doi:10.1111/evo.12139

LANDSCAPE GENOMICS IN ATLANTIC SALMON (*SALMO SALAR*): SEARCHING FOR GENE–ENVIRONMENT INTERACTIONS DRIVING LOCAL ADAPTATION

Bourret Vincent,[1,2] Mélanie Dionne,[3] Matthew P. Kent,[4] Sigbjørn Lien,[4] and Louis Bernatchez[1]

[1]*Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, 1030 avenue de la Médecine, Québec, Québec G1V 0A6, Canada*

[2]*E-mail: vincent.bourret.1@ulaval.ca*

[3]*Direction de la faune aquatique, Ministère du Développement durable, de l'Environnement, de la Faune et des Parcs, Québec, G1S 4×4, Canada*

[4]*Department of Animal and Aquacultural Sciences, Centre for Integrative Genetics (CIGENE), Norwegian University of Life Sciences PO Box 5003, 1432 Aas, Norway*

Received September 15, 2012
Accepted March 27, 2013
Data Archived: Dryad doi: 10.5061/dryad.2pj70

A growing number of studies are examining the factors driving historical and contemporary evolution in wild populations. By combining surveys of genomic variation with a comprehensive assessment of environmental parameters, such studies can increase our understanding of the genomic and geographical extent of local adaptation in wild populations. We used a large-scale landscape genomics approach to examine adaptive and neutral differentiation across 54 North American populations of Atlantic salmon representing seven previously defined genetically distinct regional groups. Over 5500 genome-wide single nucleotide polymorphisms were genotyped in 641 individuals and 28 bulk assays of 25 pooled individuals each. Genome scans, linkage map, and 49 environmental variables were combined to conduct an innovative landscape genomic analysis. Our results provide valuable insight into the links between environmental variation and both neutral and potentially adaptive genetic divergence. In particular, we identified markers potentially under divergent selection, as well as associated selective environmental factors and biological functions with the observed adaptive divergence. Multivariate landscape genetic analysis revealed strong associations of both genetic and environmental structures. We found an enrichment of growth-related functions among outlier markers. Climate (temperature–precipitation) and geological characteristics were significantly associated with both potentially adaptive and neutral genetic divergence and should be considered as candidate loci involved in adaptation at the regional scale in Atlantic salmon. Hence, this study significantly contributes to the improvement of tools used in modern conservation and management schemes of Atlantic salmon wild populations.

KEY WORDS: **Conservation, local adaptation, population genetics, salmonids, single nucleotide polymorphism.**

The environment can influence evolutionary trajectories of living organisms by imposing selective pressures and limiting migration. The last decade has witnessed the birth of landscape genetics, a field devoted to understanding the contribution of environmental conditions on the evolutionary processes shaping population genetic structure in the wild (Manel et al. 2003). Although the

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

emerging phase of landscape genetics mainly focused on linking neutral genetic divergence with ecological constraints to gene flow (e.g., Petren et al. 2005; Leclerc et al. 2008), a limited number of studies have also accounted for adaptive divergence (Bonin et al. 2006; Gaggiotti et al. 2009; Manel et al. 2010a, b). The use of relatively low genomic coverage and anonymous markers often limited the potential for functional inferences. Today, the availability of high throughput genomic tools combined with genome scan facilitates the investigation of adaptive genomic divergence. Ecological and landscape genomics now have greater power to disentangle adaptive from neutral genetic divergence and identify the environmental factors driving divergent selection (Bonin 2008). Furthermore, the possible links between newly developed single nucleotide polymorphism (SNP) markers and functional genes can reveal key biological processes or functions targeted by environmental selective pressures (Bonin 2008; Parisod and Holderegger 2012).

In widely distributed species, local populations often experience heterogeneous environmental conditions, and may evolve in these different environments for thousands of years. In such cases, these environmental conditions are suspected to have shaped distinct genetic composition among populations, and local adaptation of genetically based phenotypic traits. In general, local adaptation results in the superior fitness of indigenous individuals compared to emigrants (Kawecki and Ebert 2004). Nevertheless, the geographic extent of local adaptation can vary depending on the nature of the adaptation, the strength of the selection, the extent of gene flow between populations, and their effective population size (Lacy 1997; Hansen et al. 2002). However, the genomic extent of local adaptation can also differ among populations depending on the degree of genetic isolation (Feder and Nosil 2010) or the complexity of the trait under selection (e.g., single vs. multilocus, pleiotropy, epistasis). Even in light of recent technical advances, investigating the adaptive divergence of wild populations occupying vast heterogeneous environments is a daunting task, especially on nonmodel species.

Although Atlantic salmon is not a classical model organism, the population genetic structure of this species has been extensively studied (e.g., Vasemägi et al. 2005; King et al. 2007; Palstra et al. 2007; Tonteri et al. 2009; Bourret et al. 2013). In North America, the most extensive study on this species was performed by Dionne et al. (2008) and involved a comprehensive landscape genetics approach aiming to elucidate the environmental parameters influencing the genetic structure of 51 populations. Dionne et al. (2008) found that temperature regime and coastal distance from a southern reference influenced neutral divergence among populations, which suggested a regional component to local adaptation. A hierarchical structure analysis grouped populations into seven regional groups based upon differentiation at microsatellite markers. The recent development of a large panel of SNPs, largely discovered from coding regions, offers a more powerful way to gain insights into the possible role of environmentally induced selective pressures in shaping patterns of adaptive divergence and further assess the geographic scale of local adaptation in Atlantic salmon (Bourret et al. 2013).

Local adaptation in salmonids has been recognized as a key evolutionary process driving phenotypic and genetic divergence among populations (Taylor 1991; Garcia de Leaniz et al. 2007; Fraser et al. 2011). However, local adaptation for these species is still equivocal in front of the difficulties of performing common garden and reciprocal transplant experiments, especially on large geographic scales. In Atlantic salmon, different populations exhibit divergence in morphological traits, migratory tactics, and reproductive strategies that could be associated with local adaptation since they have been shown to be heritable (Claytor et al. 1991; Palstra et al. 2007; Vähä et al. 2007; Paez et al. 2010). Moreover, recent studies have proposed that genetic diversity at MHC class-IIB genes could represent local adaptation to cope with pathogen diversity in rivers with different thermal regimes (Dionne et al. 2007, 2009).

The main objective of this study was to further investigate the environmental factors shaping patterns of genetic divergence and the scale of local adaptation in Atlantic salmon. We accomplished this by increasing the number of genetic markers used by 150-fold and including a larger set of environmental variables. More specifically, we first revisited the population genetic structure of Atlantic salmon to assess the congruence of a wider genomic coverage with previously shown genetic structure. We then tested for significant associations between variation in 49 climate, geological, and river specific characteristics and regional genetic structure at markers identified as being potentially under divergent selection. Third, using mapping information, we documented the genomic distribution of these outlier markers. Finally, using available gene annotations, we examine the functional implications of adaptive and environmental divergence among populations and regional groups of Atlantic salmon.

Materials and Methods

SAMPLES

Samples of adult anadromous Atlantic salmon were collected in the summer of 2004. Methods for tissue collection, storage, DNA extraction, and microsatellite analyses for samples collected from 51 rivers were previously detailed in Dionne et al. (2008). In this study, we also added samples from three new rivers. DNA was extracted from fin clips as described by Dionne et al. (2008) from a total of 1341 individuals from 54 rivers in Eastern Canada (Tables 1 and 3; Fig. 1). SNP genotyping was conducted using two different approaches. We first performed a single individual

Table 1. Description of regional groupings and parameters associated with sample sites composing the groups: latitude and longitude, number of individuals genotyped (*N*), average call rate per population (CR), and average expected (HE) and observed (HO) heterozygosities per population.

Regional groups	Population ID	Code	Latitude	Longitude	*N*	CR	HO	HE
Southern Québec	Miramichi	MIR	47.09	− 65.32	25	0.99	0.181	0.179
Matapédia	MAP	47.97	− 66.93	25	0.97	0.204	0.202	
Grande Cascapédia	CS	48.21	− 65.90	25	0.98	0.195	0.191	
St-Jean Gaspésie	SJQG	48.77	− 64.43	25	0.99	0.205	0.202	
Sainte-Anne	SA	49.12	− 66.50	25	0.99	0.199	0.196	
Matane	MAT	48.85	− 67.53	25	0.98	0.203	0.202	
Québec City	Malbaie	ML	47.65	− 70.13	25	0.99	0.182	0.176
Du Gouffre	DGO	47.43	− 70.49	25	0.99	0.181	0.175	
Sainte-Marguerite	SM	48.25	− 69.93	25	0.99	0.178	0.173	
Higher North Shore	Trinité	TRI	49.42	− 67.30	25	0.99	0.179	0.179
Moisie	MOI	50.20	− 66.07	25	0.99	0.181	0.176	
St-Jean Côte-Nord	SJQC	50.28	− 64.33	25	0.99	0.185	0.181	
Natashquan	NAT	50.12	− 61.80	25	0.99	0.170	0.174	
Lower North Shore	Musquaro	MUS	50.22	− 61.07	25	0.99	0.163	0.168
Etamamiou	ET	50.27	− 59.97	25	0.99	0.166	0.169	
Gros Mécatina	MEC	50.77	− 59.08	25	0.98	0.162	0.162	
Anticosti	Jupiter	JU	49.47	− 63.58	25	0.99	0.202	0.197
Aux Sumons	SU	49.42	− 62.23	25	0.98	0.196	0.191	
Chaloupe	CHA	49.13	− 62.53	23	0.97	0.201	0.198	
Labrador	Napetipi	NAP	51.31	− 58.06	25	0.99	0.185	0.185
Saint-Paul	STP	51.47	− 57.70	25	0.98	0.180	0.182	
Vieux-Fort	VF	51.32	− 58.02	25	0.99	0.190	0.188	
Southwest Brook	SW	53.42	− 57.23	25	0.99	0.177	0.174	
Ungava	George	GE	58.82	− 66.17	18	0.99	0.160	0.156
Koksoak	KO	58.53	− 68.17	25	0.99	0.165	0.163	
Aux Feuilles	AF	58.77	− 70.07	25	0.99	0.159	0.157	

Table 2. Analysis of molecular variance (AMOVA) using neutral markers ($n = 3016$) and divergent outlier markers ($n = 68$). **P*-value < 0.001.

Source of variation	df	Neutral SNPs	Divergent SNPs
Among groups	6	6.42*	28.04*
Among populations within groups	19	3.40*	3.00*
Within populations	1254	90.18*	68.95*

genotyping approach for 641 fish from 26 rivers representative of the seven previously identified regional groups, with an average of 25 individuals per river. Second, to increase the number of populations analyzed while limiting the cost of genotyping, we performed a bulk genotyping approach for 700 individuals from 28 rivers. Previous studies have shown that reliable SNP allele frequency estimates could be obtained by this method (Macgregor et al. 2008; Craig et al. 2009). To prepare bulk assays, DNA from 25 individuals per river was quantified in triplicate using Quant-iT PicoGreen dsDNA Assays (Life Technologies, Carlsbad, CA). The DNA concentrations for all 25 individuals were standardized to the concentration of the individual with the lowest DNA concentration. For each river, those 25 individuals were then pooled into a single 50-μL bulk assay and used as individual for the genotyping steps.

GENOTYPING QUALITY CONTROL

From the 26 populations genotyped individually, four were genotyped using version one (V1) of the SNP array (described in Bourret et al. 2013) as they were used in the first assessment study of the array. The remaining populations (22 in individual genotyping and 28 in bulk assays) were genotyped using version two (V2) of the SNP array developed by the Centre for Integrative Genetics (CIGENE, Âs, Norway). A total of 5349 SNP markers on V2 were selected from V1 for their high quality and 219 SNPs were added from new sequence data. These additional markers were assessed in the same fashion as markers on V1 and discovery and quality control methods for all 5568 SNPs on V2 can be found

Table 3. Discriminant analysis population assignment results for bulks assays in contrast to previous classification of Dionne et al. (2008). NA refers to population not previously classified.

Number	Population ID	Previous classification	Bulk assigned to
1	Ouelle	Québec City	Southern Québec
2	Laval	Higher North Shore	Southern Québec
3	Patapédia	Southern Québec	Southern Québec
4	Bonaventure	Southern Québec	Southern Québec
5	Petite Cascapédia	Southern Québec	Southern Québec
6	Cap-Chat	Southern Québec	Southern Québec
7	York	Southern Québec	Southern Québec
8	Madeleine	Southern Québec	Southern Québec
9	Grand Pabos	Southern Québec	Southern Québec
10	Darmouth	Southern Québec	Southern Québec
11	Causapscal	Southern Québec	Southern Québec
12	Mitis	Southern Québec	Southern Québec
13	Upsalquitch	Southern Québec	Southern Québec
14	Little Main	Southern Québec	Southern Québec
15	Kegwick	Southern Québec	Southern Québec
16	Des Escoumins	NA	Québec City
17	Jacques Cartier	Québec City	Québec City
18	Petit Saguenay	Québec City	Québec City
19	Kecarpui	NA	Lower North Shore
20	Corneille	Lower North Shore	Lower North Shore
21	Muddy Bay	Labrador	Labrador
22	Sand Hill	Labrador	Labrador
23	Eagle	Labrador	Labrador
24	Godbout	Higher North Shore	Higher North Shore
25	Aux Anglais	Higher North Shore	Higher North Shore
26	Watshishou	Higher North Shore	Higher North Shore
27	Aganus	Higher North Shore	Higher North Shore
28	Aux Rochers	Higher North Shore	Higher North Shore

in Bourret et al. (2013). Genotyping was performed according to the manufacturer's instructions using the Illumina infinium assay (Illumina, San Diego, CA).

Samples with greater than 85% call rate (CR; proportion of SNPs successfully genotyped) were retained for future analyses. Markers absent from V2, but genotyped on V1 in four populations, were excluded and subsequent quality control steps were then performed on the remaining V2's 5568 SNPs. Using Illumina's Genotyping Module software, we assessed each SNP's cluster pattern using all individual of North American available populations ($n = 900$ individuals). Visual inspection allowed for the classification of SNPs into different categories: (i) single SNP, (ii) failed, (iii) monomorphic, and (iv) paralogous sequence variants (PSVs), and (v) multisite variants (MSVs; Table S1). Markers falling in categories other than "single SNP" were excluded from further analyses as well as markers with minor allele frequency less than 1 percent (MAF < 0.01). Ascertainment bias was assessed in Bourret et al. (2013) and was suggested to be minimal in North American populations based on an L-shape distribution of MAF (high proportion of low-frequency markers rapidly decreasing toward low proportion of high-frequency markers).

POPULATION STRUCTURE ON INDIVIDUAL SAMPLES

Using individually genotyped samples (26 rivers), we measured global and per SNP observed and expected heterozygosity (H_O and H_E) within each population. To exclude markers potentially under divergent or balancing selection from the basic population structure analyses, we then used hierarchical F_{dist} (Excoffier et al. 2009), a genome scan analysis implemented in ARLEQUIN 3.5 (Excoffier and Lischer 2010). This hierarchical method can detect outlier loci among groups of populations (F_{CT}) and represents the most appropriate method for detection of selected markers in our system because previous studies found a hierarchical structure among North American populations (Dionne et al. 2008, 2009). To this end, we classified rivers in seven regional groups previously identified by Dionne et al. (2008) for the detection of selection analysis. Markers with F_{CT} P-values > 0.01 were then used as the neutral basis for population structure analyses.

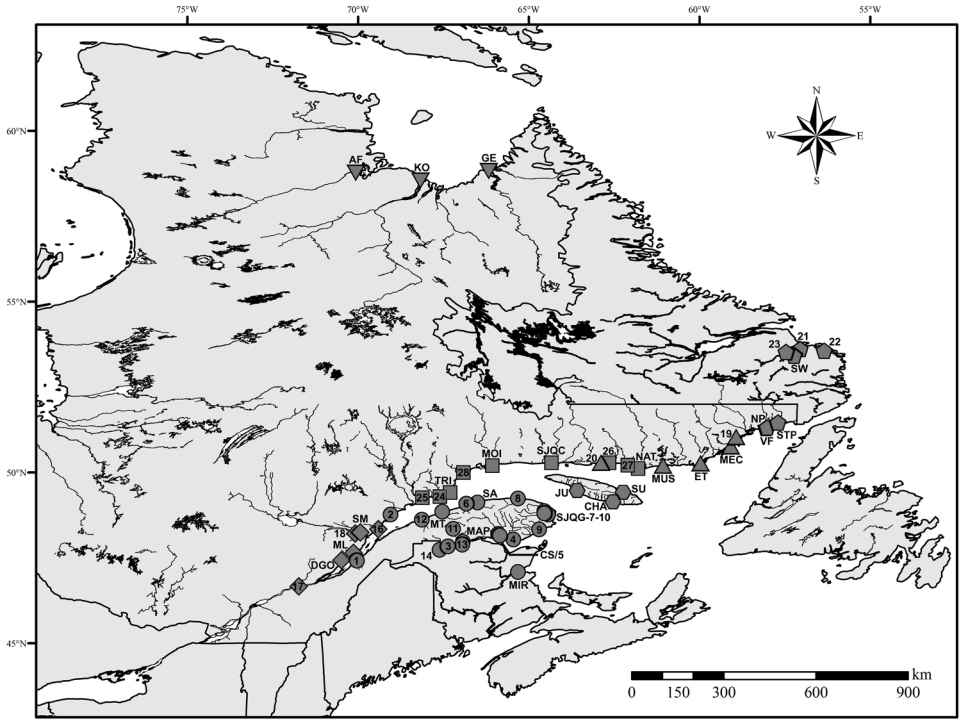

Figure 1. **Map showing sample sites. Populations are linked to the river codes in Table 1 for individually genotyped populations (n = 26) and numbers in Table 3 for bulk assays. Tick forms relate to regional groups as reported in Dionne et al. (2008).**

Pairwise genetic differentiation between populations was estimated by the F_{ST} estimate of Weir and Cockerham (1984) using ARLEQUIN 3.5 with 10,000 permutations to determine statistical significance. To confirm the regional classification proposed by Dionne et al. (2008), a principal component analysis on individual genotypes was carried out using the SmartPCA program implemented in the EIGENSOFT package (Patterson et al. 2006) and fitted using R (R core development team). Furthermore, using the confirmed regional structure, two analyses of molecular variance (AMOVAs) were performed using ARLEQUIN 3.5, one on markers identified as potentially under divergent selection and a second on neutral markers.

BULK ASSAYS POPULATION STRUCTURE

In the Illumina's Genotyping Module software used to call genotypes, red and green signals (representing homozygotes AA and BB) are normalized to θ values of 0 and 1. The expected θ value for a heterozygote is 0.5, but generally, clusters are biased toward 0 or 1. To accurately predict allele frequencies from bulk assays, we calculated a k correction factor for each SNP, which is the average θ value for heterozygotes with $k = \theta$ hets/$(1 - \theta$ hets). We selected the heterozygotes from genotyped individuals with CR > 0.99 to estimate the correction factor as accurately as possible. We then calculated the B allele frequency (BAF) on markers retained following the previously described filtering procedure while also excluding markers without heterozygotes among individuals (CR > 0.99) with BAF = θ bulkassay/(θ bulkassay + k(1 − θ bulkassay). To test the robustness of this genotyping method, bulk assays from five rivers that were previously genotyped using the single individual genotyping approach were also genotyped using the same individuals pooled in a single bulk assay per population (two replicates). A population's allele frequencies from both bulk assays and individual genotypes for each SNP were then contrasted using a simple regression and fitted using R. To further confirm the Dionne et al. (2008), regional classification and classify the additional three populations genotyped as bulk assays, we used a discriminant analysis implemented in SAS 9.1 (proc discrim; SAS Institute Inc., Cary, NC), which acts as a custom population assignment method. In the first step of this analysis, the population's allele frequencies for the 26 individually genotyped rivers are used to build a discriminant rule that classifies population in their regional group based on statistical distances estimated by the procedure. Then, using this discriminant rule, bulk assays are given a probability of belonging to either one of the regional groups based on estimated population allele frequencies (BAF).

ENVIRONMENTAL STRUCTURE

Rivers (n = 26) used in individual genotyping were also characterized for 49 environmental parameters distributed among three main categories, namely climate, river properties, and geological variables. A detailed list of environmental parameters is available in Table 4. Climate variables (n = 19) were extracted from 35

Table 4. Description and summary of environmental parameters loadings on 10 retained principal component (PC) factors after PCA on 49 parameters. Parameters are ordered according to their primary PC factor loading. Gray and white areas refers to alternance between PC factors row associations.

Parameter descriptions	PC factor 1	PC factor 2	PC factor 3	PC factor 4	PC factor 5	PC factor 6	PC factor 7	PC factor 8	PC factor 9	PC factor 10
Climate—Average temperature May–September in °C	0.98	−0.12	0.02	0.03	0.00	0.03	−0.02	0.03	0.09	0.00
Climate—Degree-days >10°C from April to October	0.96	−0.14	0.01	0.06	0.16	0.05	0.12	0.05	−0.01	−0.01
Climate—Degree-days >5°C from April to October	0.95	−0.08	0.07	0.11	0.14	0.04	0.15	0.10	0.02	−0.03
Climate—Yearly average temperature	0.94	−0.04	−0.03	0.18	0.11	−0.11	0.08	0.17	0.01	−0.01
Climate—Degree-days >15°C from April to October	0.93	−0.21	−0.06	−0.01	0.24	0.06	0.11	−0.02	−0.06	−0.01
Climate—Degree-days >0°C from April to October	0.91	−0.06	0.14	0.13	0.15	0.04	0.21	0.15	0.02	−0.05
Climate—Days with temperature >20°C in a year	0.89	−0.32	−0.07	−0.05	0.16	0.07	0.17	−0.13	−0.02	−0.01
Climate—Degree-days >18°C from April to October	0.88	−0.18	−0.06	−0.06	0.38	0.06	0.10	−0.05	−0.09	−0.01
River property—Longitude in decimal	0.80	−0.27	0.12	−0.01	−0.26	0.21	−0.20	−0.01	0.28	0.09
Climate—Average temperature December–March in °C	0.72	0.07	−0.08	0.33	0.24	−0.23	0.20	0.33	−0.07	0.04
Climate—Average temperature October–November in °C	0.67	0.10	−0.19	0.18	0.22	−0.39	0.18	0.34	−0.18	−0.04
Geological—RockType—Sedimentary	0.59	−0.03	−0.04	0.58	0.08	0.22	0.45	0.21	−0.03	−0.07
Geological—Era—Paleozoic	0.59	−0.03	−0.04	0.58	0.08	0.22	0.45	0.21	−0.03	−0.07
Climate—Yearly snowfall in cm	−0.59	−0.08	0.12	−0.04	0.23	0.30	0.14	−0.44	−0.36	−0.04
River property—Coastal distance from Miramichi River	−0.77	0.01	−0.04	−0.37	−0.08	−0.15	−0.35	−0.06	−0.16	−0.04
River property—Latitude in decimal	−0.96	−0.02	−0.02	−0.09	0.06	−0.07	0.04	−0.01	−0.07	−0.08
Climate—Rainfall from May to October in mm	−0.17	0.95	0.01	0.08	0.02	0.13	−0.02	−0.10	0.08	0.05
Climate—Days with rainfall >10 mm	−0.16	0.94	−0.08	−0.03	−0.06	0.07	−0.04	−0.14	0.02	0.09
Climate—Yearly rainfall in mm	0.12	0.92	0.08	0.20	0.13	−0.16	−0.04	−0.07	0.07	−0.10
Climate—Days with rainfall > 25 mm	0.19	0.83	0.24	0.30	−0.07	−0.07	−0.07	−0.17	0.11	−0.08
Climate—Days with rainfall > 5 mm	−0.44	0.76	−0.20	−0.18	−0.07	0.13	−0.05	0.00	0.09	0.18
Climate—Rainfall from July to September in mm	−0.40	0.75	0.03	−0.03	0.07	0.37	0.11	−0.11	−0.02	0.11

Continued.

Table 4. Continued

Parameter descriptions	PC factor 1	PC factor 2	PC factor 3	PC factor 4	PC factor 5	PC factor 6	PC factor 7	PC factor 8	PC factor 9	PC factor 10
Climate—Yearly precipitation (rain + snow) in mm	−0.32	0.70	0.17	0.12	0.29	0.10	0.06	−0.37	−0.21	−0.06
Geological—SubRockType—Paragneiss	−0.13	0.59	−0.22	−0.33	−0.13	−0.14	−0.02	0.16	−0.15	−0.19
River property—River length	−0.14	0.12	0.94	−0.12	0.05	0.05	0.02	−0.02	0.04	−0.15
River property—Migration difficulty (slope)	−0.07	0.03	0.92	−0.10	0.03	0.08	0.04	0.03	−0.04	−0.26
River property—Exploitation capacity	−0.08	0.03	0.92	−0.08	0.03	−0.01	−0.10	−0.16	0.04	0.07
River property—Maximum altitude	0.13	−0.16	0.64	−0.16	−0.04	0.23	0.48	0.21	0.23	−0.16
River property—Percentage of MSW	0.38	−0.12	0.62	0.03	−0.32	−0.05	0.07	0.03	0.46	0.28
River property—Percentage of Grilse	−0.38	0.12	−0.62	−0.03	0.32	0.05	−0.07	−0.03	−0.46	−0.28
Geological—Province—Plate-forme du Saint-Laurent	0.16	0.06	−0.18	0.91	−0.11	−0.10	−0.21	0.05	−0.06	−0.06
Geological—Period—Silurian	−0.06	0.15	−0.15	0.90	−0.03	−0.07	−0.05	−0.02	0.09	−0.01
Geological—SubRockType—Undivided sedimentary	0.41	0.04	−0.05	0.70	−0.29	−0.18	0.42	−0.13	−0.06	−0.07
Geological—SubRockType—Nonmarine sedimentary	0.43	0.06	−0.01	−0.10	0.88	−0.05	−0.03	−0.04	−0.02	−0.02
Geological—Period—Carboniferous	0.43	0.06	−0.01	−0.10	0.88	−0.05	−0.03	−0.04	−0.02	−0.02
Climate—Days with rainfall > 30 mm	0.64	−0.07	−0.04	−0.03	0.72	−0.08	0.05	−0.11	−0.08	−0.05
Geological—Cambrian-Ordovician	0.05	0.10	0.01	−0.01	−0.02	0.98	0.04	−0.02	−0.01	0.00
Geological—SubRockType—Oceanic domain miogeoclinal	0.05	0.10	0.01	−0.01	−0.02	0.98	0.04	−0.02	−0.01	0.00
River property—Index described in Dionne et al. (2008)	0.32	−0.04	0.41	−0.14	−0.18	0.61	0.09	0.06	−0.24	−0.15
Geological—Period—Devonian	0.38	−0.01	0.14	−0.08	−0.27	−0.14	0.81	−0.24	−0.02	−0.03
Geological—Province—Apalachian Orogen	0.53	−0.08	0.11	−0.14	0.19	0.33	0.69	0.19	0.02	−0.03
Geological—Period—Mesoproterozoic	−0.06	0.06	0.27	−0.44	−0.25	−0.15	−0.58	−0.20	0.38	−0.18
Geological—SubRockType—Mix	0.02	−0.32	0.02	−0.06	−0.01	0.03	0.13	0.87	0.10	0.02
Geological—Period—Ordovician	0.31	−0.33	−0.05	0.12	−0.12	−0.03	−0.12	0.72	−0.11	−0.05

Continued.

Table 4. Continued

Parameter descriptions	PC factor 1	PC factor 2	PC factor 3	PC factor 4	PC factor 5	PC factor 6	PC factor 7	PC factor 8	PC factor 9	PC factor 10
Geological—RockType—Intruded	−0.14	0.21	0.10	−0.07	0.06	−0.06	−0.04	0.08	0.78	0.03
River property—Level of exploitation	0.39	−0.13	0.12	0.34	−0.17	−0.03	0.07	−0.13	0.49	−0.15
Geological—SubRockType—Undivided gneiss	−0.21	−0.30	−0.06	−0.32	−0.10	−0.10	−0.35	−0.27	−0.08	0.69
Geological—Period—Paleoproterozoic–Mesoproterozoic	−0.46	0.22	−0.24	−0.19	0.10	−0.13	0.07	0.08	−0.42	0.56
Geological—SubRockType—Orthogneiss	−0.39	−0.18	0.27	−0.10	0.08	−0.04	−0.14	−0.12	−0.20	−0.71

meteorological stations where values of numerous temperature and precipitation parameters were collected over a period of approximately 30 years (1971–2000, Environment Canada, http://climate.weatheroffice.gc.ca/Welcome_e.html). We used data directly from a station to characterize a river when it was located within 50 km of the mouth of the river. Otherwise, we estimated values for a subset or all climate parameters using a Kriging interpolation with the Geostatistical Analyst in ArcGIS 9.2 (ESRI Inc., Redlands, CA), a common geostatistical technique to predict values at unmeasured locations. River properties ($n = 11$) were obtained from the Ministère des Ressources naturelles et de la Faune du Québec (MRNF) and the Canada3D database available through the GéoGratis website of the Ministry of Natural Resources Canada (NRC; http://geogratis.cgdi.gc.ca) as described in Dionne et al. (2008). When a river property parameter was missing for a given river, it was replaced by the average across all populations for that parameter (total of 23/437 cases for 6/19 parameters). Six categorical geological parameters were then considered and divided in 19 subcategories. The dominant subcategory (scored as 1) in a river's watershed for each parameter was determined by estimating the area covered by polygons of the geological layer associated with each of the subcategories and identifying the subcategory with the highest coverage. This was realized by examining the intercept between a watershed layer provided by the MRNF and a geological layer from the GeoScape Canada database available on the Natural Research Council of Canada (NRC) website (http://www.nrcan.gc.ca/earth-sciences/products-services/mapping-product/geoscape/6032). Each geological parameter was then transformed into presence/absence scores for each subcategory where only the dominant subcategory of a given river was scored as 1. This transformation was performed to include these categorical parameters along with continuous variables (climate and river properties) in further analyses. To minimize the colinearity among all 49 environmental parameters, we performed a principal component analysis on populations using SAS 9.1 (proc factor; rotation = varimax). As a surrogate for environmental parameters, we then used a number of principal component factors (PC factors) equal to the number of eigenvalues greater than 1, a widely used statistical rule known as the Kaiser–Guttman criterion (Yeomans and Golder 1982).

GENETIC–ENVIRONMENT ASSOCIATIONS

Association between population genetic structure and environmental parameters was assessed *via* a redundancy analysis (RDA), which is a special case of canonical correlation analysis (CCA). The CCA is a statistical test used to relate information from two different data tables. Here, using rivers as subjects, we specifically tested if the independent parameters (environmental PC factors) could predict the dependent parameters (allele frequencies).

An analysis of variance (ANOVA; 1000 permutations) was then performed to assess the global significance of the RDA and a marginal ANOVA (1000 permutations) was also run to determine if environmental PC factors were significantly correlated with allele frequencies. We also estimated the effect size of the relationship between the two datasets using Wilks' Lambda, which is a parameter analogous to the correlation coefficient (R^2). We then computed Pearson's correlation coefficients for all markers used in the RDA and environmental PC factors. Moreover, to localize potential genomic regions under environmentally divergent selection, we used a linkage map for the North American Atlantic salmon (Brenna-Hansen et al. 2012) built using the same SNP array to map markers potentially under divergent selection. We mapped each SNP according to the results from the hierarchical F_{dist} as well as the Pearson's correlation coefficient values for each significantly associated environmental factor. Analyses from this section were all performed in R.

GENE ONTOLOGY AND SNP ANNOTATION

Blast2go (Gotz et al. 2008) was used to associate gene ontology (GO) annotation terms to all SNPs retained for genomic analyses (3118 SNPs; see Results). A homology search was first completed by performing a BLAST (Altschul et al. 1990) search of the available flanking sequences for each SNP on the NCBI nr public database with the *e*-value threshold set to 1×10^{-10}. Blast2go then retrieved GO terms associated with the obtained BLAST hits. To determine if the biological processes, molecular functions or cellular components of the markers potentially under divergent selection were over-, equally, or underrepresented among outlier markers when compared to the entire retained SNP dataset, we performed an enrichment analysis using Fisher's exact test corrected for multiple tests by applying a false discovery rate (FDR) of 0.05 (Benjamini and Hochberg 1995).

Results

GENOTYPING AND QUALITY CONTROL

One individual sample (SU-14) with a CR < 0.85 (0.77) was excluded from the dataset. After initial quality control and classification of genotypes obtained from 900 samples, we classified 3974 markers out of 5568 SNPs features on the V2 array as single locus and polymorphic SNPs (i.e., diploid SNPs) for North American Atlantic salmon. Among the 3974 "good" SNPs, 3118 markers showed an overall MAF > 0.01. Therefore, besides the individuals genotyped in the bulk assays, 3118 SNPs and 640 individuals were kept for further analyses. Table 1 shows summary data for call rates (CR), observed (H_O), and expected heterozygosity (H_E) across populations (Table S2 across markers).

POPULATION STRUCTURE ON INDIVIDUAL SAMPLES

When we followed the regional grouping proposed by Dionne et al. (2008), the average F_{CT} across 3118 loci was 0.057 (ranging from -0.026 to 0.535). At the 0.01 and 0.05 significance level respectively, 68 and 179 markers were identified as potentially under divergent selection and 34 and 208 markers were identified as potentially under balancing selection (Table S3). Removing outlier markers at the 0.01 significance level (68 divergent and 34 balanced), 3016 markers were used as the basis for the neutral pairwise differentiation estimates, PCA and AMOVA. All pairwise comparisons of genetic differentiation between populations were highly significant ($P < 0.001$; Table S4). In a principal component analysis (PCA) on individual genotypes, five principal component (PC) factors determined at least 1% of the variation each, and together explained 10.5% of the total genetic variation among individuals. Principal components 1–3 accounted for 4.0%, 2.3%, and 1.6%, respectively. Principal components 1–3 differentiated populations into the seven regional groups that Dionne et al. (2008) previously defined (Fig. 2). Both AMOVAs on 3016 neutral SNPs and 68 divergent SNPs showed significant genetic variation among groups (Table 2). For neutral markers, the genetic variation attributed to differences among groups accounted for 6.42% whereas this percentage increased to 28.06% for divergent markers. Differentiation among populations within groups was similar with 3.04% and 3.00% for neutral and divergent SNPs, respectively. Thus, intergroup differentiation was about 4.5 times more pronounced at outlier markers than neutral loci whereas interpopulation differentiation within groups remained the same.

BULK ASSAYS POPULATION STRUCTURE

Of the available individuals genotyped ($n = 900$), 728 had CR > 0.99 and were thus used to estimate the marker's k correction factors. Of the 3118 markers used in previous analyses, 287 markers did not have heterozygotes to estimate the correction factor and were excluded from bulk assays analyses. Correlations of population allele frequencies estimated from bulk assays to those measured with actual individual genotypes yielded correlation coefficients (R^2) ranging from 0.898 (MAP) to 0.921 (DGO) with all *P*-values < 0.001. Globally, population allele frequencies estimated from bulk assays were highly correlated with those estimated from individual genotypes with an overall $R^2 = 0.909$ (*P*-value < 0.001; Fig. S1). Based on the allele frequencies of the 26 populations genotyped on an individual basis and on the regional group they belonged to, we were able to build a powerful discriminant rule. Regional distances estimated *via* the discriminant analysis allowed us to assign all 28 bulk-assayed populations to one of the seven regional groups with a probability of 1.00 (Table 3). Out of the 26 previously classified populations (Dionne et al. 2008), only the Ouelle and Laval rivers were both

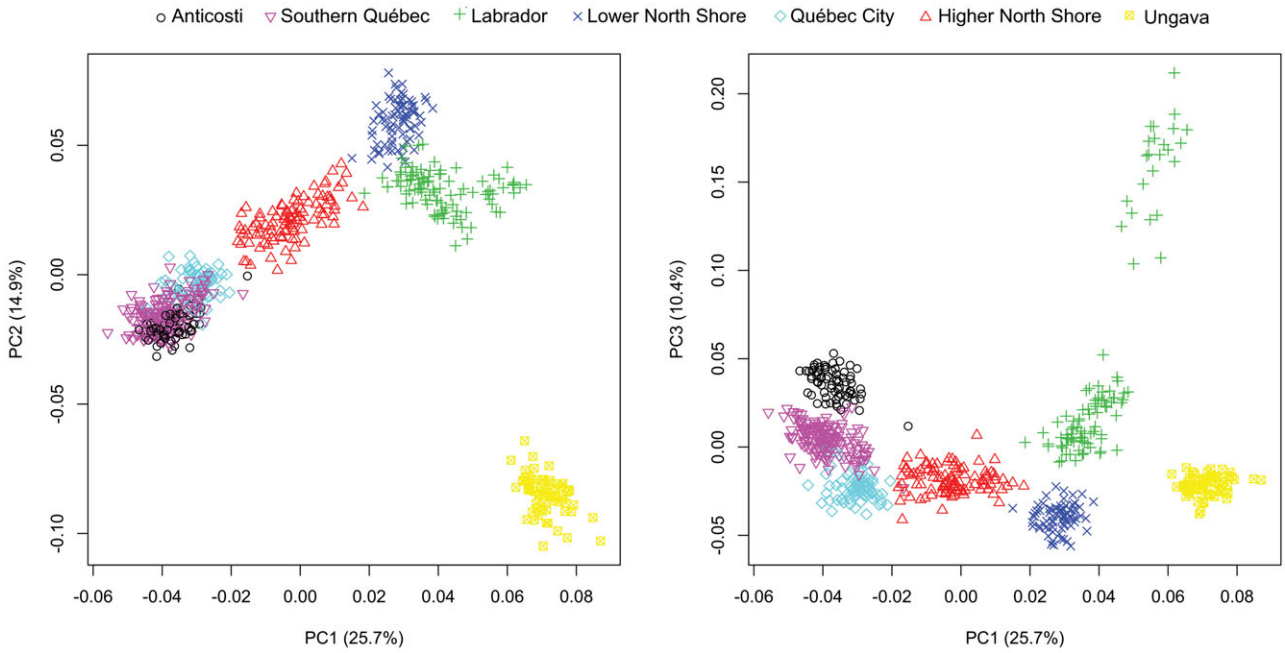

Figure 2. Principal components analysis of genetic differentiation among individuals based on 3016 SNP markers (each point represents one individual) with principal component 1 (PC1: 4.02% of variance) against PC2 (2.33% of variance) on the left panel and PC1 against PC3 (1.62% of variance) on the right panel. Color and tick form reflect population's regional groups as reported in Table 1.

"missassigned" to the Southern Québec region instead of the Québec City region for the former and the Higher North Shore region for the latter population.

ENVIRONMENTAL STRUCTURE

In the remote Ungava region, compiled climate data were available for only a single meteorological station and the next closest stations were located over 1000 km away from the mouth of the rivers. Therefore, for these rivers, the interpolation yielded values outside a reasonable confidence interval (data not shown). For this reason and considering that these populations encounter extreme climatic conditions, the three rivers from Ungava (AF, GE, and KO) were left out of environmental analyses to avoid confounding factors due to biased outlier values. Of the 23 rivers left, 11 were located within close distance to a meteorological station and did not require interpolated climate data. We then used the interpolated climate data partially for three rivers (for degree-days) and completely for the remaining nine rivers. A PCA on environmental parameters presented 10 PC factors with eigenvalues greater than 1, and together explained 92.1% of the total genetic variation among individuals. Principal components factors 1–4 accounted for 33.8%, 12.8%, 11.3%, and 8.3%, respectively. PC factor 1, PC factor 3, and PC factor 4 differentiated the populations along the spatial axes whereas PC factor 2 offered a similar dispersion of populations as PC factor 1 (not shown). In general, populations tended to group according to previously identified regional groups but several populations showed less clear patterns (Fig. 3). Different PC factors were dominated by different categories of environmental parameters. As indicated by the loadings of parameters on the 10 PC factors retained (Table 4), PC factor 1 was primarily loaded with climate data related to temperature (10/16), PC factor 2 by climate data related to precipitation (7/8), PC factor 3 by river properties (6/6), PC factors 4 and 7 by geological data (3/3 and 3/3, respectively).

GENETIC–ENVIRONMENT ASSOCIATIONS

Using the 23 rivers as subjects and the 10 environmental PC factors as explanatory variables, Figure 4 shows a RDA performed on the 179 SNPs potentially under divergent selection (0.05 significance level) as the response variables. We used this outlier threshold to be consistent with correlation and annotation analyses where a less stringent threshold for outliers was used. Moreover, RDA results using more or fewer markers were similar, and adding more markers increased variance and *P*-values (data not shown). Globally, the RDA was highly significant with a P-value < 0.001 (ANOVA, $F = 5.025$). The first 10 RDA axes accounted for 80.7% of the variation, whereas the RDA axes 1 and 2 represented 49.7% and 12.4%, respectively. The marginal ANOVA showed that PC factors 1, 2, 4, and 7 were significant predictors of the populations' allele frequencies with P-values < 0.001 (respective $F =$ 20.617, 3.379, 10.023, and 4.536; Fig. 4). Principal component factors 3, 9, and 10 presented significant relationships but to a lesser degree (P-values between 0.015 and 0.038). The correlation between the multivariate environmental and genetic structure

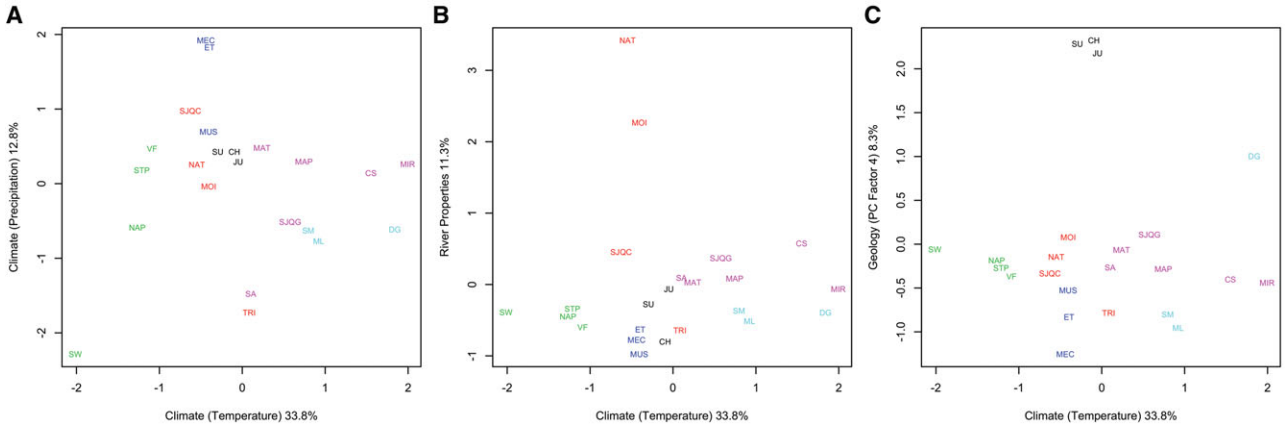

Figure 3. Principal components analysis of environmental parameters ($n = 49$) among populations ($n = 23$) with: (A) climate-related PC factor 1 (temperature; 33.8% of variance) against climate-related PC factor 2 (precipitation; 12.8% of variance); (B) PC factor 1 against PC factor 3 (river properties; 11.3% of variance); (C) PC factor 1 against PC factor 4 (geology; 8.3% of variance). Population locations on the spatial axes are marked by their code name and colors reflect population's regional groups as reported in Table 1.

was highly significant (P-value < 0.001) with an effect size of 0.980 (analogous to R^2). The distribution of environmental PC factors 1, 2, 4, and 7 Pearson's correlation coefficients for the 179 outlier markers used are shown in Figure 5. A higher frequency of correlation coefficients greater than 0.6 is found for PC factor 1, whereas an increasing frequency of lower coefficients generally characterized the PC factors with decreasing F-statistics values. The genomic distribution of F_{CT} values and PC factor 1 Pearson's correlation coefficients among all 3118 SNPs are shown in Figure 6. Overall, outlier markers are widely distributed among all linkage groups. Moreover, markers highly correlated with PC factor 1 are present on all linkage groups, whereas identified genetic outliers are often those with the highest correlation coefficient to PC factors on a given linkage group.

GENE ONTOLOGY AND SNP ANNOTATION

The BLAST and annotation steps in Blast2go yielded 1119 SNPs with annotations (Table S5). Using an FDR of 0.05, an enrichment analysis did not indicate significant over- or underrepresentation of any biological pathway among the 208 markers potentially under balancing selection. However, 12 GO-terms were overrepresented among the 179 markers potentially under divergent selection, which were associated with 12 SNPs (Table S6). Molecular functions, biological processes, and cell compartments associated with identified GO-terms suggested that these markers were associated with growth. The particular categories highlighted by the enrichment analysis include: positive regulation of JNK cascade, ephrin receptor binding, syndecan binding, frizzled binding. It should be noted that these categories were highlighted from a common hit of four markers for the syntenin-1 protein of Atlantic salmon (GenBank: ACI33400.1). We then plotted the relationship between population allele frequencies of these four SNPs against the environmental parameter with the highest loading on PC factor 1 (average temperature between May and September). All four models (generalized linear models) showed a similar significant regression (Fig. 7 shows one example regression). For PC factors 1 and 4, we then divided outliers into quartiles according to their correlation coefficients and used the fourth quartile (highly correlated markers) to perform enrichment analyses on the set of 179 outliers as reference. The test for temperature correlated markers identified similar functions and the same four markers reported above, which all had correlation coefficients greater than 0.70 with PC factor 1. However, no terms were overrepresented for geological-correlated markers for FDR = 0.05, whereas the enrichment was significant with a less stringent 0.05 P-value threshold.

Discussion

By surveying more than 3000 SNP markers widely distributed across the Atlantic salmon genome from 54 populations in combination with a thorough examination of 49 environmental factors in 23 North American rivers, we have been able to complete one of the most extensive landscape genomics analyses reported to date. The innovative statistical framework presented here demonstrated a very strong correlation between genetic and environmental structure characterized by significant associations between potentially adaptive divergence and climate. Geological parameters were also found to be important factors associated with potentially adaptive divergence. Our results suggest a regional component to local adaptation in Atlantic salmon that is associated with both climatic and geological factors. Furthermore, among markers potentially under divergent selection, we observed an enrichment of GO-terms associated with growth-related functions, suggesting

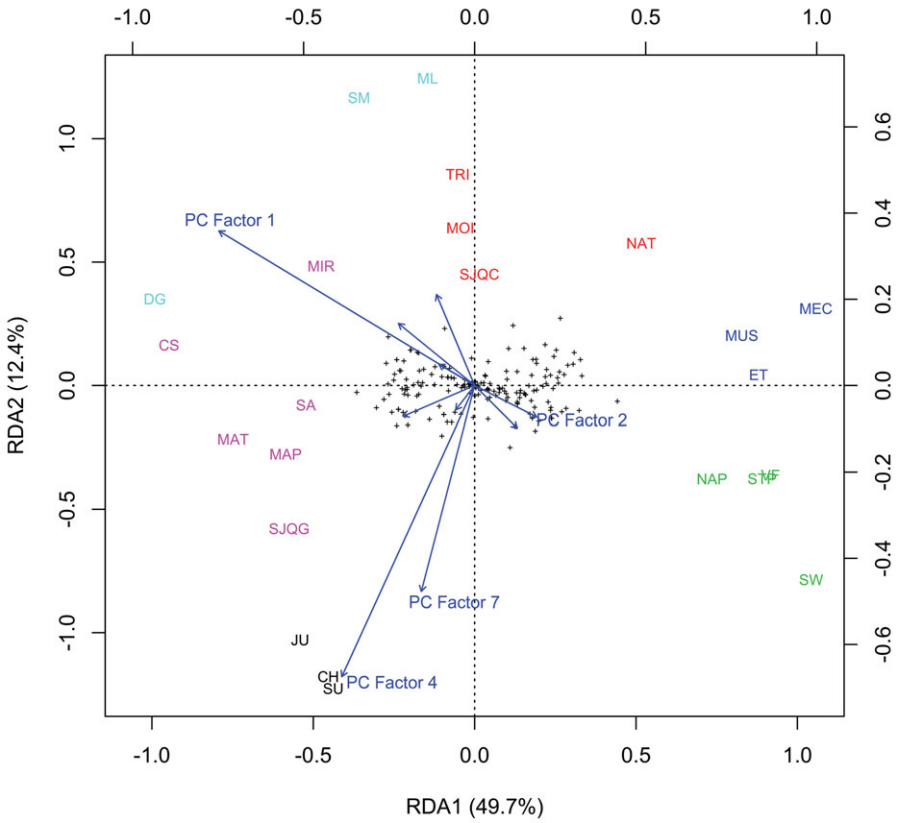

Figure 4. **Redundancy analysis axes 1 (49.7% of variance) and 2 (12.4% of variance) showing the position of allele frequency vectors for the 179 SNP markers potentially under divergent selection at the 0.05 significance level (plus marks) and related environmental PC factors as blue arrows. Only environmental PC factors significantly associated with genetic markers are identified (*P*-values < 0.001). Markers' positions relate to scales on the bottom and left axes. Environmental PC factors positions relate to scales on top and right axes. Population locations on the spatial axes are marked by their code name and colors reflect population's regional groups as reported in Table 1.**

a role for these biological functions in the adaptive divergence among populations and regional groups.

GENETIC DIVERGENCE

The first objective of this study was to revisit the population genetic structure of Atlantic salmon in this system with a new set of SNP markers and confirm whether the regional structure revealed with microsatellites was supported with SNP-array genotypes. Two results strongly suggest that the neutral genetic structure supported by SNP markers is similar to that of the microsatellite markers. First, pairwise F_{ST} estimates obtained for both types of markers are highly correlated (data not shown). Second, the distribution of populations along the first three principal components of the PCA shows a regional organization identical to the one reported in Dionne et al. (2008). Regional differences were such that by using the population allele frequencies of 2831 markers, we were able to build a powerful discriminant rule to classify bulk-assayed populations to their regional group. Out of 26 bulk assays for which populations had already been associated with a regional group in Dionne et al. (2008), 92% were concordantly assigned to the same regional group using our discriminant rule and SNPs. In the only two discordant populations, the Ouelle River population was originally classified as part of the Québec City region but included in the Southern Québec region with SNP markers. This river is actually located on the western limit of the Southern Québec group, which is bordered by the Québec City group. It thus seems plausible that this discordant regional assignment is at least partially due to an improved assignment power resulting from the increased number of markers. The Laval river population, however, was previously classified in the Higher North Shore regional group and is now assigned to the Southern Québec group, which is not geographically congruent. The only reasonable explanation we can offer at this stage is that the Laval, which is a small population (N < 100 individuals), might have been seriously affected by drift effects and occasional straying, causing this mixed signal in the population. However, two of the new populations analyzed with the bulk assays, Kecarpui and Des Escoumins, were assigned to regional groups concordant with

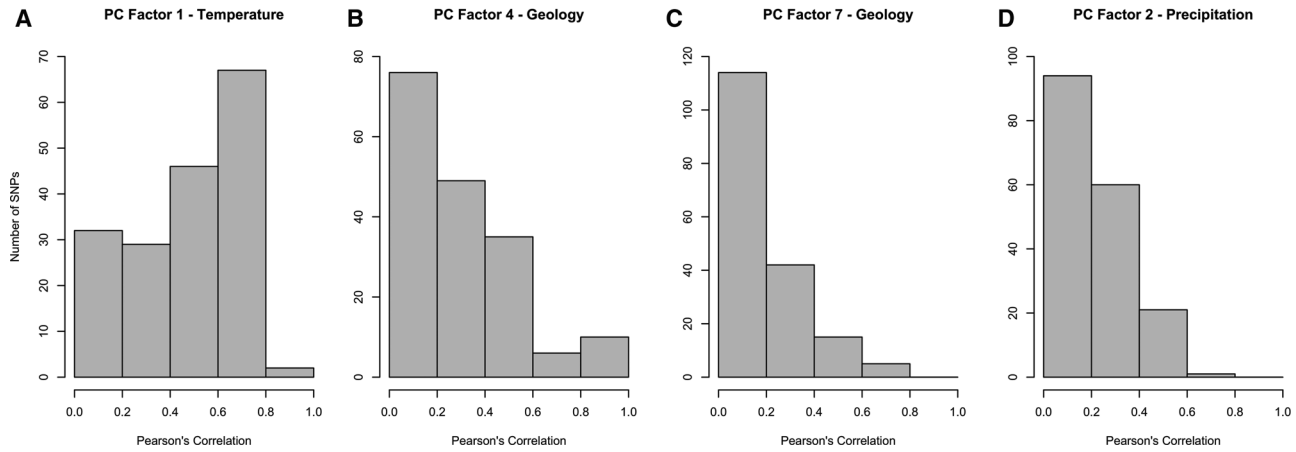

Figure 5. **Pearson's correlation coefficient distribution for the 179 SNP markers potentially under divergent selection at the 0.05 significance level when correlated with environmental PC factors significantly associated with genetic markers (*P*-values < 0.001). Presented in order of decreasing *F*-static value of analysis of variance from left to right where: (A) PC factor 1 (climate-temperature), (B) PC factor 4 (geology), (C) PC factor 7 (geology), and (D) PC factor 2 (climate-precipitation).**

geographical location. Overall, given the convincing regional assignment probabilities provided by the discriminant analysis, we argue that further interpretations concerning the regional genetic structure in the system could be extrapolated to bulk-assayed populations.

Apart from generally confirming the regional structure observed in Dionne et al. (2008), a second contribution of the new genomic dataset was the detection of 68 outliers potentially under divergent selection among populations. For this relatively small set of markers, the genetic variation attributed to differences among regions was more than four times the observed proportion for neutral SNPs (28.04% and 6.42%, respectively). Comparing indirect estimates of migration rates between and within regional groups, Dionne et al. (2008) previously hypothesized that local adaptation at the regional scale was driving higher population differentiations for intergroup rather than intragroup population comparisons even when the distance between populations was similar in either comparisons. This agrees with theoretical expectations that local selection in subdivided populations enhances between-deme genetic diversity (Charlesworth et al. 1997). It is also in agreement with our observation of an increased level of divergence between regions at outliers relative to neutral SNP markers but not between populations within a given regional group. In summary, our results provide further evidence for a possible role of selection in shaping regional population structure in Atlantic salmon.

GENETIC–ENVIRONMENT ASSOCIATIONS

Once the putative targets of divergent selection were identified from the genome scan, the second step was to identify the particular environmental variables acting as potential selective agents driving adaptive genetic divergence among regional groups. We found a strong association between the overall regional genetic groups and environmental structure. This indicated that these regional genetic groups also differed in ecological settings and that several environmental factors could represent selective agents leading to regional local adaptation. In particular, four PC factors were found to be strong predictors of genetic divergence. Temperature-related PC factor 1 and precipitation-related PC factor 2 were both climate factors correlated with genetic divergence, but displayed opposite direction vectors on the RDA axes, whereas geological-related PC factors 4 and 7 oriented in relatively similar directions to each other. Owing to the orthogonal relationship between climate and geological vectors on the RDA, we chose to discuss environmental factors as two main axes correlated with potentially adaptive divergence, namely climate and geology. Furthermore, because PC factor 2 was the fourth factor in decreasing order of statistical significance and negatively correlated with temperature on the RDA axes, we hereafter refer to temperature as a proxy for climate conditions.

Temperature regime was identified as the most important selective agent in the system, as this climatic factor has the highest *F*-statistic for PC factor 1 in the RDA and the highest correlation coefficient with outlier markers. Numerous studies have suggested that temperature regime is an important variable influencing local adaptation in Atlantic salmon (reviewed in Taylor 1991; Garcia de Leaniz et al. 2007) and that growth (Claytor et al. 1991; Nicieza et al. 1994a b; Paez et al. 2010) and immune-related functions (Dionne et al. 2007, 2009) could be important targets of local selection. Southern Atlantic salmon populations live in warmer conditions and are known to grow faster and migrate to sea at younger age (Power 1981; Metcalfe and Thorpe 1990), which could be linked to living in a more productive environment. However, Nicieza et al. (1994a, b) observed a significantly

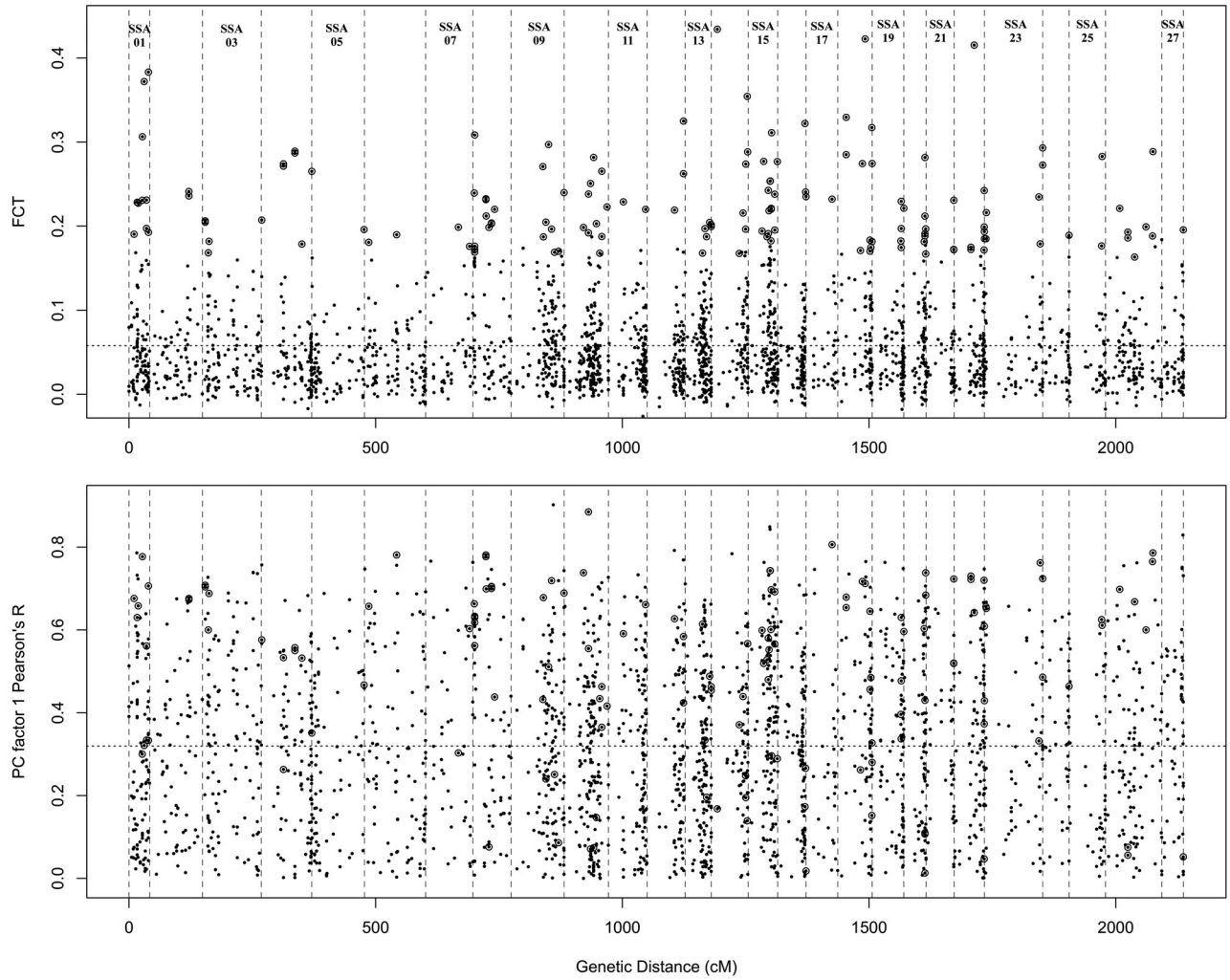

Figure 6. Genetic linkage map showing the distribution of regional differentiation (F_{CT}) of each SNP marker (n = 3118) on the top panel and Pearson's correlation coefficients (R) related to PC factor 1 (climate-temperature) on the bottom panel. Gray and white rectangles separated by vertical dashed lines represent linkage groups (named SSA). On the top panel, circled dots indicate outlier markers (significance level P = 0.05) and the horizontal dotted line indicates the average F_{CT} among markers (0.058). On the bottom panel, circled dots indicate outlier markers (significance level P = 0.05) and the horizontal dotted line indicates the average Pearson's correlation coefficient (R) among markers (0.320).

higher digestion and growth rate for salmon from higher latitude (Scotland) compared to salmon from southern latitudes (Spain) when reared at the same temperature. These results argue that countergradient selection has resulted in selection for more efficient growth in northern latitudes to compensate for a shorter growing season. Therefore, rather than local adaptation for high growth rate in the southern latitudes, the countergradient theory (reviewed in Conover and Schultz 1995) suggests that cold temperatures, a proxy for short growing season length, might actually be selecting for more efficient growth in higher latitudes. Furthermore, temperature regime was found to be closely related to bacterial diversity in the wild, which in turn was associated with genetic diversity of an immune-competence gene, the major histocompatibility complex class-IIB gene (Dionne et al. 2007, 2009). Genetic diversity at this locus is suggested to be involved in local adaptation of Atlantic salmon to different pathogen communities associated with different thermal regimes.

Geological parameters were, after climate, the environmental factors with the strongest links to genetic divergence. These were primarily loaded by the geological provinces category, geological periods associated with rock formations, and some specific rock or substrate types. All of these geological categories emphasized a division between populations from rivers draining either on the north or on the south shore of the St. Lawrence River, and supports the distinctiveness of Anticosti Island populations. Thus, rivers south of the St. Lawrence River and Anticosti

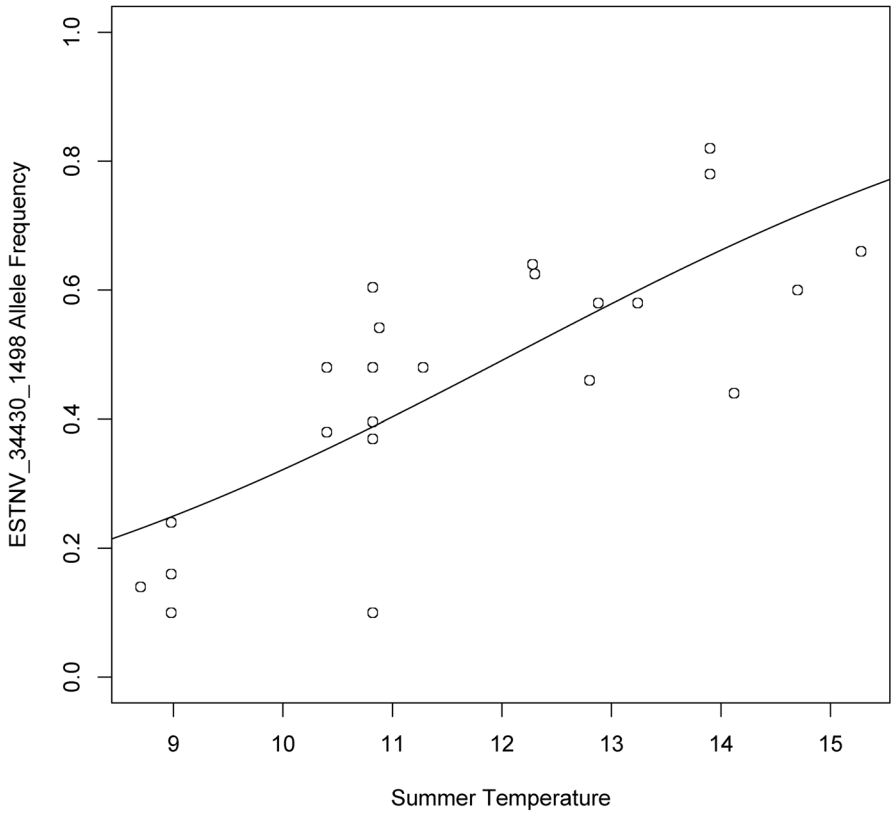

Figure 7. Generalized linear models illustrating the relation between marker population allele frequency and average seasonal temperatures between May and September.

belonged to the *Plate-forme du St-Laurent* and *Appalachian Orogen* geological provinces that are characterized by sedimentary rock formations dating to Silurian and Devonian periods of the Paleozoic era. In contrast, rivers draining on the north shore of the St. Lawrence River belong to the Grenville Province, which is mostly characterized by rock formation of gneiss type dating to the Mesoproterozoic period of the Precambrian era. Many factors can influence water chemistry, but aside from anthropogenic impacts, geology has been shown to be a dominant factor (Stallard and Edmond 1983; Johnson et al. 1997). Accordingly, the Southern Québec and Anticosti rivers are characterized by alkaline water. Because fish are surrounded by their environment, constant osmotic, ionic, and pH regulation are required to maintain homeostasis. Although water pH outside neutrality represents a stress for most fish, many species have adapted to alkaline or acidic waters (Pritchard 2003). We thus propose that regional specificity of geological parameters may interact with water chemistry of rivers to represent potential selective agents driving local adaptation in Atlantic salmon populations. Furthermore, as reported in Perrier et al. (2011), which found geological areas to significantly correlate with Atlantic salmon population genetic structure in France, geological substrate is suspected to be instrumental in the propensity of salmon to return to their natal river to spawn (Stabell 1984; Dittman et al. 1996). Geological characteristics of rivers may be an important factor influencing the accuracy of homing behavior and consequently reduce straying among regional groups, which would further contribute to maintain a regional component of local adaptation.

EXTENT OF LOCAL ADAPTATION

By studying climatic and geological factors, we now have evidence to argue for environmental selection driving adaptive divergence at the regional level. In fact, the strong regional genetic differentiation associated with distinct environmental features allowed us to use bulks assays to assign an additional 28 populations to regional groups without requiring us to perform individual genotyping. We emphasize that bulk assays provided additional indications for potential regional local adaptation and better defined the boundaries of this component in our system. The significant associations of temperature and geology with regional population genetic structure thus suggests these parameters are among the most important environmental selective agents delineating the geographical scale of local adaptation.

We observed a genome-wide distribution of divergent outliers and environmentally correlated markers, which is not unexpected given the diversity of putative environmental selective

agents identified. In fact, as discussed earlier, climate and geological characteristics can be used as proxies for many indirect ecological differences among populations from different regional groups, such as growing season length, pathogen diversity, and water chemistry. Thus, numerous targets of selection spread across the genome are more likely to emerge than a small number of targets in localized islands of adaptive divergence (Nosil et al. 2009). Theory predicts that local selection acting in a species with hierarchical population structure can lead to increased differentiation between demes. Furthermore, Charlesworth et al. (1997) found that under such conditions, local selection produced very high differentiation values for loci close to targets of selection but also high values for distant neutral loci with no selection. This is precisely what we observed when exploring the environmental correlation with either all 3118 SNPs, only neutral SNPs or with only highly divergent markers. The RDA systematically indicated an overall association between genetic and environmental divergence. Because the association was strongest when considering only the fewest outliers markers (rather than diminishing the value of the observed association), we argue that it reinforces the environmental selection arguments. Albeit weaker than those of putative targets of local adaptation, concordant differentiation patterns across the genome might be indicative of what Thibert-Plante and Hendry (2010) referred to as a generalized barrier to gene flow. Other recent studies, for instance Cooke et al. (2012), also reported a concordant but amplified genetic divergence when comparing potentially selected markers correlated with contrasting environmental conditions against the neutral genetic divergence of characin fish (*Triportheus albus*) in Amazonia. However, in some cases contrasting patterns of adaptive versus neutral divergence were detected due to a strong association between adaptive divergence and environmental condition not reflected in the neutral differentiation patterns (e.g., Gaggiotti et al. 2009; Bradbury et al. 2010; Lee and Mitchell-Olds 2011). In such cases, some outlier loci could also be linked to genetic incompatibilities revealing ancestral divergence rather than actual markers associated with selected genes to exogenous (environmental) factors (Bierne et al. 2011).

FUNCTIONAL IMPLICATIONS

Although interpretation of available sequence annotation should be made with caution, especially in an ecological context involving nonmodel species (Pavlidis et al. 2012; Pavey et al. 2012), they remain a useful tool for determining possible functional targets of selection, identifying candidate genes and framing hypotheses to link environmental selective agents and adaptive divergence at the genome level. The importance of temperature-related selection was predominant among the markers potentially associated with adaptive divergence. Indeed, the GO categories overrepresented among divergent outlier markers when compared to the 3118 markers were the same than those overrepresented among the outlier markers highly correlated with temperature (fourth quartile) when compared to the complete set of divergent outliers ($P = 0.05$). Enriched annotations were primarily associated with biological processes and functions linked to growth (e.g., regulation of JNK cascade, ephrin, and frizzled binding). As reported earlier, the sequences for markers revealing the most enriched GO categories blasted to a syntenin-1 sequence from Atlantic salmon. This protein functions as a binding protein for syndecan, a transmembrane domain protein with an important growth-factor-receptor activation function for one of its four forms (Carey et al. 1997). Given the regional differentiation at these SNP markers located in growth-related genes, we propose that they might bear a signature of thermal local adaptation linked to countergradient selection imposed by growth season length. Thus, thermal regimes may act as a selective agent driving local adaptation on growth potential. This hypothesis should be tested by measuring the impact of outlier allelic variants on enzyme function and growth rate.

Conclusion

In summary, this study offers new insights on the geographic and genomic extent of local adaptation in Atlantic salmon by combining population genomics with landscape genetics. As in any landscape genetics study, the relationships between environmental conditions and genetic divergence presented are correlational and not necessarily causal, but nevertheless provide testable hypotheses for future studies, such as reciprocal transplants or specific genotypes impacts on enzyme function, growth, or range-wide fitness measurements in wild Atlantic salmon. Owing to a wide geographic and genome coverage, we were able to rigorously confirm a hierarchical genetic structure, and found a strong regional component of both neutral and potentially adaptive divergence among the 54 populations we studied. We also found that this regional genetic structure was significantly correlated with an ecological structure described by a set of 49 environmental parameters. We found specific associations between environmental factors related to climate (temperature) and geology with markers potentially under divergent selection. This allowed us to propose putative environmental selective agents and candidate genes potentially involved in the process of local adaptation in Atlantic salmon. We also found that markers potentially under divergent selection were distributed throughout the genome. Finally, we were able to use annotations to infer a plausible causal link for environmental selection associated with growth-related functions.

Although we reported very strong support for a regional component of local adaptation at the geographic level, we lacked the genomic coverage to investigate the extent of adaptive hitchhiking surrounding our genome-wide potential targets of

selection at the genomic level. An enhanced genomic coverage could allow the investigation of precise genomic location potentially under the effect of divergent selection and document more precisely the location and size of islands of adaptive divergence (Feder and Nosil 2010). Finally, although our results call for experimental confirmation of the adaptive hypotheses that they propose, this study illustrates how landscape population genomics contribute to improve our knowledge of the evolutionary processes affecting populations and may help develop conservation tools integrating both genetic and environmental parameters and their interactions (Funk et al. 2012).

ACKNOWLEDGMENTS

The authors are grateful to G. Daigle for incredible support and insights on statistical analyses, B. Landry and technical assistance from the Ministère du Développement durable, de l'Environnement, de la Faune et des Parcs du Québec (MDDEFP) in the identification of available environmental information, A. C. Dalziel for language editing and three anonymous reviewers for their comments on the original manuscript. The SNP discovery, array development, and genotyping were performed by Centre for Integrative Genetics 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 in close collaboration with the MRNF represented by MD and the Department of Fisheries and Oceans Canada (DFO). 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. The authors declare no conflict of interest.

LITERATURE CITED

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate—a practical and powerful approach to multiple testing. J. R. Stat. Soc. 57:289–300.

Bierne, N., J. Welch, E. Loire, F. Bonhomme, and P. David. 2011. The coupling hypothesis: why genome scans may fail to map local adaptation genes. Mol. Ecol. 20:2044–2072.

Bonin, A. 2008. Population genomics: a new generation of genome scans to bridge the gap with functional genomics. Mol. Ecol. 17:3583–3584.

Bonin, A., P. Taberlet, C. Miaud, and F. Pompanon. 2006. Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). Mol. Biol. Evol. 23:773–783.

Bourret, V., M. P. Kent, C. R. Primmer, A. Vasemägi, S. Karlsson, K. Hindar, P. McGinnity, E. Verspoor, L. Bernatchez, and S. Lien. 2013. SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). Mol. Ecol. 22:532–551.

Bradbury, I. R., S. Hubert, B. Higgins, T. Borza, S. Bowman, I. G. Paterson, P. V. R. Snelgrove, C. J. Morris, R. S. Gregory, D. C. Hardie, et al. 2010. Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. Proc. R. Soc. B 277:3725–3734.

Brenna-Hansen, S., J. Li, M. P. Kent, E. G. Boulding, S. Dominik, W. S. Davidson, and S. Lien. 2012. Chromosomal differences between European and North American Atlantic salmon discovered by linkage mapping and supported by fluorescence in situ hybridization analysis. BMC Genom. 13, doi:10.1186/1471-2164-1113-1432.

Carey, D. J. 1997. Syndecans: multifunctional cell-surface co-receptors. Biochem. J. 327:1–16.

Charlesworth, B., M. Nordborg, and D. Charlesworth. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res. 70:155–174.

Claytor, R. R., H. R. Maccrimmon, and B. L. Gots. 1991. Continental and ecological variance-components of European and North-American Atlantic salmon (*Salmo salar*) phenotypes. Biol. J. Linn. Soc. 44:203–229.

Conover, D. O., and E. T. Schultz. 1995. Phenotypic similarity and the evolutionary significance of countergradient variation. Trends Ecol. Evol. 10:248–252.

Cooke, G. M., N. L. Chao, and L. B. Beheregaray. 2012. Divergent natural selection with gene flow along major environmental gradients in Amazonia: insights from genome scans, population genetics and phylogeography of the characin fish *Triportheus albus*. Mol. Ecol. 21:2410–2427.

Craig, J. E., A. W. Hewitt, A. E. McMellon, A. K. Henders, L. Ma, L. Wallace, S. Sharma, K. P. Burdon, P. M. Visscher, G. W. Montgomery, et al. 2009. Rapid inexpensive genome-wide association using pooled whole blood. Genome Res. 19:2075–2080.

Dionne, M., K. M. Miller, J. J. Dodson, F. Caron, and L. Bernatchez. 2007. Clinal variation in mhc diversity with temperature: evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. Evolution 61:2154–2164.

Dionne, M., F. Caron, J. J. Dodson, and L. Bernatchez. 2008. Landscape genetics and hierarchical genetic structure in Atlantic salmon: the interaction of gene flow and local adaptation. Mol. Ecol. 17:2382–2396.

Dionne, M., K. M. Miller, J. J. Dodson, and L. Bernatchez. 2009. MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. Philos. Trans. R. Soc. B 364:1555–1565.

Dittman, A. H., T. P. Quinn, and G. A. Nevitt. 1996. Timing of imprinting to natural and artificial odors by coho salmon (*Oncorhynchus kisutch*). Can. J. Fish. Aquat. Sci. 53:434–442.

Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10:564–567.

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

Feder, J. L., and P. Nosil. 2010. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. Evolution 64:1729–1747.

Fraser, D. J., L. K. Weir, L. Bernatchez, M. M. Hansen, and E. B. Taylor. 2011. Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. Heredity 106:404–420.

Funk, W. C., J. K. McKay, P. A. Hohenlohe, and F. W. Allendorf. 2012. Harnessing genomics for delineating conservation units. Trends Ecol. Evol. 27:489–496.

Gaggiotti, O. E., D. Bekkevold, H. B. H. Jorgensen, M. Foll, G. R. Carvalho, C. Andre, and D. E. Ruzzante. 2009. Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study. Evolution 63:2939–2951.

Garcia de Leaniz, C., I. A. Fleming, S. Einum, E. Verspoor, W. C. Jordan, S. Consuegra, N. Aubin-Horth, D. Lajus, B. H. Letcher, A. F. Youngson, et al. 2007. A critical review of adaptive genetic variation in Atlantic salmon: implications for conservation. Biol. Rev. 82:173–211.

Gotz, S., J. M. Garcia-Gomez, J. Terol, T. D. Williams, S. H. Nagaraj, M. J. Nueda, M. Robles, M. Talon, J. Dopazo, and A. Conesa. 2008.

High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36:3420–3435.

Hansen, M. M., D. E. Ruzzante, E. E. Nielsen, D. Bekkevold, and K. L. D. Mensberg. 2002. Long-term effective population sizes, temporal stability of genetic composition and potential for local adaptation in anadromous brown trout (*Salmo trutta*) populations. Mol. Ecol. 11:2523–2535.

Johnson, L. B., C. Richards, G. E. Host, and J. W. Arthur. 1997. Landscape influences on water chemistry in Midwestern stream ecosystems. Freshwater Biol. 37:193–208.

Kawecki, T. J., and D. Ebert. 2004. Conceptual issues in local adaptation. Ecol. Lett. 7:1225–1241.

King, T. L., E. Verspoor, A. P. Spidle, R. Gross, R. B. Phillips, M. L. Koljonen, J. A. Sanchez, and C. L. Morrison. 2007. Biodiversity and population structure. Pp. 117–166 *in* E. Verspoor, L. Stradmeyer, and J. L. Nielsen, eds. The Atlantic Salmon–genetics, conservation and management. Blackwell Publishing, Oxford, U.K.

Lacy, R. C. 1997. Importance of genetic variation to the viability of mammalian populations. J. Mammal. 78:320–335.

Leclerc, E., Y. Mailhot, M. Mingelbier, and L. Bernatchez. 2008. The landscape genetics of yellow perch (*Perca flavescens*) in a large fluvial ecosystem. Mol. Ecol. 17:1702–1717.

Lee, C.-R., and T. Mitchell-Olds. 2011. Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. Mol. Ecol. 20:4631–4642.

Macgregor, S., Z. Z. Zhao, A. Henders, N. G. Martin, G. W. Montgomery, and P. M. Visscher. 2008. Highly cost-efficient genome-wide association studies using DNA pools and dense SNP arrays. Nucleic Acids Res. 36:e35.

Manel, S., M. K. Schwartz, G. Luikart, and P. Taberlet. 2003. Landscape genetics: combining landscape ecology and population genetics. Trends Ecol. Evol. 18:189–197.

Manel, S., S. Joost, B. K. Epperson, R. Holderegger, A. Storfer, M. S. Rosenberg, K. T. Scribner, A. Bonin, and M.-J. Fortin. 2010a. Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. Mol. Ecol. 19:3760–3772.

Manel, S., B. N. Poncet, P. Legendre, F. Gugerli, and R. Holderegger. 2010b. Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. Mol. Ecol. 19:3824–3835.

Metcalfe, N. B., and J. E. Thorpe. 1990. Determinants of geographical variation in the age of seaward migrating salmon, *Salmo salar*. J. Anim. Ecol. 59:135–145.

Nicieza, A. G., L. Reiriz, and F. Brana. 1994a. Variation in digestive performance between geographically disjunct populations of Atlantic salmon—countergradient in passage time and digestion rate. Oecologia 99:243–251.

Nicieza, A. G., F. G. Reyesgavilan, and F. Brana. 1994b. Differentiation in juvenile growth and bimodality patterns between northern and southern-populations of Atlantic salmon (*Salmo salar* L). Can. J. Zool. 72:1603–1610.

Nosil, P., D. J. Funk, and D. Ortiz-Barrientos. 2009. Divergent selection and heterogeneous genomic divergence. Mol. Ecol. 18:375–402.

Paez, D. J., M. Morrissey, L. Bernatchez, and J. J. Dodson. 2010. The genetic basis of early-life morphological traits and their relation to alternative male reproductive tactics in Atlantic salmon. J. Evol. Biol. 23:757–768.

Palstra, F. P., M. F. O'Connell, and D. E. Ruzzante. 2007. Population structure and gene flow reversals in Atlantic salmon (*Salmo salar*) over contemporary and long-term temporal scales: effects of population size and life history. Mol. Ecol. 16:4504–4522.

Parisod, C., and R. Holderegger. 2012. Adaptive landscape genetics: pitfalls and benefits. Mol. Ecol. 21:3644–3646.

Patterson, N., A. L. Price, and D. Reich. 2006. Population structure and eigenanalysis. PLoS Genet. 2:2074–2093.

Pavey, S. A., and J. J. Dodson. 2012. At the interface of behaviour, ecology and evolution: Insights from the world of fishes. Current Zool. 58.

Pavlidis, P., J. D. Jensen, S. Wolfgang, and A. Stamatakis. 2012. A critical assessment of storytelling: Gene ontology categories and the importance of validating genomic scans. Mol. Biol. Evol. 29:3237–3248.

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

Petren, K., P. R. Grant, B. R. Grant, and L. F. Keller. 2005. Comparative landscape genetics and the adaptive radiation of Darwin's finches: the role of peripheral isolation. Mol. Ecol. 14:2943–2957.

Power, G. 1981. Stock characteristics and catches of Atlantic salmon (*Salmo salar*) in Québec, and Newfoundland and Labrador in relation to environmental variables. Can. J. Fish. Aquat. Sci. 38:1601–1611.

Pritchard, J. B. 2003. The gill and homeostasis: transport under stress. Am. J. Physiol.-Reg. I. 285:R1269–R1271.

Stabell, O. B. 1984. Homing and olfaction in salmonids—a critical review with special reference to the Atlantic salmon. Biol. Rev. Camb. Philos. 59:333–388.

Stallard, R. F., and J. M. Edmond. 1983. Geochemistry of the Amazon 2. The influence of geology and weathering environment on the dissolved-load. J. Geophys. Res.-Oc. Atm. 88:9671–9688.

Taylor, E. B. 1991. A review of local adaptation in salmonidae, with particular reference to Pacific and Atlantic salmon. Aquaculture 98:185–207.

Thibert-Plante, X., and A. P. Hendry. 2010. When can ecological speciation be detected with neutral loci? Mol. Ecol. 19:2301–2314.

Tonteri, A., A. J. Veselov, A. V. Zubchenko, J. Lumme, and C. R. Primmer. 2009. Microsatellites reveal clear genetic boundaries among Atlantic salmon (*Salmo salar*) populations from the Barents and White seas, northwest Russia. Can. J. Fish. Aquat. Sci. 66:717–735.

Vähä, J.-P., J. Erkinaro, E. Niemela, and C. R. Primmer. 2007. Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. Mol. Ecol. 16:2638–2654.

Vasemägi, A., J. Nilsson, and C. R. Primmer. 2005. Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). Mol. Biol. Evol. 22:1067–1076.

Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population-structure. Evolution 38:1358–1370.

Yeomans, K. A., and P. A. Golder. 1982. The Guttman-Kaiser criterion as a predictor of the number of common factors. Statistician 31: 221–229.

Associate Editor: K. Petren

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Figure S1. Within population allele frequencies estimated from bulk assays for 2831 SNPs on the x-axis and corresponding allele frequencies for the same populations ($n = 5$) as measured by individual genotyping of the same individual ($n = 25$ individuals per population).

Table S1. Single nucleotide polymorphism (SNP) markers' categories.

Table S2. Single nucleotide polymorphism (SNP) markers observed (H_O) and expected (H_E) heterozygosities per population.

Table S3. Summary of the detection of markers potentially under selection using hierarchical F_{dist} genome scans implemented in ARLEQUIN 3.5 (Excoffier and Lischer 2009).

Table S4. Pairwise measures of genetic differentiation (F_{ST}).

Table S5. Blast results from BLAST2GO and gene ontology (GO) terms annotation.

Table S6. Enrichment analysis results testing over- or under-representation of gene ontology (GO) terms.