leaniz C, Fleming IA, Einum S *et al.* (2007) A critical review of adaptive genetic variation in Atlantic salmon: implications for conservation. *Biological Reviews*, **82**, 173–211.
- Götz S, García-Gómez JM, Terol J *et al.* (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, **36**, 3420–3435.
- Gutierrez AP, Lubieniecki KP, Davidson EA *et al.* (2012) Genetic mapping of quantitative trait loci (QTL) for body-weight in Atlantic salmon (*Salmo salar*) using a 6.5 k SNP array. *Aquaculture*, **358**, 61–70.
- Hale MC, Thrower FP, Berntson EA, Miller MR, Nichols KM (2013) Evaluating adaptive divergence between migratory and nonmigratory ecotypes of a salmonid fish, *Oncorhynchus mykiss*. *G3: Genes Genomes Genetics*, **3**, 1273–1285.
- Hansen LP, Jonsson N, Jonsson B (1993) Oceanic migration in homing Atlantic salmon. *Animal Behaviour*, **45**, 927–941.
- Hara T, Nagase T, Ozaki A *et al.* (2007) A genetic linkage map of amago salmon (*Oncorhynchus masou ishikawae*) and mapping of quantitative trait loci associated with smoltification. *Aquaculture*, **272**, S266–S267.
- Hecht BC, Campbell NR, Holecek DE, Narum SR (2013) Genome-wide association reveals genetic basis for the propensity to migrate in wild populations of rainbow and steelhead trout. *Molecular Ecology*, **22**, 3061–3076.
- Hendry AP, Letcher BH, Gries G (2003) Estimating natural selection acting on stream-dwelling atlantic salmon: implications for the restoration of extirpated populations. *Conservation Biology*, **17**, 795–805.
- Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced rad tags. *PLoS Genetics*, **6**, 1–23.
- Hohenlohe PA, Bassham S, Currey M, Cresko WA (2012) Extensive linkage disequilibrium and parallel adaptive divergence across threespine stickleback genomes. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, **367**, 395–408.
- Hubert S, Le Provost O, O'Keeffe AM, Voisin M, Cotter D, Cairns MT (2007) Gene expression patterns in Atlantic salmon (*Salmo salar*): gene expression in intestine tissue during smoltification. *Aquaculture*, **272**, S270.
- Jones FC, Chan Yingguang F, Schmutz J *et al.* (2012) A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Current Biology*, **22**, 83–90.
- Jonsson B, Jonsson N (2011) *Migrations Ecology of Atlantic Salmon and Brown Trout*, pp. 247–325. Springer, Netherlands.
- Kaeuffer R, Peichel CL, Bolnick DI, Hendry AP (2012) Parallel and nonparallel aspects of ecological, phenotypic, and genetic divergence across replicate population pairs of lake and stream stickleback. *Evolution*, **66**, 402–418.
- Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**, 1225–1241.
- Kazakov RV, Veselov AJ (1998) Atlantic salmon (*Salmo salar* L.) catches in Russia. *Journal of Applied Ichthyology*, **14**, 65–68.
- King TL, Verspoor E, Spidle AP *et al.* (2007) Biodiversity and population structure. In: *The Atlantic Salmon: Genetics, Conservation and Management* (eds E Verspoor, L Stradmeyer and J Nielsen), pp. 117–166. Blackwell Publishing Ltd, Oxford, UK.

- Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, **21**, 1548–1566.
- Liedvogel M, Åkesson S, Bensch S (2011) The genetics of migration on the move. *Trends in Ecology & Evolution*, **26**, 561–569.
- Lien S, Gidskehaug L, Moen T *et al.* (2011) A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. *BMC Genomics*, **12**, 615.
- Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, **21**, 2128–2129.
- McCairns RJS, Bernatchez L (2012) Plasticity and heritability of morphological variation within and between parapatric stickleback demes. *Journal of Evolutionary Biology*, **25**, 1097–1112.
- Narum SR, Hess JE (2011) Comparison of Fst outlier tests for SNP loci under selection. *Molecular Ecology Resources*, **11**, 184–194.
- Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular-data.2. Gene-frequency data. *Journal of Molecular Evolution*, **19**, 153–170.
- Nichols KM, Edo AF, Wheeler PA, Thorgaard GH (2008) The genetic basis of smoltification-related traits in *O. mykiss*. *Genetics*, **179**, 1559–1575.
- Nielsen EE, Hemmer-Hansen J, Poulsen NA *et al.* (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology*, **9**, 276.
- Nilsen TO, Ebbesson LOE, Stefansson SO (2003) Smolting in anadromous and landlocked strains of Atlantic salmon (*Salmo salar*). *Aquaculture*, **222**, 71–82.
- Nilsen TO, Ebbesson LOE, Madsen SS *et al.* (2007) Differential expression of gill Na⁺, K⁺ -ATPase alpha- and beta-subunits, Na⁺, k⁺, 2Cl⁻-cotransporter and CFTR anion channel in juvenile anadromous and landlocked Atlantic salmon *Salmo salar*. *Journal of Experimental Biology*, **210**, 2885–2896.
- Norman JD, Robinson M, Glebe B, Ferguson MM, Danzmann RG (2012) Genomic arrangement of salinity tolerance QTLs in salmonids: a comparative analysis of Atlantic salmon (*Salmo salar*) with arctic charr (*Salvelinus alpinus*) and rainbow trout (*O. mykiss*). *BMC Genomics*, **13**, 420.
- Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.
- Olson-Manning CF, Wagner MR, Mitchell-Olds T (2012) Adaptive evolution: evaluating empirical support for theoretical predictions. *Nature Reviews Genetics*, **13**, 867–877.
- Ozerov M, Veselov A, Lumme J, Primmer C (2010) Genetic structure of freshwater Atlantic salmon (*Salmo salar*; L.) populations from the lakes Onega and Ladoga of northwest Russia and implications for conservation. *Conservation Genetics*, **11**, 1711–1724.
- Potvin C, Bernatchez L (2001) Lacustrine spatial distribution of landlocked Atlantic salmon populations assessed across generations by multilocus individual assignment and mixed-stock analyses. *Molecular Ecology*, **10**, 2375–2388.
- Power G, Power MV, Dumas R, Gordon A (1987) Marine migrations of Atlantic salmon from rivers in Ungava bay, Quebec. *American Fisheries Society Symposium*, **1**, 13.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Prunier J, GÉRARDI S, Laroche J, Beaulieu J, Bousquet J (2012) Parallel and lineage-specific molecular adaptation to climate in boreal black spruce. *Molecular Ecology*, **21**, 4270–4286.
- Quinn TP (1993) A review of homing and straying of wild and hatchery-produced salmon. *Fisheries Research*, **18**, 29–44.
- Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during evolutionary diversification as revealed in replicate lake–stream stickleback population pairs. *Molecular Ecology*, **21**, 2852–2862.
- Saisa M, Koljonen ML, Gross R *et al.* (2005) Population genetic structure and postglacial colonization of Atlantic salmon (*Salmo salar*) in the Baltic sea area based on microsatellite DNA variation. *Canadian Journal of Fisheries and Aquatic Sciences*, **62**, 1887–1904.
- Schaffer WM, Elson PF (1975) Adaptive significance of variations in life-history among local populations of Atlantic salmon in North-America. *Ecology*, **56**, 577–590.
- Schluter D (2000) *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford, UK.
- Schluter D, Conte GL (2009) Genetics and ecological speciation. *Proceedings of the National Academy of Sciences*, **106**, 9955–9962.
- Seear PJ, Carmichael SN, Talbot R, Taggart JB, Bron JE, Sweeney GE (2010) Differential gene expression during smoltification of Atlantic salmon (*Salmo salar* L.): a first large-scale microarray study. *Marine Biotechnology*, **12**, 126–140.
- Seeb LW, Templin WD, Sato S *et al.* (2011) Single nucleotide polymorphisms across a species' range: implications for conservation studies of Pacific salmon. *Molecular Ecology Resources*, **11**, 195–217.
- Smith SJ, Campana SE (2010) Integrated stock mixture analysis for continuous and categorical data, with application to genetic-otolith combinations. *Canadian Journal of Fisheries and Aquatic Sciences*, **67**, 1533–1548.
- Stabell OB (1984) Homing and olfaction in salmonids - a critical-review with special reference to the Atlantic salmon. *Biological Reviews of the Cambridge Philosophical Society*, **59**, 333–388.
- Tamura K, Dudley J, Nei M, Kumar S (2007) Mega4: molecular evolutionary genetics analysis (mega) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.
- Taylor EB (1991) A review of local adaptation in Salmonidac, with particular reference to Pacific and Atlantic salmon. *Aquaculture*, **98**, 185–207.
- Tessier N, Bernatchez L (1999) Stability of population structure and genetic diversity across generations assessed by microsatellites among sympatric populations of landlocked Atlantic salmon (*Salmo salar* L.). *Molecular Ecology*, **8**, 169–179.
- Tessier N, Bernatchez L (2000) A genetic assessment of single versus double origin of landlocked Atlantic salmon (*Salmo salar*) from lake Saint-Jean, Québec, Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 797–804.
- Tonteri A, Veselov AJ, Titov S, Lumme J, Primmer CR (2007) The effect of migratory behaviour on genetic diversity and population divergence: a comparison of anadromous and freshwater Atlantic salmon *Salmo salar*. *Journal of Fish Biology*, **70**, 381–398.
- Verspoor E, Cole LJ (1989) Genetically distinct sympatric populations of resident and anadromous Atlantic salmon, *Salmo salar*. *Canadian Journal of Zoology*, **67**, 1453–1461.

- Ward HB (1932) The origin of the landlocked habit in salmon. *Proceedings of the National Academy of Sciences of the United States of America*, **18**, 569–580.
- Weir BS, Cockerham CC (1984) Estimating f-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- Wu C-I (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

C.P. and L.B. conceived the study, M.P.K. performed genotyping, and C.P. and V.B. analysed the data. All authors contributed to the interpretation of the results and to the writing of the manuscript. C.P. is postdoc in L.B. laboratory and is interested in various aspects of evolutionary biology and conservation genetics of fish populations. V.B. is currently studying the genomic basis of adaptive divergence and sea mortality of Atlantic salmon as a Ph.D. candidate under the supervision of L.B. His main interests are in population, landscape and conservation genomics. M.P.K.'s research interests revolve around the use of next-generation sequencing tools and data analysis for SNP discovery and genome assembly in non-model organisms, and the subsequent development of high-density SNP arrays and critical analysis and interpretation of raw SNP data. L.B.'s research focuses on understanding of the patterns and processes of molecular and organismal evolution as well as their significance to conservation.

Data accessibility

Data available at Dryad Digital Repository. doi:10.5061/dryad.7163d.

Supporting information

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

Table S1 Number and proportion of outlier SNPs detected potentially under divergent selection using hierarchical or pairwise genome scans and number of times outliers have been found potentially under divergent selection.

Fig. S1 LD (D') over all SNPs.

Fig. S2 Frequencies in freshwater and anadromous populations of the most frequent allele in freshwater for SNPs found under divergent selection between anadromous and freshwater populations.

Fig. S3 Frequencies in freshwater and anadromous populations of the most frequent allele in freshwater for SNPs repeatedly found under divergent selection between anadromous and freshwater populations.

Fig. S4 Sliding window analysis on LG8.