MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 4444–4457 doi: 10.1111/mec.12798

FROM THE COVER

Detecting genotypic changes associated with selective mortality at sea in Atlantic salmon: polygenic multilocus analysis surpasses genome scan

VINCENT BOURRET,* MÉLANIE DIONNE† and LOUIS BERNATCHEZ*

Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, 1030 avenue de la Médecine, Québec, Québec, Canada G1V 0A6, †Service de la faune aquatique, Ministère des Forêts, de la Faune et des Parcs du Québec, Québec, Canada G1S 4X4

Abstract

Wild populations of Atlantic salmon have declined worldwide. While the causes for this decline may be complex and numerous, increased mortality at sea is predicted to be one of the major contributing factors. Examining the potential changes occurring in the genome-wide composition of populations during this migration has the potential to tease apart some of the factors influencing marine mortality. Here, we genotyped 5568 SNPs in Atlantic salmon populations representing two distinct regional genetic groups and across two cohorts to test for differential allelic and genotypic frequencies between juveniles (smolts) migrating to sea and adults (grilses) returning to freshwater after 1 year at sea. Given the complexity of the traits potentially associated with sea mortality, we contrasted the outcomes of a single-locus F_{ST} based genome scan method with a new multilocus framework to test for genetically based differential mortality at sea. While numerous outliers were identified by the single-locus analysis, no evidence for parallel, temporally repeated selection was found. In contrast, the multilocus approach detected repeated patterns of selection for a multilocus group of 34 covarying SNPs in one of the two populations. No significant pattern of selective mortality was detected in the other population, suggesting different causes of mortality among populations. These results first support the hypothesis that selection mainly causes small changes in allele frequencies among many covarying loci rather than a small number of changes in loci with large effects. They also point out that moving away from the a strict 'selective sweep paradigm' towards a multilocus genetics framework may be a more useful approach for studying the genomic signatures of natural selection on complex traits in wild populations.

Keywords: conservation, local adaptation, mortality at sea, salmonids, single nucleotide polymorphism

Received 21 January 2014; revision received 14 May 2014; accepted 15 May 2014

Introduction

In the past decades, we have witnessed a range-wide decline in wild Atlantic salmon populations. Despite the ban on commercial fishing, the majority of natural stocks are still declining. In North America, this decline has been characterized by a remarkable reduction in marine survival (Gibson *et al.* 2011; ICES 2013). Habitat degradation, the impacts of the salmon farming industry and climatic changes are commonly invoked to explain much of the decline in the species as a whole (Hansen & Quinn 1998). The Atlantic salmon is an anadromous species that usually spends one or more years in freshwater before the juveniles physiologically prepare for seaward migration via the process of smoltification and then migrate towards their marine feeding grounds off the coast of Newfoundland-Labrador (for

Correspondence: Vincent Bourret, Fax: 418-656-7176; E-mail: vincent.bourret.1@ulaval.ca

North American populations) and Greenland (for both North American and European populations). After one or more years at sea, adults return to reproduce in their natal river (Stabell 1984). Studying Atlantic salmon during the marine phase is critical for determining the reasons for population declines, but has proven very difficult. Therefore, we still know very little about this species' life in the marine environment. However, multiple monitoring studies have shown a sharp decline in sea survival indexes, suggesting that mortality at this life history stage is contributing to the observed population decline and needs to be better studied (Dionne *et al.* 2013; ICES 2013).

The marine environment is challenging to salmon for a number of reasons, such as a higher risk of predation and the need to undertake a long distance migration. In fact, the seaward migration of Atlantic salmon is one of the most deadly phases of this species' life cycle as the bulk of mortality is thought to occur when entering the marine environment (Friedland *et al.* 2000; Thorstad *et al.* 2012). Differential survival at sea is most likely to involve a complex variety of traits associated with predator avoidance, osmoregulation, growth and even disease resistance as reported in previous studies on migratory salmon (Miller *et al.* 2011). Moreover, the accentuated decline over the last 30 years suggests that unusual or new selective forces may be acting more strongly on Atlantic salmon relative to the past. In addition, variable return rates vary among salmon from different geographical regions suggests differential mortality at sea among distinct populations (Dionne *et al.* 2013). Investigating genomic changes occurring during the marine phase of salmon may help to determine whether reduced survival at sea is occurring randomly or is due to specific selective pressures.

From a fundamental perspective, disentangling neutral and selective changes occurring at the genome level is a crucial yet challenging goal in evolutionary biology. When it comes to identifying the specific targets of selection, some methods have received much criticism. Nevertheless, despite the numerous pitfalls and maladapted models extensively discussed in recent years (Bierne *et al.* 2011; Narum & Hess 2011; Vilas *et al.* 2012; De Mita *et al.* 2013; Lotterhos & Whitlock 2014), methods based on the identification of markers with unusually high levels of genetic differentiation (e.g. genome scans) have proved to be successful in many situations. On the other hand, as reported in a growing number of quantitative genetic studies (Mackay *et al.* 2009; Hancock *et al.* 2010; Yang *et al.* 2010; Nadeau & Jiggins 2011; Daub *et al.* 2013), outlier detection is essentially restricted to the investigation of single loci or genomic regions of large effects. However, as selection is expected to generate linkage disequilibrium between loci for complex quantitative traits, relatively low and subtle allelic changes among covarying loci is expected to yield a combined effect greater than individual loci on phenotypes (Latta 1998; Le Corre & Kremer 2003). Therefore, it might be time for a paradigm shift from the quest for a strong selective sweep on individual loci to the consideration of polygenic soft sweeps when investigating potential targets of selection at the genomic level (Le Corre & Kremer 2012).

Although theory predicts that a beneficial new allele or standing genetic variation can be swept by a rapid shift in frequency caused by selection, Pritchard & Di Rienzo (2010) have argued that reducing adaptation to 'selective sweeps at key loci is too limited'. Obviously, there is a giant leap between models based upon selective sweeps and those which model polygenic adaptation, and fortunately, they are not mutually exclusive as both scenarios are probably occurring in nature. However, regardless of the traits investigated, recent studies have relied on single-locus genome scans to find selective sweeps, even if that means surfacing with a large number of false-positive targets. While selective sweep models might be effective for simple traits, quantitative genetic theory makes a strong case for polygenic adaptation in complex traits, as best exemplified by the genetic study of human height (Yang *et al.* 2010).

In this study, our main goal was to test expectations of single-locus and multilocus models on a complex trait such as marine survival that is predicted to be under recently occurring, novel selective pressures caused by recent environmental changes in the Atlantic Ocean (Mills *et al.* 2013). Over the past 30 years, two rivers in the Canadian province of Québec have been annually monitored to evaluate the evolution of intrinsic characteristics, abundance and rates of freshwater survival and adult returns in Atlantic salmon (Dionne *et al.* 2013). Every year, young Atlantic salmon leaving the rivers for the first time (called smolts) and adults returning after one (called grilse) or more years at sea (called multisea winter) are captured. This system offers a unique opportunity to follow spatially and temporally replicated cohorts of Atlantic salmon and to examine the genetic changes occurring during the marine phase of their life cycle. Here, we take advantage of this unique system to investigate the genomic changes occurring due to natural selection during the first year at sea in Atlantic salmon. This was achieved by contrasting single-locus F_{ST} based methods and a customized polygenic multilocus approach to identifying genomic targets of selection and test for spatial and temporal variation in patterns of selection.

Materials and methods

Samples

Fin clips were collected and preserved in 95% ethanol for 25 Atlantic salmon smolts leaving freshwater for the sea. Smolts were sampled across the peak of outmigration from freshwater to the marine environment in May and June of 2004 and 2005 in the Saint-Jean River, in Gaspésie, Québec, Canada and de la Trinité River, in the North Shore region of Québec, Canada. In addition to these 100 smolts, 100 fins clips were then collected from 25 adult Atlantic salmon returning after one year at sea (hereafter named grilses) in 2005 and 2006 from both rivers (Fig. 1). These two rivers represent two previously identified regional genetic groups of salmon (Dionne *et al.* 2008) shown to be potentially locally adapted to distinct environmental conditions (Bourret *et al.* 2013b). Also, different return rates from the sea have been reported between salmon from these two regions, indicating that they suffer differential mortality at sea, which may indicate differences in the strength or identity of selective forces (Dionne *et al.* 2013). Smolts were captured using rotary traps at 8 and 9 km upstream of the river mouth in the Saint-Jean and Trinité rivers respectively. Adults were collected over the entire return migration period, from local anglers in the Saint-Jean River and from a fish ladder located at the river mouth in the Trinité River. For all samples total length, fork length and weight were measured, and scales were also collected to determine age, although none of these parameters differed significantly between rivers (results not shown). DNA was extracted from fin clips using the QIAGEN DNeasy Tissue Kit following the manufacturer's guidelines (Qiagen, Valencia, CA, USA). Samples were sexed using a molecular sexing technique (Yano *et al.* 2012).

Genotyping quality control

Genotyping for all 200 DNA samples was performed on 5568 SNPs with the SNP-array developed by the Centre for Integrative Genetics (CIGENE, Norway) and reagents from the Illumina infinium assay (Illumina, San Diego, CA, USA) following the manufacturer's instructions. Detailed methods for SNP discovery and quality control can be found in Bourret *et al.* (2013a, b). The quality of individual samples was assured by only using individuals genotyped at a greater than 0.85 call rate (*CR*: proportion of genotyped SNPs). Markers with minor allele frequencies less than 1% (MAF < 0.01) and markers missing in more than 1% of individuals were excluded from our analyses. Ascertainment bias was assessed by Bourret *et al.* (2013a) and was shown to be minimal when comparing North American populations.

Signatures of selection

Single-locus genome scan. To contrast single-locus and multilocus approaches for detecting selection during life at sea in Atlantic salmon grilses, we first performed a single-locus F_{ST} outlier approach using LOSITAN (Antao *et al.* 2008). LOSITAN is a selection detection workbench based on the FdistE method of Beaumont &

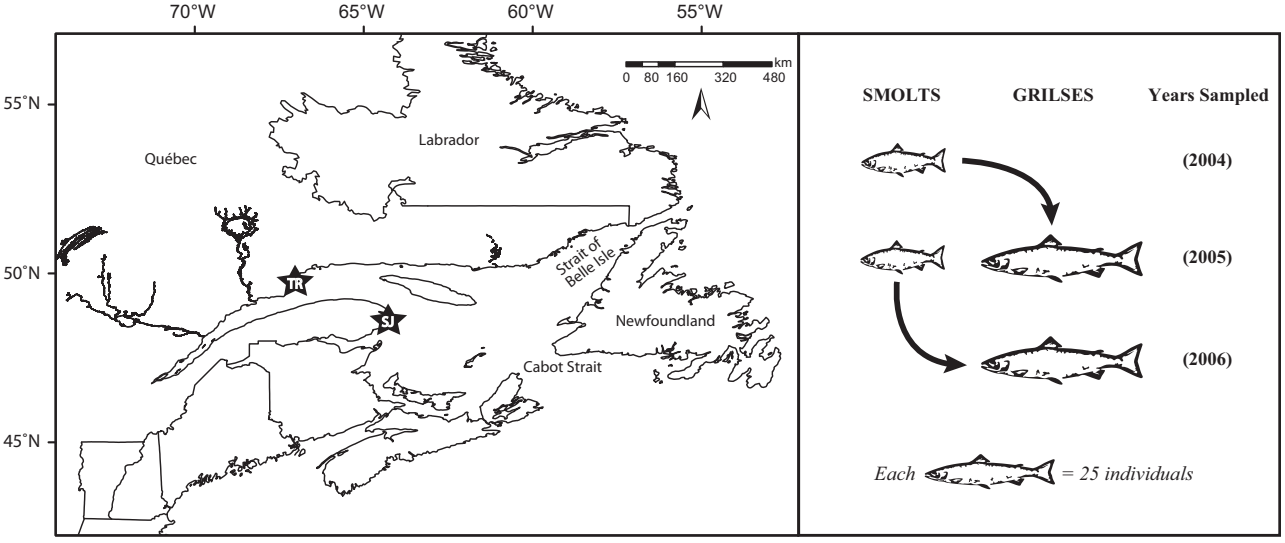

Fig. 1 Map showing sample sites on the left panel: Trinité River (TR) and Saint-Jean River (SJ). The right panel is a schematic representation of the experimental design in each river. Twenty-five individuals were sampled for each life stages in each cohort in both rivers for a total of 200 individuals.

Nichols (1996). For each cohort from each river, we identified F_{ST} outliers between smolts and grilses for a total of four runs of the program. Each run was carried out using the recommended calculation of 'Neutral mean F_{ST}', which removes potentially selected loci from the calculation of the mean F_{ST} after a first run of simulation and using the 'Forcing mean F_{ST}' option. We used 50 000 simulations, at a significance level of 0.01. Markers identified as potentially under divergent selection and with a reported heterozygosity >0.1 were then compared between cohorts in a given river, between rivers in a given cohort and overall to detect any significant parallel patterns of selection between cohorts and/or rivers for nonrandom selective mortality at sea.

Polygenic multilocus approach. Here, the number of individual markers was reduced into sets of independently, covarying groups of markers based on their genotypic distribution. For each marker, we coded individual genotypes as 0, 1 and 2, which referred to the count of one allele for each SNP. A principal component analysis (PCA) was first performed on the retained set of markers (see Results) for all 200 individuals using the PROC FACTOR and VARIMAX rotation procedure implemented in SAS 9.3 (SAS Institute Inc). The set of markers was then reduced to a number of factors identified using the parallel analysis (PA) criterion (Horn 1965). The PA criterion is a Monte-Carlo based simulation method that compares the observed eigenvalues with those obtained from uncorrelated normal variables. It is based on the rationale that eigenvectors from the data set should have larger eigenvalues than parallel eigenvectors from random data sets with the same sample size and number of variables (Ford *et al.* 1986; Lautenschlager 1989). A factor or component was retained if the associated eigenvalue was bigger than the 99th of the distribution of eigenvalues derived from a random data. The PA was then performed on 1000 random data sets (bootstraps) and a significance level of 0.01 to determine the number of factors to be retained in subsequent analyses. We could then determine the specific markers significantly correlated with a given principal component using a conservative significance level of 0.001, which corresponds to an absolute loading weight superior to 0.23 on a given factor.

In the second step, a canonical discriminant analysis based on the retained factor's scores was performed to examine the power to distinguish individuals from rivers, cohorts and life stages using the PROC CAN procedure implemented in SAS 9.3. For statistically significant canonical axes, we then performed a generalized linear model to identify the sources of discrimination using the PROC GLM procedure in SAS 9.3. The model considered rivers, cohorts, life stages, interactions among rivers and cohorts, and interactions among rivers and life stages as possible sources of correlation with the canonical axis. We then determined the PCA factors significantly correlated with either discriminant canonical axis using a conservative significance level of 0.001, which corresponds to an absolute loading weight superior to 0.23 on a given factor, as mentioned above. These factors thus represented independent multilocus factors potentially under selection according to the source of variation.

Covariance of allelic effects

Based on quantitative genetics theory, we predicted positive covariance (i.e. linkage disequilibrium) to have built up across loci under selection and associated with survival at sea (Latta 1998; Le Corre & Kremer 2003). We therefore estimated the covariance of allelic effects across populations and within each population for all pairwise combinations of SNPs. The covariance of allelic effects was calculated based on Equation 2 of Ma *et al.* (2010): $4a_ia_jD_{ij}$; where a_i and a_j are the additive effects of loci i and j, respectively and D_{ij} is the linkage disequilibrium (LD) between the two loci. As a surrogate of the scale of additive effects for each locus on survival at sea, we used the absolute loading weight on identified PC factors correlated with the canonical axis, which presented life stage as a source of discrimination among individuals (see Results). LD was calculated using the locus pairwise multiallelic D in POWERMARKER (Liu & Muse 2005). We then compared the distribution of allelic effect covariance across the two populations for loci loading on significantly correlated PC factors potentially under selection to the distribution of mean covariance of allelic effect for a similar number of markers randomly chosen among all markers 1000 times. All of these analyses were performed in R (R Development Core Team).

Strength of selection

As a surrogate for fitness estimates, we thus used delta p (Δp), which is the absolute difference in frequency of the major allele between smolts and grilses. The allele frequency change between life stages can be a direct response to differential survival, and in such as case, measuring Δp for a given marker is related to estimating its fitness values (Barrett *et al.* 2008). Therefore, to measure the strength of selection on a targeted group of markers we compared the distribution and mean of Δp for loci identified with the multilocus approach to those loci identified as outliers in the single-locus outlier detection and the distribution for retained markers. Those comparisons were made by considering both rivers together and separately as well as cohorts pooled

within each river. To localize potential genomic regions associated with mortality at sea, we used a linkage map of North American Atlantic salmon (Brenna-Hansen *et al*. 2012) to position the distribution of Δp as well as the loading weights of all markers according to their chromosomal distances on the 27 Atlantic salmon linkage groups. These analyses were all performed in R (R Development Core Team).

Gene ontology and SNP annotation

BLAST2GO (Gotz *et al*. 2008) was used to associate gene ontology (GO) annotation terms to all SNPs. Homology search was first conducted with 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}. To determine if the biological processes, molecular functions or cellular components were over-, equally or under-represented among outlier markers or SNPs reported on a multilocus group potentially under selection, we performed enrichment analyses using Fisher's exact test corrected for multiple tests by applying a false discovery rate of 0.05 (FDR; Benjamini & Hochberg 1995).

Results

Genotyping and quality control

All individuals had call rates >95% so all samples were included in our analyses. After initial quality control of genotypes, we identified 3380 polymorphic and reliable SNPs of which 457 were discarded because they showed an overall MAF < 0.01 or were missing genotypes at a proportion greater than 0.01. Therefore, from 5568 SNPs on the SNP-array, 2923 SNPs from all 200 individuals were used in our analyses.

Signatures of selection

Single-locus genome scan. Many outliers were identified in each of the pairwise comparisons between smolts and grilse for each cohort in each river (Table S1, Supporting information). However, we found no evidence for parallelism among them. In the Saint-Jean River, 34 SNPs were identified as outliers in the 2004 cohort (F_{ST} range: 0.082–0.231) and 42 outliers were identified in the 2005 cohort (F_{ST} range: 0.093–0.198). Among those outliers potentially under divergent selection, only one (GCR_cBin9358_Ctg1_120) was found in both 2004 and 2005 cohorts. In the Trinité River, 34 (2004; F_{ST} range: 0.085–0.180) and 20 (2005; F_{ST} range: 0.096–0.184) SNPs were identified as outliers and none were the same in both cohorts. Two markers (ESTNV_22435_210, ESTV_16844_203) out of 64 were found to be parallel divergent outliers in the 2004 cohort between the Saint-Jean and Trinité populations, whereas one SNP (ESTNV_34235_752) out of 62 was parallel between both populations for the 2005 cohort when comparing smolts and grilses. No outliers were common to more than two comparisons. A summary of all four F_{ST} based genome scans is presented in Fig. 2.

Polygenic multilocus approach. Based on the parallel analysis (PA) criterion, 48 principal component factors (PC factors) were kept for subsequent analyses (Table S2, Supporting information). These PC factors accounted for 41% of the variation observed among

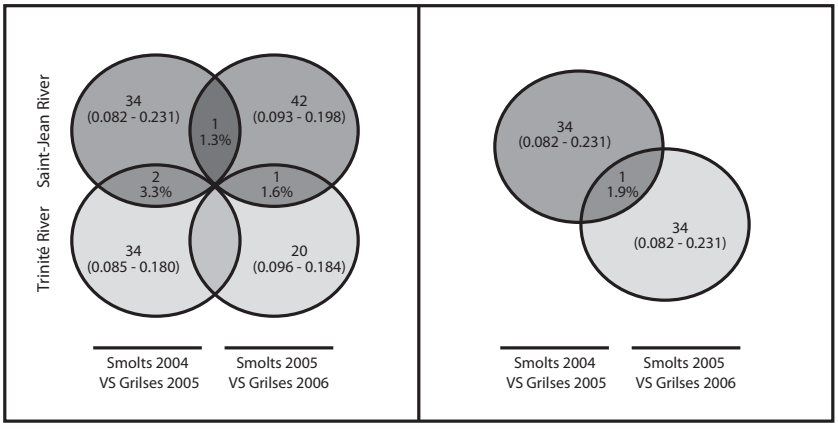

Fig. 2 Summary of single-locus genome scans results. Each circle represents a genome scan comparing juveniles (smolts) and adults (grilses) of a given population and cohort. Numbers in circles correspond to the number of loci under potential divergent selection in each genome scans with their given F_{ST} range in parenthesis. Numbers in overlapping regions correspond to common outliers found in two genome scans with the proportion of markers this number represent on the total number of outliers for both scans. Where there are no overlapping regions, common outliers were not found. Genome scan plots are presented on Fig. S1 (Supporting information).

genetic markers and the top three factors each accounted for at least 1% (PC1 = 5.06%; PC2 = 1.05%; PC3 = 1.00%). The first two axes of the canonical discriminant analysis displayed significant canonical correlations for populations, cohorts or life stages ($P < 0.001$; Table 1). Fig. 3 illustrates the spatial distribution of individuals along these two axes. Generalized linear models indicated that population of origin (Saint-Jean *vs*. Trinité) was the only source of variation discriminated on the first axis ($F = 16\,543$; $P < 0.001$; Table 2a). Among the 48 PC factors initially retained, only PC factor 1 significantly associated with the first axis of canonical correlation with a loading weight of 0.418.

The second axis contained three significant (Table 2b) sources of variation. In particular, the interaction term between populations and life stages ($F = 135.33$; $P < 0.0001$), life stages ($F = 98.52$; $P < 0.0001$), and more marginally the cohort (2004 *vs*. 2005; $F = 12.24$; $P = 0.0006$) were correlated with PC factors loading on this second axis. Therefore, we found genetic differences between life stages which are repeated in two independent cohorts but the interaction term indicates that this effect differs between populations. Indeed, the second axis clearly discriminates smolts and grilses from the Saint-Jean population but not from the Trinité population, a pattern repeated for both cohorts. One PC factor (PC factor 40) was also found to be significantly associated with the second axis of canonical correlation. PC factor 40 presented a canonical loading weight of 0.320 on which 34 markers have PC loading weights > 0.23 (Table S3, Supporting information). Therefore, these markers are significantly associated with the discriminant axis differentiating life stages in the Saint-Jean population. Among these 34 markers, only two were also divergent outliers in the single-locus genome scans (ESTNV_22435_210 and ESTV_16844_203).

Allelic effects and strength of selection

The quantile–quantile plot clearly demonstrates that the distribution of allelic effect covariance among markers significantly correlated with PC factor 40 is skewed towards higher values compared to a random set of markers (Fig. 4). Thus, except for an outlier point at the end of the distribution, the upper range of the distribution of values for multilocus markers was more than three times higher than that obtained for random markers. Consequently, the markers associated with genetic differences between life stages in the Saint-Jean River displayed significantly greater covariance in allelic effect than expected by chance, further confirming their association with different life stages repeated for two independent cohorts. Also, the distributions of Δp among all 2923 markers was skewed towards low values with a very high proportion of values between 0 and 0.02 and exponentially reducing in proportion for higher values up to 0.36 in cohort 2 (2005) of the Saint-Jean River and identical medians of 0.030 for both rivers (Fig. 5a). We then found that markers identified with the multilocus approach had Δp distributions which were more uniform with medians of 0.045 and 0.030 for the Saint-Jean and Trinité rivers, respectively, translating into modest change in frequencies for these markers (Fig. 5b). These small changes on a set of covarying SNPs are consistent with the polygenic selection hypothesis acting on many genomic targets and causing small changes of allele frequencies. In contrast, the single-locus genome scan identified markers with medians of 0.090 and 0.075 for the Saint-Jean and Trinité rivers, respectively (Fig. 5c), which is consistent with single large effect markers or genes putatively being highlighted by single-locus methods. The genomic distribution of Δp and loading weight presented a widely spread distribution of values along the different linkage groups (Fig. 6). The 34 markers identified using the multilocus approach were distributed among 14 different linkage groups with a maximum of 4 SNPs being observed on a single linkage group (Fig. 6).

Gene ontology and SNP annotation

The BLAST and annotation steps in BLAST2GO yielded annotations for 1119 SNPs. Using a FDR of 0.05, an enrichment analysis at the significance level of 0.05 did

Table 1 Summary of the canonical discriminant analysis. For each axis, eigenvalue, proportion of variance and cumulative variance accounted for, degrees of freedom, *F* statistic and *P*-value is given

Axis	Eigenvalue	Proportion	Cumulative	d.f.	F_{ST}	*P*-value
1	86.7251	0.9652	0.9652	336	5.03	<0.0001
2	1.3629	0.0152	0.9804	282	1.47	<0.0001
3	0.7548	0.0084	0.9888	230	1.08	0.228
4	0.3852	0.0043	0.9931	180	0.82	0.948
5	0.2960	0.0033	0.9964	132	0.69	0.994
6	0.2291	0.0025	0.9989	86	0.56	0.999
7	0.0966	0.0011	1.0000	42	0.35	0.999

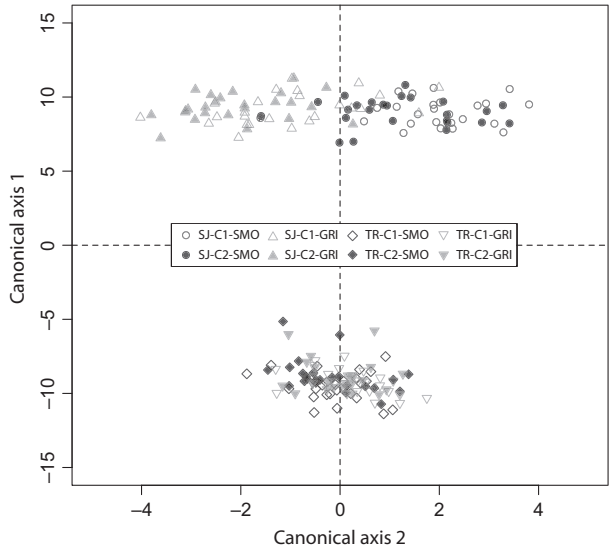

Fig. 3 Canonical discriminant analysis of genetic differentiation among individuals based on 2923 markers grouped in 48 principal components. Canonical axis 1 is significantly correlated with genetic variation among rivers, and Canonical axis 2 is significantly correlated with life stages and cohorts. (SJ, Saint-Jean River, TR, Trinité River, C1, cohort 2004, C2, cohort 2005, SMO, smolts and GRI, grilses).

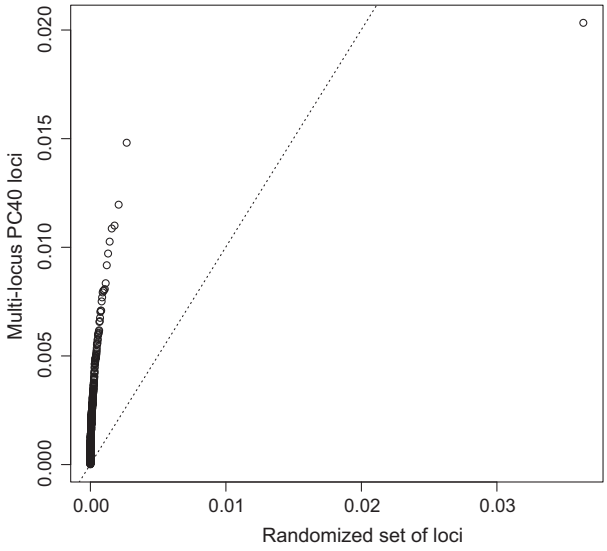

Fig. 4 Quantile–quantile plot of the distribution of allelic effect covariance among 34 markers significantly correlated with PC factor 40 on the *Y*-axis and a 34 marker randomly chosen among all markers 1000 times on the *X*-axis. The dotted line represents X = Y relation.

Table 2 Summary of the generalized linear models. For each of both significant canonical discriminant axis (a) axis 1, and (b) axis 2; the degrees of freedom, sum of squares, *F* statistics and *P*-values are given for each source of variation tested in the model. Significant values are bold and italicized

Source	d.f.	Sum square	F_{ST}	*P*-value
(a)				
Rivers	***1***	***16 641***	***16 543***	***<0.0001***
Life stages	1	2.436	2.42	0.121
Cohorts	1	3.613	3.59	0.060
Life stages*Cohorts	1	0.426	0.42	0.516
Rivers*Life stages	1	0.270	0.27	0.605
(b)				
Rivers	1	0.038	0.04	0.848
Life stages	***1***	***102***	***98.52***	***<0.0001***
Cohorts	1	12.60	12.24	0.0006
Life stages*Cohorts	1	0.385	0.37	0.542
Rivers*Life stages	***1***	***139***	***135.33***	***<0.0001***

not indicate the significant over- or under-representation of any biological pathway among the markers under divergent selection compared to all other markers in any of the individual genome scans or for the 130 divergent outliers from all genome scans taken as a test group. Similarly, the enrichment analysis did not reveal any over- or under-representation of biological functions or processes for the 34 markers identified using the multilocus approach compared to all other markers. Nevertheless, many annotations were associated with mitochondrial functions. Annotations and GO-terms identified for the 34 markers significantly correlated with PC factor 40 are reported in Table S4 (Supporting information).

Discussion

Detecting selective mortality – single-locus genome scan versus polygenic multilocus approach

We assessed for the first time whether nonrandom, genetically based selection could contribute to differential mortality during the first year at sea of salmon. We used a population genomic approach and sampled two genetically distinct Atlantic salmon populations by comparing outmigrating smolts and returning adults of the same cohorts. In doing so, we also compared the outcomes of a single-locus outlier approach vs. polygenic multivariable statistical method in detecting repeated (parallel) signals of selection at the genome level.

Both the single and multilocus approaches identified markers potentially under divergent selection during the first year at sea, but only the multilocus approach identified a clear pattern of nonrandom differential mortality that was repeated across two independent cohorts. Specifically, the single-locus F_{ST} based genome scans found 0.7% to 1.4% of SNP markers to be potentially under the effect of divergent selection in any of

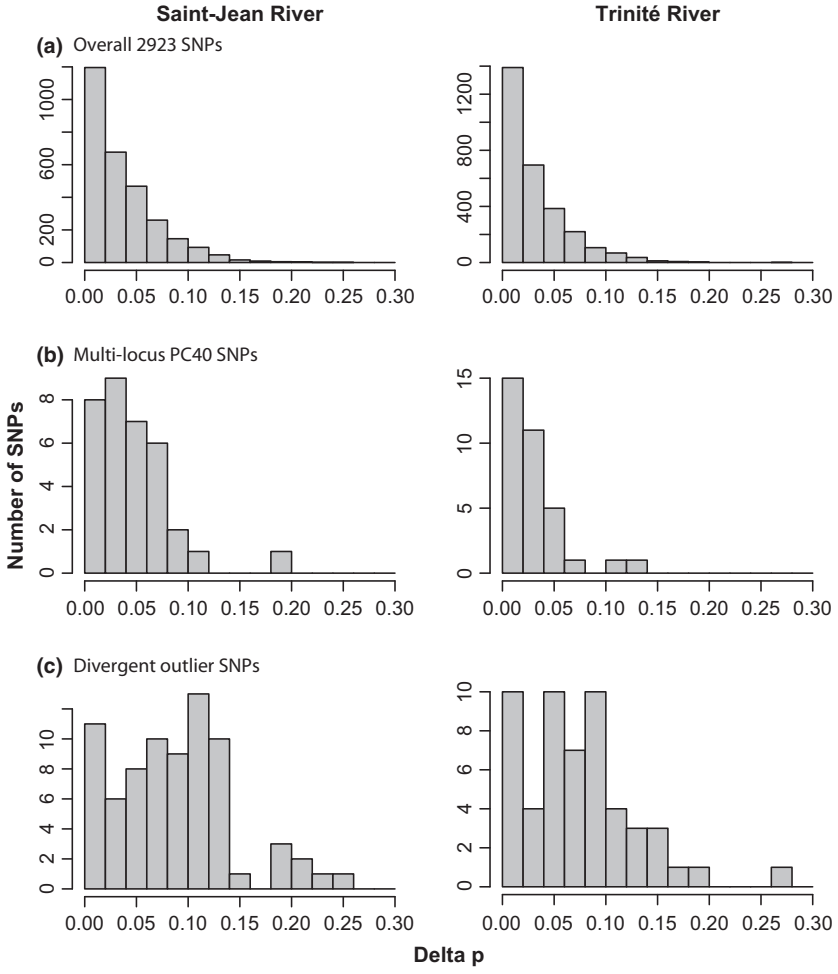

Fig. 5 Distribution of allelic frequency changes (Δp) between life stages in each population with the two cohorts grouped together: a) over all 2923 markers, b) among 34 markers significantly correlated with PC factor 40, and c) among single-locus genome scan divergent outliers.

the four comparisons of smolts and grilses. While these markers were highly differentiated between life stages in independent comparisons, of the 130 markers identified as outliers by the single-locus method, no more than two were significant in more than one analysis and none were parallel in more than two. Overall, the single-locus genome scan approach was not able to detect any convincing pattern of genetically based differential mortality between smolts and grilses. We cannot entirely refute the possibility that nonparallelism in these outliers translate different selective causes of mortality that varied in time and space, or that other factors influenced the different results obtained for the two populations. For example, samples were collected differently from each location which could influence the results and selective forces for mortality probably differ. Yet, it would seem exceptional that these factors would differ to the point that essentially no markers would be a common target of selection in either cohorts or populations. Instead, we propose that such pronounced nonparallelism in outliers detected mainly translate the random effect of factors other than selection (Le Corre & Kremer 2012).

In contrast, the multilocus approach identified 34 covarying SNPs that loaded on a multivariate axis of differentiation that could distinguish between smolts and grilses based on the multilocus genotype composition in the Saint-Jean River. The same pattern of differentiation was also repeated for both cohorts ruling out the possibility that the observed pattern between smolts and grilse was random or sample biased. These covarying markers also displayed small changes in allele frequencies between life stages rather than large changes translating into high levels of genetic differentiation. Thus, the distribution of allele frequency change was offset towards low values, especially when compared to genome scan outliers from our single-locus comparisons. Moreover, allele frequencies at these 34 SNP covaried more than expected by chance. Based on the contrasting results between single-locus genome scans *vs.* polygenic mutlivariate analysis, we argue that the complexity of a phenomenon such as genetically based

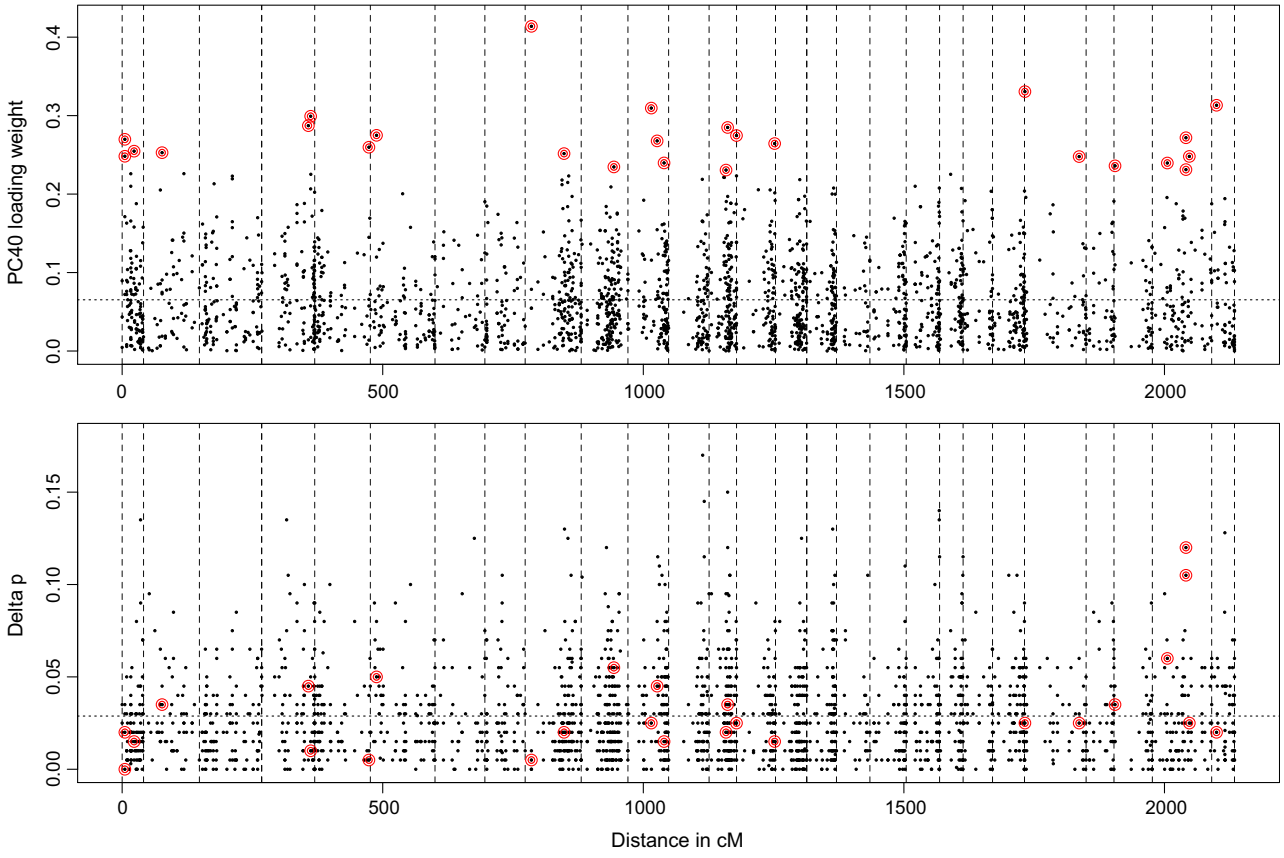

Fig. 6 Genetic linkage map showing the distribution of principal component 40 (PC40) loading weight of each of the 2923 SNP markers on the top panel and overall allelic frequency changes between life stages on the bottom panel. Vertical dashed lines distinguish linkage groups. On both panels, circled dots indicate the 34 SNPs significantly correlated with PC factor 40 and the horizontal dotted line indicates the average loading weight (top) and delta p (bottom) among all markers.

differential mortality at sea is best explored via a multilocus genetic statistical framework, as predicted from quantitative genetics theory and in agreement with Le Corre & Kremer (2012). Indeed, since many selective agents are probably acting to cause mortality at sea and the phenotypic traits potentially under selection are most likely polygenic, it is plausible that many genes across the genome, associated with different biological functions are individually under the effect of relatively mild selection rather than a few genes of large effects. This is supported by the annotation analysis, which revealed that many biological functions were represented among the covarying SNP set although no specific biological function was over-represented. Instead of detecting potential selective mortality as highlighted by the multilocus approach, single-locus genome scans revealed outlier markers randomly changing in time and space. Given the small number of markers common to both methods, which are suggestive of false positives, our results are consistent with the high rate of false positives that F_{ST} based methods can yield in experiments lacking other ecological and functional support. The conditions needed to detect parallelism in outliers might be restricted to simple phenotypic traits encoded by genes with very large effects (e.g. Cresko *et al.* 2004), which emphasizes the need for a multivariate approach to identifying targets of selection of small effects, especially for complex traits and/or when multiple agents of selection are thought to be involved.

The use of genome-wide screens to detect selection in natural populations has become a very popular pursuit, particularly relying on the detection of outlier loci to uncover signatures of selection. Yet, these signatures are generally interpreted without questioning basic model assumptions (Lotterhos & Whitlock 2014). Evidence is accumulating that this may lead to erroneous conclusions due to false positives (through recombination hotspots, population stratification, endogenous incompatibilities) or false negatives (e.g. weak selection relative to migration or drift) (Bierne *et al.* 2011; Le Corre & Kremer 2012). Somewhat surprisingly, studies to date have generally ignore the fact that most selected

loci may be involved in polygenic adaptation in which case theory would predict small changes in allelic frequencies (Pritchard & Di Rienzo 2010). To our knowledge, only a handful of studies have addressed this is issue using a different analytical framework than the one we propose here. For instance, Hancock *et al.* (2010) contrasted the results of approaches based on haplotype structure and differentiation of allele frequencies to those from a method for identifying SNP strongly correlated with environmental variables. Their results suggested that the first group of approaches tended to identify only variants with relatively strong phenotypic effects, whereas the environmental correlation methods can detect variants that make smaller contributions to an adaptive trait. More recently, Daub *et al.* (2013) proposed to evidence polygenic selection in humans by detecting signals of adaptation at the pathway or gene-set level instead of analysing single independent genes. Using a gene-set enrichment test to identify genome-wide signals of adaptation among human populations, the authors found that most pathways globally enriched for signals of positive selection are either directly or indirectly involved in immune response. Although not directly related to the detection of selection but instead in finding genotype–phenotype association, other studies have shown the potential superiority of a multivariate analytical framework over traditional genome-wide association studies (GWAS) in explaining phenotypic variation. Yang *et al.* (2010) pointed out that SNPs discovered by genome-wide association studies (GWASs) account for only a small fraction of the genetic variation of complex traits in human populations, raising the question: 'Where is the remaining heritability?' They thus estimated the proportion of variance for human height explained by SNP variation using a multivariate linear model analysis. They showed that 45% of variance could be explained by considering all SNPs simultaneously, and concluded that most of the heritability in human height is not missing but has not previously been detected because the individual effects are too small to pass stringent significance tests. Finally, in perhaps the only such studies in nonhuman, wild populations performed to date, Ma *et al.* (2010) examined genetic variation in genes from the photoperiodic pathway in a tree species (*Populus tremula*) for signatures of diversifying selection in response to varying light regimes across a latitudinal gradient. While they failed to identify any obvious outlier SNPs using a genome scan approach, they observed a pronounced covariance in allelic effects across populations for growth cessation. Their results thus suggested that spatially variable selection could be affecting genes from the photoperiod pathway even if selection is not strong enough to cause individual loci to be identified as outliers. Overall, then studies in humans and that of Ma *et al.* (2010) corroborate the results of our study in pointing the potential superiority of a polygenic multilocus analytical framework in detecting signal of selection acting on complex traits.

Differential mortality at sea and evolutionary changes

Many factors have been suggested to contribute to the large declines in Atlantic salmon populations across the globe. The majority of the identified factors focus on the freshwater and coastal phase of the species' life history where monitoring is more accessible (e.g. the impacts of fish farming, freshwater habitat degradation and warming, barrier to migration, etc.; reviewed in Aas *et al.* (2011)). However, little is known about the factors contributing to mortality at sea, despite the observed declines in sea survival and many hypotheses for what is contributing to these declines. For example, climate change, changes in trophic levels and increased predation have all been proposed as hypothesis for increased mortality at sea (Beaugrand & Reid 2012; Sheehan *et al.* 2012). It is also unclear if such factors act on populations randomly or in a more deterministic manner. Given the repeated pattern of genotypic changes detected between smolts and grilse in both cohorts of the Saint-Jean River by the canonical discriminant analysis, our results strongly support the hypothesis of genetically based selective mortality at sea in Atlantic salmon from the Saint-Jean River population. In fact, the temporal replication of similar genetically based selective mortality for that population strengthens the hypothesis of nonrandom evolutionary changes occurring during the marine phase of this population. However, these sustained changes are also of concern because North American populations of Atlantic salmon have been established for at least 11 000 years (Verspoor *et al.* 2007) and it is likely that these populations have been adapting to marine migration since this time. Therefore, given the time elapsed and their large population sizes, one could argue that any loci increasing marine survival in Atlantic salmon will have been fixed long ago, so observing current selective mortality is unlikely. However, given the fluctuating conditions occurring in the marine environment attributable to natural events such as the North Atlantic Oscillation (NAO) (Stenseth *et al.* 2003), we argue for a type of balancing selection maintaining diversity among traits underlying marine survival. Nevertheless, we did observe subtle yet significant and repeatable changes in allele frequencies for one of the two populations investigated. The presence of recently selected 'marine' loci suggests that genetic variation for alleles contributing to marine survival has been maintained in these populations.

This may be due to selective factors of recent origin, perhaps associated with recent changes in environmental conditions. For instance, important rise in sea temperature was shown as tightly linked to growth and survival of migrating Atlantic salmon (Beaugrand & Reid 2012; Mills *et al.* 2013).

On the other hand, results obtained for the Trinité River suggest that the same selective pressures did not affect this population, as observed genomic changes between life stages in this population were apparently random, with no evidence for repeatable genetically based selective mortality. Under the hypothesis of genetically based selective mortality at sea, the contrasting patterns of selection between the Saint-Jean and Trinité rivers populations could possibly be explained by different migratory routes to the feeding grounds and/or migration to different feeding grounds off the coast of Newfoundland and Labrador. Indeed Atlantic salmon from populations of the North shore of the Gulf of St. Lawrence (where the Trinité River population is located) likely migrate out to sea through the Strait of Belle Isle at the northern tip of Newfoundland (Dutil & Coutu 1988). On the other hand, rivers further south from the Gaspésie Peninsula (where the Saint-Jean River population is located) most likely use both out migrating routes of the Gulf of St. Lawrence, namely the northern Strait of Belle Isle and the southern passage through the Cabot Strait (Belding 1940; Lefèvre *et al.* 2012). Moreover, it has been reported that Atlantic salmon from different regions in the United Kingdom segregate in distinct feeding grounds (MacKenzie *et al.* 2012). Populations from Saint-Jean and Trinité rivers represent two regional genetic groups of salmon previously identified (Dionne *et al.* 2008) and shown to be potentially locally adapted to distinct environmental settings (Bourret *et al.* 2013b). In particular, early post-smolt life stage definitely occurs in different estuarine areas of the Gulf of St-Lawrence River for these populations. Also, there is evidence for high mortality rate during the early marine migration, especially in the estuaries and the river mouths, reaching up to 71% (reviewed in Thorstad *et al.* 2012). Also, Plantalech manel-la *et al.* (2011) reported differences in early (estuarine) migratory behaviour of distinct populations of Atlantic salmon that could have fitness consequences. Other studies have also found a negative correlation between marine water temperatures and post-smolt survival of migrating Atlantic salmon (Friedland *et al.* 2000; Todd *et al.* 2008; Hvidsten *et al.* 2009). The Saint-Jean River estuary is located in the southern part of Québec and experiences higher temperatures than the Trinité River (Dionne *et al.* 2013). The estuary of the Saint-Jean River is also much larger than that of the Trinité River (Dionne *et al.* 2013) and as a result, smolts from the Saint-Jean River would need to spend more time crossing this environment which is typically associated with high mortality and selective pressures (e.g. predation, high temperature exposure; Thorstad *et al.* 2012). However, one must also consider that the adults were collected using different methods for the two populations. In the Saint-Jean River adults were collected by fishermen, which could represent some form of selective harvesting whereas in the de la Trinité River, fish were passively sampled in a fish ladder, which is undoubtedly a potentially less selectively biased means of collecting returning salmon. Thus, one could argue that adults genotyped from the Saint-Jean River might not truly represent surviving marine migrants, but selection during their reproductive upstream migration. Nevertheless, our observations still raise the hypothesis that any recent change in the environmental conditions encountered along their two migratory routes may have differentially impacted the Saint-Jean and Trinité populations. More information on local early marine migration and monitoring of oceanic environmental conditions is needed to further test this hypothesis.

Alternatively, one could argue that the observed pattern of differentiation between smolts and grilse may be attributable to a sex-biased sampling between life stages since it is known that the majority of Atlantic salmon returning after one winter at sea in both the Saint-Jean and Trinité rivers are males, whereas most multiseawinter (MSW) fish are females (Dionne *et al.* 2013). Thus, the observed genetically based differentiation might reveal divergence between sexes rather than a result of differential mortality at sea. For this to explain the repeated differences observed between both populations would require that they differ in the proportion of the two sexes in returning grilses. Yet, grilses were composed of a majority of males in both populations (100% in SJ and 86% in TR), while the proportion of both sexes in smolts was relatively similar in both rivers (average of 74% males in SJ and 72% males in TR). Therefore, it is unlikely that the differences observed between populations reflect a global effect of sex.

Concluding remarks

This study is the first attempt to use population genomics to study the evolutionary changes associated with the global declines in adult Atlantic salmon returns. Detecting a specific causative factor for mortality was beyond the scope of this study. Indeed, much more information about the physical and environmental conditions encountered by Atlantic salmon during their marine phase and the genomic composition of other life stages will be required to reach this goal. Furthermore, given the complexity of the trait under selection and

the diversity of potential selective forces in the marine environment, we acknowledge that increasing the number of individuals genotyped from each population as well as the geographical and genomic coverage will be a necessary step in future studies. Namely, a more comprehensive study would allow a finer characterization of markers and interactions underlying selective mortality across the genome. Indeed, genome-wide association studies (GWAS) aiming at detecting association between variation at the genome and phenotypic level for complex traits typically involve the use of many more markers than what we could use here. Thus, even if we reach a genome coverage of about 100 markers per chromosome, which is much more than typical 'saturated' genetic maps published until very recently, it is very likely that the relatively small number of markers we used led to missing many of the associations between genotyped markers and the actual target of selection in outbred salmon populations with limited linkage disequilibrium. Nevertheless, we were able to detect nonrandom, temporally repeatable, genetically based selective mortality occurring between the smolts and grilse life stages in one of two rivers. This represents a major step towards elucidating the dynamics of differential mortality at sea between salmon from different population origin. This pattern could only be detected with a multivariate genetic framework inspired from quantitative genetics theory and designed to consider markers as covarying multilocus entities. In contrast, the F_{ST} based single-locus approach was unable to reveal any pattern of allele frequency change putatively associated with differential mortality at sea, further supporting arguments in favour of a paradigm shift in the analysis of selection that has been largely based thus far on the detection of strong selective effects causing large allele frequency on single markers to searching for polygenic selection of smaller effects causing modest by covaried changes in allele frequencies at many genomic regions, which is predicted to be more commonly acting on complex, fitness-related traits in natural populations.

Acknowledgements

We warmly acknowledge the contribution of Dr. Matthew P. Kent in providing his expertise for the genotyping of analysed individuals at CIGENE. We thank the Ministère du Développement durable, de l'Environnement, de la Faune et des Parcs du Québec (MDDEFP) for sampling Atlantic salmon. We are grateful to Marie-Hélène Perreault and Geneviève Ouellet-Cauchon for technical assistance in the laboratory, Gaétan Daigle for support in statistical analyses, and Pierre-Alexandre Gagnaire and Anne C. Dalziel for constructive insights on the manuscript. We are also grateful to associate editor S. Narum as well as anonymous referees for their valuable suggestions that improved a previous version of the paper. The SNP discovery, array development and genotyping were 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 in close collaboration with the MDDEFP 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.

References

Aas O, Einum S, Klemetsen A, Skurdal J (2011) *Atlantic Salmon Ecology*. Wiley-Blackwell, Oxford.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a F(st)-outlier method. *BMC Bioinformatics*, **9**, 323, doi:10.1186/1471-2105-9-323.

Barrett RDH, Rogers SM, Schluter D (2008) Natural selection on a major armor gene in threespine stickleback. *Science*, **322**, 255–257.

Beaugrand G, Reid PC (2012) Relationships between North Atlantic salmon, plankton, and hydroclimatic change in the Northeast Atlantic. *ICES Journal of Marine Science*, **69**, 1549–1562.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **263**, 1619–1626.

Belding DL (1940) Migration of the Atlantic salmon (*Salmo salar*) in the Gulf of St Lawrence as determined by tagging experiments. *Transactions of the American Fisheries Society*, **69**, 290–295.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate - A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological*, **57**, 289–300.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, **20**, 2044–2072.

Bourret V, Kent MP, Primmer CR *et al.* (2013a) 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.

Bourret V, Dionne M, Kent MP, Lien S, Bernatchez L (2013b) Landscape genomics in Atlantic salmon (*Salmo salar*): searching for gene-environment interactions driving local adaptation. *Evolution*, **67**, 3469–3487.

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, doi:10.1186/1471-2164-1113-1432.

Cresko WA, Amores A, Wilson C *et al.* (2004) Parallel genetic basis for repeated evolution of armor loss in Alaskan three-

spine stickleback populations. *Proceedings of the National Academy of Sciences, USA*, **101**, 6050–6055.
Daub JT, Hofer T, Cutivet E *et al.* (2013) Evidence for polygenic adaptation to pathogens in the human genome. *Molecular Biology and Evolution*, **30**, 1544–1558.
De Mita S, Thuillet A-C, Gay L *et al.* (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, **22**, 1383–1399.
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.
Dionne M, Cauchon V, Harnois N, Fournier D (2013) Écologie et évolution des populations témoins de saumon atlantique au Québec: Rapport de recherche 2012. Pp. 80 in D. g. n. r. d. l. e. s. l. f. e. s. habitats, ed. Ministère des Ressources naturelles et de la Faune, Québec.
Dutil JD, Coutu JM (1988) Early marine life of Atlantic salmon, *Salmo salar*, postsmolts in the northern Gulf of St-Lawrence. *Fishery Bulletin*, **86**, 197–212.
Ford JK, Maccallum RC, Tait M (1986) The application of exploratory factor analysis in applied psychology - A critical review and analysis. *Personnel Psychology*, **39**, 291–314.
Friedland KD, Hansen LP, Dunkley DA, MacLean JC (2000) Linkage between ocean climate, post-smolt growth, and survival of Atlantic salmon (*Salmo salar* L.) in the North Sea area. *ICES Journal of Marine Science*, **57**, 419–429.
Gibson AJF, Bowlby HD, Hardie DC, O'Reilly PT (2011) Populations on the brink: low abundance of southern upland Atlantic Salmon in Nova Scotia, Canada. *North American Journal of Fisheries Management*, **31**, 733–741.
Gotz S, Garcia-Gomez JM, Terol J *et al.* (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research*, **36**, 3420–3435.
Hancock AM, Alkorta-Aranburu G, Witonsky DB, Di Rienzo A (2010) Adaptations to new environments in humans: the role of subtle allele frequency shifts. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **365**, 2459–2468.
Hansen LP, Quinn TR (1998) The marine phase of the Atlantic salmon (*Salmo salar*) life cycle, with comparisons to Pacific salmon. *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 104–118.
Horn JL (1965) A rationale and test for the number of factors in factor analysis. *Psychometrika*, **30**, 179–185.
Hvidsten NA, Jensen AJ, Rikardsen AH *et al.* (2009) Influence of sea temperature and initial marine feeding on survival of Atlantic salmon *Salmo salar* post-smolts from the Rivers Orkla and Hals, Norway. *Journal of Fish Biology*, **74**, 1532–1548.
ICES 2013. Report of the Working Group on North Atlantic Salmon (WGNAS). Pp. 378, Copenhagen, Denmark.
Latta RG (1998) Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *The American Naturalist*, **151**, 283–292.
Lautenschlager GJ (1989) A comparison of alternatives to conducting Monte-Carlo analyses for determining parallel analysis criteria. *Multivariate Behavioral Research*, **24**, 365–395.
Le Corre V, Kremer A (2003) Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics*, **164**, 1205–1219.
Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, **21**, 1548–1566.
Lefèvre MA, Stokesbury MJW, Whoriskey FG, Dadswell MJ (2012) Atlantic salmon post-smolt migration routes in the Gulf of St. Lawrence. *ICES Journal of Marine Science*, **69**, 981–990.
Liu KJ, Muse SV (2005) POWERMARKER: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, **21**, 2128–2129.
Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, **23**, 2178–2192.
Ma X-F, Hall D, St Onge KR, Jansson S, Ingvarsson PK (2010) Genetic differentiation, clinal variation and phenotypic associations with growth cessation across the *Populus tremula* photoperiodic pathway. *Genetics*, **186**, 1033–1044.
Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics*, **10**, 565–577.
MacKenzie KM, Trueman CN, Palmer MR *et al.* (2012) Stable isotopes reveal age-dependent trophic level and spatial segregation during adult marine feeding in populations of salmon. *ICES Journal of Marine Science*, **69**, 1637–1645.
Miller K, Li S, Kaukinen KH *et al.* (2011) Genomic signatures predict migration and spawning failure in wild Canadian salmon. *Science*, **331**, 214–217.
Mills KE, Pershing AJ, Sheehan TF, Mountain D (2013) Climate and ecosystem linkages explain widespread declines in North American Atlantic salmon populations. *Global Change Biology*, **19**, 3046–3061.
Nadeau NJ, Jiggins CD (2011) A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends in Genetics*, **26**, 484–492.
Narum SR, Hess JE (2011) Comparison of *FST* outlier tests for SNP loci under selection. *Molecular Ecology Resources*, **11**, 184–194.
Plantalech manel-la N, Chittenden CM, Okland F *et al.* (2011) Does river of origin influence the early marine migratory performance of *Salmo salar*? *Journal of Fish Biology*, **78**, 624–634.
Pritchard JK, Di Rienzo A (2010) Adaptation - not by sweeps alone. *Nature Reviews Genetics*, **11**, 665–667.
Sheehan TF, Reddin DG, Chaput G, Renkawitz MD (2012) SALSEA North America: a pelagic ecosystem survey targeting Atlantic salmon in the Northwest Atlantic. *ICES Journal of Marine Science*, **69**, 1580–1588.
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.
Stenseth NC, Ottersen G, Hurrell JW *et al.* (2003) Studying climate effects on ecology through the use of climate indices: the North Atlantic Oscillation, El Nino Southern Oscillation and beyond. *Proceedings of the Royal Society B-Biological Sciences*, **270**, 2087–2096.
Thorstad EB, Whoriskey F, Uglem I, Moore A, Rikardsen AH, Finstad B (2012) A critical life stage of the Atlantic salmon *Salmo salar*: behaviour and survival during the smolt and initial post-smolt migration. *Journal of Fish Biology*, **81**, 500–542.

Todd CD, Hughes SL, Marshall CT, Maclean JC, Lonergan ME, Biuw EM (2008) Detrimental effects of recent ocean surface warming on growth condition of Atlantic salmon. *Global Change Biology*, **14**, 958–970.

Verspoor E, Stradmeyer L, Nielsen JL (2007) *The Atlantic Salmon: Genetics, Conservation, and Management*. Blackwell Publishing, Oxford.

Vilas A, Perez-Figueroa A, Caballero A (2012) A simulation study on the performance of differentiation-based methods to detect selected loci using linked neutral markers. *Journal of Evolutionary Biology*, **25**, 1364–1376.

Yang JA, Benyamin B, McEvoy BP *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, **42**, 565–569.

Yano A, Guyomard R, Nicol B *et al.* (2012) An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. *Current Biology*, **22**, 1423–1428.

M.D., L.B. and V.B. designed the study. M.D. organized the collection of samples. V.B. analyzed the data. V.B., M.D. and L.B. wrote the manuscript and contributed to revisions.

Data accessibility

Genotype file, input data file and SAS script for multilocus discriminant analysis, input file and script for covariance of allelic effect analysis, and input files and script for delta p analysis are available at Dryad doi:10.5061/dryad.j86v9.

Supporting information

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

Fig. S1 Differentiation (F_{ST}) as a function of heterozygosity as estimated by LOSITAN (Antao *et al.* 2008) for four comparison of smolts and grilses in: a) Saint-Jean River cohort 1, b) Saint-Jean River cohort 2, c) Trinité River cohort 1, and d) Trinité River cohort 2. In each panel, outliers markers ($P < 0.01$) are marked by circled dots, dashed lines represent upper and lower 99% confidence levels and dotted lines indicates the average F_{ST} across all loci.

Table S1 Summary of the detection of markers potentially under selection following genome scans implemented in LOSITAN (Antao *et al.* 2008).

Table S2 Summary of principal component analysis performed on 2923 SNP markers for 200 samples.

Table S3 Summary of markers' absolute loading weight on PC factor 40 in the Abs(F40) column.

Table S4 Blast results from BLAST2GO with blast e-value threshold of 1×10^{-3} and gene ontology (GO) terms annotation for blast of e-value inferior to 1×10^{-10}.