Conservation genomics of anadromous Atlantic salmon across its North American range: outlier loci identify the same patterns of population structure as neutral loci

JEAN-SÉBASTIEN MOORE,*[1] VINCENT BOURRET,*†[1] MÉLANIE DIONNE,‡ IAN BRADBURY,§ PATRICK O'REILLY,¶ MATTHEW KENT,** GÉRALD CHAPUT†† and LOUIS BERNATCHEZ*

*Institut de Biologie Intégrative et des Systèmes, Université Laval, 1030 Avenue de la Médecine, Québec, Québec G1V 0A6, Canada, †Direction de la Protection de la Faune, Ministère des Forêts, de la Faune et des Parcs, Québec, Québec G1S 4X4, Canada, ‡Direction de la Faune Aquatique, Ministère des Forêts, de la Faune et des Parcs, Québec, Québec G1S 4X4, Canada, §Science Branch, Fisheries and Oceans Canada, 80 East White Road, St. John's, Newfoundland A1C 5X1, Canada, ¶Science Branch, Fisheries and Oceans Canada, Bedford Institute of Oceanography,1 Challenger Drive, Dartmouth, Nova Scotia, B2Y 4A2, Canada, **Centre for Integrative Genetics (CIGENE), Department of Animal and Aquacultural Sciences (IHA), Norwegian University of Life Sciences, PO Box 5003, 1432 Aas, Norway, ††Fisheries and Oceans Canada, PO Box 5030, Moncton, New Brunswick E1C 9B6, Canada

Abstract

Anadromous Atlantic salmon (*Salmo salar*) is a species of major conservation and management concern in North America, where population abundance has been declining over the past 30 years. Effective conservation actions require the delineation of conservation units to appropriately reflect the spatial scale of intraspecific variation and local adaptation. Towards this goal, we used the most comprehensive genetic and genomic database for Atlantic salmon to date, covering the entire North American range of the species. The database included microsatellite data from 9142 individuals from 149 sampling locations and data from a medium-density SNP array providing genotypes for >3000 SNPs for 50 sampling locations. We used neutral and putatively selected loci to integrate adaptive information in the definition of conservation units. Bayesian clustering with the microsatellite data set and with neutral SNPs identified regional groupings largely consistent with previously published regional assessments. The use of outlier SNPs did not result in major differences in the regional groupings, suggesting that neutral markers can reflect the geographic scale of local adaptation despite not being under selection. We also performed assignment tests to compare power obtained from microsatellites, neutral SNPs and outlier SNPs. Using SNP data substantially improved power compared to microsatellites, and an assignment success of 97% to the population of origin and of 100% to the region of origin was achieved when all SNP loci were used. Using outlier SNPs only resulted in minor improvements to assignment success to the population of origin but improved regional assignment. We discuss the implications of these new genetic resources for the conservation and management of Atlantic salmon in North America.

Introduction

The increased availability of genomic data in nonmodel organisms has the potential to revolutionize the use of genetic tools in conservation and management (Kohn

Correspondence: Jean-Sébastien Moore, Fax: +1 418 656-7176;
E-mail: jean-sebastien.moore.1@ulaval.ca
[1]Co-first authors.

et al. 2006; Avise 2009; Primmer 2009; Allendorf et al. 2010). At the most basic level, the greater number of loci available can increase the precision and accuracy of population parameters of interest formerly estimated with a small number of neutral markers (Allendorf et al. 2010). More fundamentally, however, genomic tools can also provide new types of information not available with traditional markers. For instance, they can provide information on the genomic architecture of inbreeding and/or outbreeding depression (Allendorf et al. 2010) and highlight regions of differential introgression in the genome of native taxa under threat from hybridization (Crispo et al. 2011; Lamaze et al. 2012; Hohenlohe et al. 2013). Furthermore, genomic data can be used to infer the presence of loci or genomic regions that show higher differentiation and are putatively under directional selection (Nosil et al. 2009; Stapley et al. 2010). Such outlier loci could therefore offer an opportunity to assess patterns of local adaptation among populations, thus informing conservation and management.

Establishing conservation units based on objective and practical criteria has always been a challenge (Moritz 1994; Waples 1995; Fraser & Bernatchez 2001). While neutral markers may provide appropriate tools to delimit reproductively isolated populations (Waples & Gaggiotti 2006), their utility in determining their adaptive potential or ecological distinctiveness has been questioned (Crandall et al. 2000; Fraser & Bernatchez 2001; McKay & Latta 2002). The inclusion of phenotypic or ecological data in the definition of conservation units has been advocated by some (Waples 1991; Crandall et al. 2000), but remained impractical for most taxa of conservation concern. The analysis of genomic data combined with genome scan approaches identifying putatively adaptive loci could provide a practical solution to this problem (Bradbury et al. 2013). Recently, Funk et al. (2012) proposed a framework to integrate adaptive genetic variation from genomic data into the definition of conservation units. In short, they propose a sequential assessment of population structure using in turn putatively neutral loci and putatively adaptive loci. In principle, the inclusion of information on putatively adapted loci can help resolve questions regarding the spatial scale of local adaptation and help tailor conservation and management actions to specific conditions.

Anadromous Atlantic salmon (Salmo salar) is a species of biological as well as cultural and economic importance (Dodson et al. 1998; COSEWIC 2011). Population declines, however, have also made it a species of conservation concern (COSEWIC 2011; Chaput 2012). Historically, anthropogenic habitat destruction and overexploitation have severely reduced the range of the species in the United States (Spidle et al. 2001) and reduced abundance in southern Canada (COSEWIC

2011), with more recent declines (since the early 1990's) possibly resulting from increased mortality at sea (Friedland et al. 2005; Chaput 2012). In addition, the species is popular with the sport fishing industry, which targets adults returning from their marine migrations to spawn. Offshore fisheries also operate in Greenland, Labrador and on Saint-Pierre and Miquelon Islands targeting migrating adults in the marine phase of their life cycle (Chaput 2012). In Greenland, for example, a subsistence fishery targets adults migrating from all regions of North America to the feeding grounds of West Greenland, making it an important mixed-stock fishery (Chaput et al. 2005; Gauthier-Ouellet et al. 2009). Improving genetic tools for mixed-stock fishery analysis and individual assignments to natal rivers or to their management unit of origin is therefore of interest. Furthermore, sampling as many source populations as possible is crucial to increase the precision of estimates of stock composition and of individual assignment. The use of a large number of SNPs, and especially 'outlier' SNPs, offers the promise to improve the power of mixed-stock fishery analysis and assignment tests (Russello et al. 2011).

An important paradigm in existing conservation and management strategies for Atlantic salmon is that populations are structured hierarchically (Dodson et al. 1998; COSEWIC 2011). At a small scale, the precise homing behaviour of the species (Stabell 1984; Hendry et al. 2004) leads to genetic differentiation among rivers (King et al. 2001; Spidle et al. 2001; Dionne et al. 2008; Perrier et al. 2011; Bradbury et al. 2014a) and in some cases within rivers (Garant et al. 2000; Vähä et al. 2007; Dionne et al. 2009). At a larger scale, dispersal occurs, but effective migration is spatially restricted, leading to genetically differentiated regional groups (Dionne et al. 2008; Tonteri et al. 2009; Bourret et al. 2013a; Bradbury et al. 2014a). These genetically differentiated regional groups form the basis of management and conservation units (COSEWIC 2011), because they represent units with restricted gene flow that are at least partially demographically independent (Waples & Gaggiotti 2006). Furthermore, there is increasing evidence that these genetic groups are differentially adapted to spatially varying environmental conditions (Taylor 1991; Garcia de Leaniz et al. 2007; Fraser et al. 2011), such as pathogen diversity (Dionne et al. 2007), water pH (Fraser et al. 2008), temperature and geology (Bourret et al. 2013a). Anadromous Atlantic salmon would thus represent a good candidate species to evaluate the potential of adaptive markers to enhance the definition of conservation units.

In this study, data from a geographically extensive microsatellite database are combined with thousands of SNP markers to apply a conservation genomics

framework to anadromous populations of Atlantic salmon in North America. Our first goal was to provide a North American-wide assessment of population structure on the basis of both neutral and putatively adaptive genetic variation to inform the establishment of conservation units for management. Population structure has been examined previously for the Québec populations (Dionne *et al.* 2008), the Newfoundland populations (Bradbury *et al.* 2014a) and the Nova Scotia populations (McConnell *et al.* 1997; Vandersteen Tymchuk *et al.* 2010). A population genomics assessment conducted on the entire North American range in a single analysis, however, is lacking (but see King *et al.* (2001) for a range-wide assessment of microsatellite DNA variation based on a small number of sampling locations and Verspoor (2005) for a North American-wide study based on allozymes). By combining results from previous studies to new data, we now have the most extensive Atlantic salmon microsatellite database to date to infer population structure with high geographical resolution: the database is comprised of data from 149 sampling locations, with over 9000 individuals genotyped at 15 microsatellite loci. To evaluate the usefulness of integrating adaptive variation in the definition of conservation units, we used SNP data generated with a medium-density SNP array from a subset of the populations (Bourret *et al.* 2013b). Population structure was assessed on both putatively adaptive and neutral markers identified with two genome scan approaches. Our second goal was to test the power of the database for assigning individuals to (i) regional groups defined using the microsatellite data and (ii) their sampling location of origin. We did so with microsatellite markers and SNP markers separately, with the explicit aim of evaluating the relative performance of the two types of markers. For the SNP data, we also compared the performance of outliers vs. putative neutral markers in population assignment.

Materials and methods

Microsatellite genotyping and analysis

A total of 9142 anadromous individuals from 149 sampling locations from the entire North American range of the species were genotyped at 15 microsatellite markers (Fig. S1 and Table S1, Supporting information). Life stages sampled include parr, smolts, and in most cases, returning adults and the samples were collected between 2000 and 2010 (Table S1, Supporting information). The data represented a combination of previously analysed data sets (see Bradbury *et al.* (2014a) and Dionne *et al.* (2008) for regional analyses and further details of genotyping) and new data (Table S1, Support-

ing information). Note that some populations included have been supplemented by hatchery-reared fish and/or have been hybridized with farmed escapees (Carr *et al.* 2004; Bourret *et al.* 2011). While this will necessarily influence our results, our goal is to document present-day population structure regardless of its causes. Individuals were genotyped using an ABI 3130xl (or standardized from ABI 3100 following Gauthier-Ouellet *et al.* 2009) by three independent laboratories (Table S2, Supporting information). Allele scoring was standardized across laboratories in the Laval University laboratory using a panel of 10 standard individuals. Additional individuals ($n = 87$ total; 46 individuals from Newfoundland and Labrador; 41 from Nova Scotia and New Brunswick) were genotyped to account for unrepresented allele values and assess the entire range of allele sizes. To account for discrepancies between ABI 3100 and ABI 3130xl, a panel of 64 individuals representing all allele sizes of Québec populations previously genotyped on the ABI 3100 were analysed on the ABI 3130xl. Rescreening of samples or re-analysis of allele sizes was completed when discrepancies existed between laboratories or machines. Scoring patterns among laboratories were generally consistent and allowed standardization using simple rules (see Table S2, Supporting information).

We tested for departures from Hardy–Weinberg equilibrium (HWE) using the least-square method based on the AMOVA F_{IS} implemented in GENODIVE (Meirmans & van Tienderen 2004) with 999 permutations. We tested for linkage disequilibrium (LD) between each pair of loci in each population using Genepop (1000 iterations per batch for 100 batches; Raymond & Rousset 1995). We adjusted the *P*-values of the LD and HWE tests for multiple comparisons according to the false discovery rate (FDR) method in the function 'p.adjust' in R with an experiment-wide alpha of 0.05 (R Development Core Team 2006). Observed (H_O) and expected heterozygosity (Hs; Nei 1987) were calculated in GENODIVE (Meirmans & van Tienderen 2004). Allelic richness was calculated for each sampling location using the rarefaction method implemented in HP rare (Kalinowski 2005). The Moisie River sample was removed from this analysis because it is missing data at three loci. The HP-rare analysis was repeated twice: once on all the sampling locations (except Moisie; minimal sample size of 16) and once with the KC6 and Georges Rivers samples removed because these two locations had small sample sizes and many missing values, respectively (minimal sample size of 36). Pairwise F_{ST} values were calculated in GENODIVE, significance was assessed using 999 permutations, and *P*-values were adjusted for multiple comparisons with the FDR method.

SNP-array genotyping, filters and basic statistics

A total of 1080 anadromous individuals from 50 sampling locations across the North American range (Table S3, Supporting information) were genotyped at 5568 loci using the SNP array developed by the Centre for Integrative Genetics (CIGENE, Norway) following the manufacturer's instructions (Illumina, San Diego, CA, USA). The data from the Québec populations were previously published in (Bourret *et al.* 2013a), but data from the other regions are previously unpublished. Detailed methods for SNP discovery and quality control can be found in Bourret *et al.* (2013a). Ascertainment bias for the North American populations is minimal (Bourret *et al.* 2013b) and should not bias our results because most markers on the chip were discovered from populations not considered here (European). The quality of individual samples was assured by only using individuals genotyped at a >0.95 call rate (*CR*: proportion of genotyped SNPs). Markers with minor allele frequencies less than 1% (MAF < 0.01) across all populations and markers missing in more than 5% of individuals were excluded from our analyses. Observed (H_O) and expected heterozygosity (H_S; Nei (1987)) were calculated in GENODIVE (Meirmans & van Tienderen 2004). As with the microsatellite data, we tested for departures from HWE GENODIVE with 999 permutations and adjusted the *P*-values for FDR in R. Pairwise F_{ST} values were calculated in GENODIVE, and significance was assessed using 999 permutations.

Outlier markers detection

We used several alternative genome scan methods to detect loci with greater than expected levels of divergence among regional groups. First, we used BAYESCAN, a Bayesian approach that allows the estimation of the posterior probability of a given locus being under the effect of selection (Foll *et al.* 2008). It is based on the multinomial-Dirichlet model, and assumes that allele frequencies among demes are correlated through a common migrant gene pool, therefore allowing complex ecological scenarios to be modelled satisfactorily (Foll *et al.* 2008). Because previous studies suggested that local adaptation is probably important at the regional scale (Dionne *et al.* 2008; Bourret *et al.* 2013a), we ran BAYESCAN on the entire data set by defining groups of individuals on the basis of the microsatellite-defined regional groups (not on populations; Table S1, Supporting information and results) using all the defaults (tests runs with longer chain parameters gave identical results). Loci putatively under selection were defined as those with alpha-values significantly >0 (i.e. with *Q*-values smaller than 0.05) while loci putatively under

balancing selection had alpha-values significantly smaller than 0. All other loci were considered neutral. Second, we used hierarchical Fdist (Excoffier *et al.* 2009) implemented in Arlequin 3.5 (Excoffier & Lischer 2010). This method uses a hierarchical island model allowing for lower migration rates among groups compared to among populations within groups. This has been shown in some instances to reduce the prevalence of false positives compared to methods using the finite island model (Excoffier *et al.* 2009; Mita *et al.* 2013). We used the regional groups defined using the microsatellites (Table S1, Supporting information and results) as the higher level of population structure. Third, as an alternative way to identify loci that maximize assignment to the sampling location of origin, we conducted a 'nested Fdist' analysis, where we used a nonhierarchical Fdist on each microsatellite-defined regional groups separately, and combined all the unique markers identified by each of the individual analyses. The goal with the nested analysis was to select markers that would maximize assignment power to the population and is thus only discussed in this context. In both hierarchical and nested Fdist analyses, loci with significantly (at the 0.01 significance level) higher F_{CT} or F_{ST} values were classified as outliers potentially under directional selection among regional groups, loci with significantly lower F_{CT} or F_{ST} values were loci putatively under balancing selection, and all other loci were classified putatively neutral loci. The results of both the Bayescan and hierarchical Fdist analyses were plotted against the estimated genomic position of a subset of SNP loci for which map data were available (Brenna-Hansen *et al.* 2012) to assess whether specific genomic regions or linkage groups showed elevated amounts of divergence.

Population structure

Initial tests using the widely used Bayesian clustering program STRUCTURE (Pritchard *et al.* 2000) determined that this approach was too computationally demanding even for the microsatellite data set. Instead, we used BAPS (Corander *et al.* 2003, 2004, 2008), a model-based Bayesian clustering approach that, like STRUCTURE, infers genetic groupings that maximize Hardy–Weinberg and linkage equilibrium (Corander *et al.* 2003). Unlike STRUCTURE, however, BAPS infers the optimal number of clusters directly (instead of relying on ad hoc measures; Corander *et al.* 2004) and is computationally more efficient, therefore facilitating the analysis of large data sets. Because in this analysis we were interested in regional clustering of sampling locations, we used the 'clustering of groups of individuals'

option in BAPS. This prior is justified by the homing behaviour of Atlantic salmon (Stabell 1984; Hendry *et al.* 2004). For the microsatellite data set, 20 replicate runs of BAPS were performed with a maximum number of *K* of 149 (the number of samples in the data set). Runs with smaller values for maximum *K* (10, 20, 30, 40 and 50) were also performed to explore the effects of this prior on the final results, which were always identical.

After running BAPS on the entire data set, we examined the possibility of further hierarchical population structure, by running both BAPS and STRUCTURE on the regional groups defined by BAPS with the full data set. For these analyses on each regional groups, STRUCTURE was run with the admixture model with correlated allele frequencies, 50 000 burn-in and 100 000 MCMC repetitions, *K* 1–10 with 10 iterations of each *K*. Alpha-values were examined to ensure that convergence had been attained after the burn-in (Pritchard *et al.* 2000). For the BAPS analysis, we ran 20 independent runs with max *K* = 50 on each of the regional groupings separately.

For the SNP data set, BAPS was run on all SNPs, on the neutral SNPs identified by Bayescan, on the outlier SNPs identified by Bayescan and on the outlier SNPs identified by Fdist. In all cases, 20 independent runs were performed with a maximum number of *K* of 50 (again, the number of sampling locations).

Finally, to infer the relationship of regional groups among each other, we constructed an unrooted neighbour-joining tree of Cavalli-Sforza chord distances for the microsatellite data using the software package Phylip (Felsenstein 1993). The Moisie was excluded from this analysis because it is missing data at three loci. We used 1000 bootstrap replicates to assess confidence in the nodes.

Individual assignment to regional groups and river of origin

We used two approaches to test and compare the relative power of the microsatellite and SNP data sets to assign individuals to their region of origin and their sampling location of origin. For the definition of regional groups, we used two hierarchical levels of population structure defined by the microsatellites (see Results). First, we used the leave-one-out approach implemented in ONCOR (Kalinowski *et al.* 2007). This approach sequentially removes each individual from the baseline and estimates its origin by determining the sampling location that has the highest probability of producing its genotype. Assignment success is then estimated as the proportion of individuals that were correctly assigned to their sampling location of origin. This

approach does not allow for missing data. We therefore filled the missing data by drawing alleles randomly from the entire data set in GENODIVE (Meirmans & van Tienderen 2004). This will reduce overall population differentiation and will in most cases provide conservative estimates of power of assignment. Second, we used the assignment test option in GENODIVE, which implements the approach of Paetkau *et al.* (1995, 2004). The assignment tests were performed on two different microsatellite data sets and on four different SNP data sets. For the microsatellites, analyses were first conducted on the entire database. However, to minimize biases due to sampling effects in the comparisons between SNPs and microsatellites, we combined the different sampling locations that represented several tributaries of the same river (see Supporting information for details). The second microsatellite data set tested only included the sampling locations for which we had SNP data. This was performed to further ensure comparability between the microsatellite and SNP assignment tests in case the inclusion of many genetically similar populations in the full microsatellite data set biased assignment success. For the SNPs, our goal was to compare the power of the entire data set to other data sets, which include only outlier loci and which have been found elsewhere to be sufficient for powerful assignment (Russello *et al.* 2011). We therefore performed assignment tests on a data set including all loci (*n* = 3192), using only the outliers from the Bayescan analysis (*n* = 106), the outliers from the hierarchical Fdist analysis (*n* = 61) and from the nested hierarchical Fdist (*n* = 293). We further evaluated the effect of the number of SNPs included in the analysis on the power of assignment tests. We tested data sets with 50, 100, 250, 500 and 1000 randomly selected SNP markers and performed assignment tests in ONCOR for three replicate data sets for each number of random SNPs at both the population and regional level. In all cases, missing values were filled in GENODIVE according to the overall allele frequencies. Note that using the same samples to identify outliers and to assess their power of assignment can results in upwardly biased estimates of power (Anderson 2010). Our results, however, show that adding those outliers did not improve our assignment success compared to using all other markers (see Results), and the potential upward bias in power only reinforces this conclusion.

Results

Microsatellites

The microsatellite loci were highly polymorphic, with the number of alleles per locus ranging from 14 (D486) to 91 (SsaD144) (average 40; total 593). Missing values

were rare (1.7% overall) and mainly the result of three loci (D486, D58, 2201) having a high failure rate for three populations (Moisie, George and Muddy Bay rivers; 22.2%, 16.9% and 10.1% missing values across all loci, respectively). We found evidence for significant departures from HWE in 230 of 2295 tests (at $\alpha = 0.05$), but none remained significant after adjustment for false discovery rate (FDR). Of the 230 significant tests before adjustments, 152 were heterozygote deficits. These heterozygote deficits were distributed more or less equally across populations and loci: the population with the most loci with heterozygote deficits (NL-SLR) had six and the loci with the most populations in heterozygote deficits (SsaD144) had 17. We also found evidence for significant LD in 1773 of 15 863 tests (at $\alpha = 0.05$), but only 471 remained significant after adjustment for FDR. No pair of loci was consistently found to be in significant LD, but many populations showed elevated numbers of pairs of loci in significant LD. This was particularly true for the Nova Scotia populations (prefix NS in Table S1, Supporting information) that accounted for 226 of the 471 significant tests. Nova Scotia populations have been particularly affected by recent declines in abundance (COSEWIC 2011), which may partly explain this pattern. Pairwise population differentiation was moderate (average $F_{ST} = 0.043$), but values observed were typically greater than zero: only 45 pairwise F_{ST} estimates of 11 628 tests were not statistically significant (at $\alpha = 0.05$ and after FDR adjustment). Nonsignificant tests mostly involved sampling locations with small sample sizes (GU-KC6 had 23 nonsignificant tests), or sampling locations found in the same watershed (e.g. the Miramichi watershed). Notably, comparisons involving sampling locations on Anticosti were either not significantly different (QC-CH was not different from both QC-SU and QC-JU) or marginally so ($P = 0.049$ for the QC-SU vs QC-JU comparison).

SNP-array genotyping, filters and basic statistics

A total of 3192 loci were retained for the analyses. We found evidence for significant departures from HWE in 2467 of 159 600 tests (at $\alpha = 0.05$), but none remained significant after adjustment for false discovery rate (FDR). Departures appeared roughly equally distributed among populations and loci. Expected heterozygosity averaged 0.179 across all populations, but varied by almost twofold within population: from 0.128 (NS-MED) to 0.215 (NL-RKR) (Table S3). Average heterozygosity also varied among regions and was highest in the Avalon Peninsula populations (0.215) and lowest on the east coast of Nova Scotia (0.138). All pairwise population differentiation estimates were statistically significant ($P < 0.01$), with an average pairwise F_{ST} of 0.108,

and ranging from 0.004 (QC-MAT vs. QC-SA) to 0.222 (QC-AF vs. GU-CRO).

Population structure: microsatellites

BAPS identified $K = 29$ as the most likely number of clusters (Fig. 1). The clusters were consistent with geography and are largely consistent with previously published regional groups (Dionne et al. 2008; Bradbury et al. 2014a). Of the 29 clusters, 17 were singletons, that is populations that clustered in their own cluster (identified with asterisks in Fig. 1). Note that the only population from the United States (Penobscot River) clustered on its own, which is consistent with previous studies suggesting U.S. populations are distinct from Canadian populations (Spidle et al. 2001). Of the other 12 clusters, 5 consisted of pairs of populations (one pair in Labrador, two pairs in the inner Bay of Fundy, one pair grouping two populations in Prince Edward Island and one pair on the Avalon Peninsula). The seven remaining clusters identify major regional groups: St. Lawrence North Shore, Gaspésie/Anticosti, Gulf of St. Lawrence, Nova Scotia east coast, Newfoundland, Labrador and Ungava. A detailed description of the regional groupings is provided in the Supporting information.

The neighbour-joining tree of chord distances was largely consistent with the results of the BAPS analysis (Fig. 2). Few of the groups, however, were monophyletic, and none of the internal branches defining regional groups were well supported by bootstrap values. The location of the singleton populations identified by BAPS in the tree was consistent with geography (except for GU-KC6; but this populations had the smallest sample size with $n = 9$), and we therefore include them in the nearest regional clusters for all further analyses on regional groups.

We also identified further levels of hierarchical population structure within some of the regional groupings defined by the BAPS analysis (details in Supporting information). A STRUCTURE analysis provided evidence of two subregional groups within the St. Lawrence North Shore regional grouping (upper vs. lower; Fig. 1 and Supporting information), and both STRUCTURE and BAPS showed that the three Anticosti populations were distinct from the Gaspésie populations (Fig. 1 and Supporting information). There was also some evidence for a separation between the north and south portions of the Nova Scotia east coast regional group, but BAPS and STRUCTURE disagreed regarding the location of the boundary (see Supporting information for details). In all other regions, there was no clear evidence for further hierarchical structuring, thus suggesting that the previous analysis was sufficient in
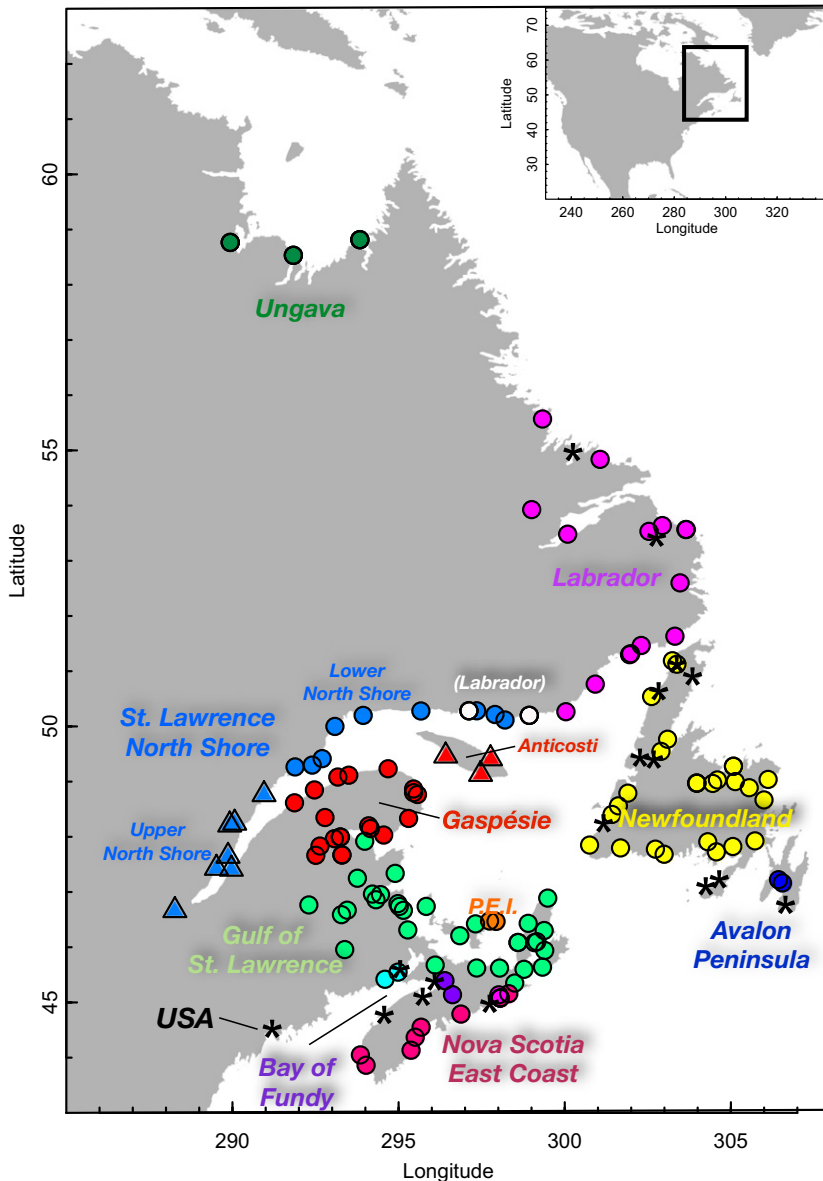
Fig. 1 North American-wide population structure of anadromous Atlantic salmon showing the 149 sampling locations analysed using the microsatellite data set. Locations correspond to the mouths of sampled rivers; in cases where locations are inland, they correspond to a tributary where it intersects the main stem. The sampling locations are coloured according to the results of the BAPS clustering analysis. BAPS clusters containing only one sampling location are denoted by asterisks (*). Two of the BAPS groupings were further divided based on the hierarchical population structure analysis and are shown with different symbols (triangles and circles). Major regional groups discussed in the text are labelled: (1) upper North Shore of St. Lawrence River, (2) lower North Shore of St. Lawrence River, (3) Gaspésie Peninsula, (4) Anticosti Island, (5) Gulf of St. Lawrence (including Prince Edward Island (P.E.I.), see text for details), (6) Nova Scotia east coast, (7) Newfoundland (excluding Avalon Peninsula), (8) Labrador, (9) Bay of Ungava, (10) inner Bay of Fundy (comprised of two BAPS clusters; see text for details), (11) Avalon Peninsula, (12) United States (only one sample location included in the analysis). Further details on the regional group designations can be found in the Supporting information.

capturing population structure, at least at the regional scale (see Supporting information for details).

Outlier markers detection

The Bayescan analysis conducted on the regional groups identified 106 loci putatively under divergent selection, 1413 loci under balancing selection and 1673 putatively neutral loci. The hierarchical Fdist analysis identified 61 loci putatively under directional selection, 32 loci putatively under balancing selection and 3099 putatively neutral loci. Of the 61 loci putatively under directional selection, 36 were also identified as such by Bayescan (Fig. 3). The outlier loci were distributed fairly evenly across the genome and were mapped to several

linkage groups (Fig. 3). The nested analysis in Fdist identified a total of 293 loci putatively under directional selection within at least one of the regional groups.

Population structure: SNPs

The BAPS analyses conducted on different subsets of the SNP data sets did not substantially differ from the analysis conducted on the microsatellite data set (Fig. 4). When all SNPs were used, BAPS uncovered 15 genetic clusters, three of which were formed of a single population. These clusters were very similar to the one based on microsatellites. Consistent with the hierarchical analysis performed on microsatellites, the SNP analysis identified two clusters in the St. Lawrence North Shore and concluded that

Fig. 2 Un-rooted neighbour-joining (NJ) tree of Cavalli-Sforza Chord distances obtained using software Phylip (note that the Moisie River population is not included because it is missing data for all individuals sampled at three loci, something Phylip does not allow). Groups of populations belonging to the same regional cluster as identified in BAPS are highlighted with the same colours as in Fig. 1. Populations identified as singletons in BAPS are still denoted with the asterisk. Bootstrap values (1000 replicates) above 50% are shown. The stars denote the only geographically inconsistent relationships (GU-KC6 and NL-BVB; see text for details). Black dotted lines group sampling locations that form regional groups on the basis of geographical proximity and genetic similarity but that did not group together in the BAPS analysis. Coloured dotted lines group sampling locations that were found to form distinct clusters in the hierarchical analysis of population structure: Anticosti and the two subregional groups within the St. Lawrence North Shore regional group.

the three Anticosti populations cluster apart from the Gaspésie sampling locations. Unlike the microsatellite analysis, however, the SNP analysis divided the Newfoundland sampling locations in two clusters, with the Northern Peninsula populations appearing distinct. The analysis on neutral loci (Bayescan) identified 20 clusters, but the increase was only due to more numerous singletons (8). The clusters formed of more than one populations were otherwise identical to those with all SNPs. Surprisingly, the analysis based on the Bayescan outlier loci gave identical results to that based on neutral SNPs. Using the hierarchical Fdist outliers, 20 clusters were

identified, 7 of which were singletons. The clusters observed were similar to those from the other analyses, with the following differences: the Avalon Peninsula cluster disappeared; two new clusters of two locations were identified in the Labrador cluster and the Gulf of St. Lawrence cluster.

Individual assignment to regional groups and river of origin

The analyses conducted using ONCOR (ON) and GENODIVE (GD) gave comparable results (Fig. 5).
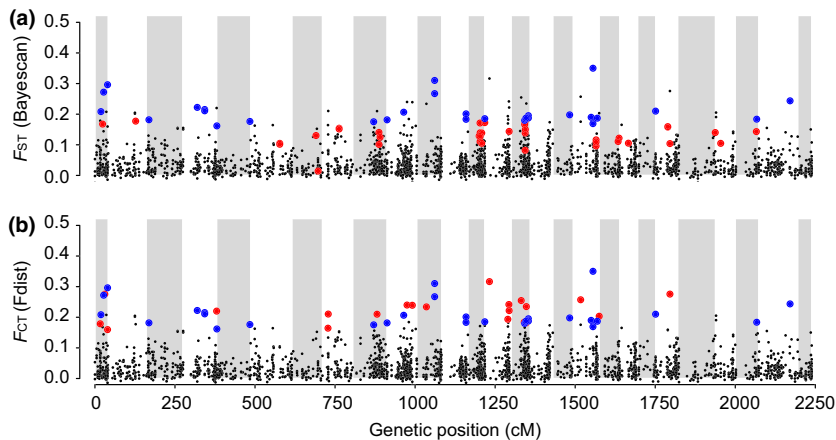
Fig. 3 Genome scan results for the 2607 SNP loci genotyped in this study for which genetic map position is available (out of a total of 3192). (a) Among-region F_{ST} per locus based on the Bayescan analysis. (b) Among-region F_{CT} per locus based on the hierarchical Fdist analysis. In both panels, black circles are putatively neutral loci (or loci putatively under balancing selection), red circles are loci identified as outliers by each analysis alone, and blue circles are loci identified as outliers in both analyses. Alternating white and shaded background shows boundaries between linkage groups.

Assignment success was high for all sets of markers at both hierarchical levels of regional groupings explored, and we therefore only discuss the results on the level of the region. Microsatellites achieved a reasonable level of success in assigning individual to their region of origin both when all sampling locations were included (ON: 77%; GD: 64%) and when only the sampling locations with comparable SNP data were included (ON: 76%; GD: 73%). Note that the reduction in the number of sampling locations only improved the results of the GENODIVE analysis. Assignment success to the regional level was higher than 80% for all the SNP data sets examined and was maximal when all SNPs were included in the analysis (ON: 100%; GD: 97%). Assignment success to the population of origin was low for microsatellite loci when all populations were included (ON: 51%; GD: 52%), but increased to a level comparable to that achieved with the outlier SNP when we used only the populations for which SNP data were available (Fig. 5). Filling missing data with overall allele frequencies did not seem to have a large influence on the ON-COR assignment success, even for populations with large amounts of missing data (e.g. Moisie River: ON: 32%; GD: 32%; George River: ON: 66.7%; GD: 77%). Assignment success to the population level also increased for the outlier loci from the nested genome scan (ON: 85%; GD: 84%) and was best when all loci were included (ON: 94%; GD: 97%).

Evaluation of the number of SNP loci included in the analysis on power of assignment suggested that the SNP data set was sufficiently large to reach maximal power (Fig. 6). For randomly selected loci, assignment success at the regional level reached a plateau at around 250 loci (~95% assignment success), while at the population level, a plateau in assignment success seemed to be attained around 500 loci. Interestingly, using outlier loci only seemed to result in marginal increases in power compared to randomly selected loci.

Indeed, the use of outlier loci only resulted in important increases in assignment success when the number of randomly selected loci was less than about 100 (Fig. 6).

Discussion

We used a conservation genomics framework to inform the definition of conservation units for anadromous Atlantic salmon in North America. We first analysed the most extensive microsatellite data set generated to date for the species to provide a North American-wide assessment of population structure based on neutral markers. This analysis confirmed with confidence previous regional assessments of population structure while also providing interesting nuances. Contrary to expectations, however, the use of SNPs (and especially outlier SNPs) did not substantially change the overall population structure identified. This result contrasts with several previous studies, and we discuss technical and biological factors that may explain these discrepancies. We argue that such patterns may be common and that putatively adaptive markers identified from genome scans are unlikely to provide a practical alternative to the use of ecological or phenotypic data when defining conservation units. The SNP data set, on the other hand, did provide greatly increased assignment success compared to microsatellites, achieving assignment success of up to 100% at the regional level and more than 90% at the population level. We discuss the utility of these tools in management, as well as strategies to achieve optimal assignment power with minimal costs.

### *Population structure: microsatellites vs. neutral SNPs*

The re-analysis of North American-wide population structure based on microsatellites did not substantially differ from previous assessments that only considered parts of the range (McConnell *et al.* 1997; Spidle *et al.*
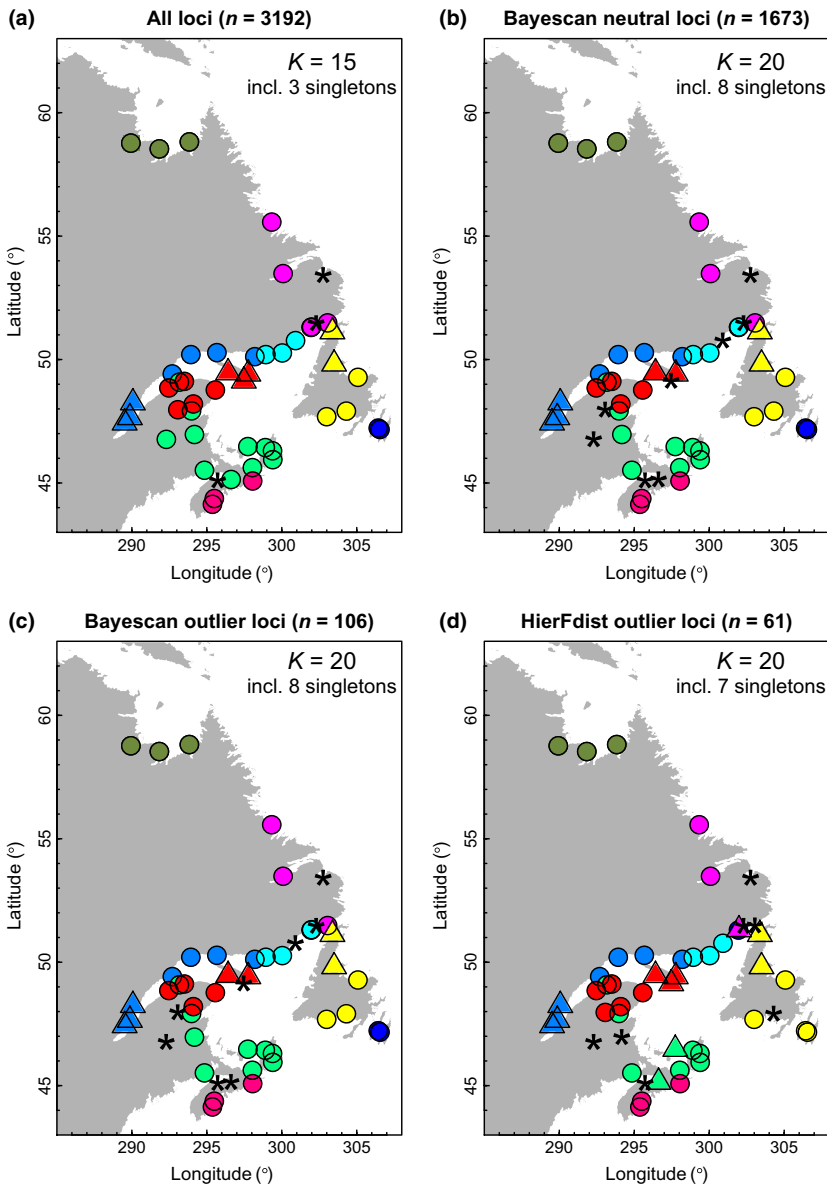
Fig. 4 North American-wide population structure of Atlantic salmon in 50 sampling locations inferred using BAPS from four different SNP data sets including: (a) all SNP loci retained after filters, (b) the putatively neutral markers from the Bayescan analysis, (c) the putatively adaptive loci from the Bayescan analysis, and (d) the putatively adaptive loci from the hierarchical Fdist analysis. The colours and shapes of the symbols denote the different BAPS-determined genetic clusters (we used the same colour-coding as in Fig. 1 and used triangles for new clusters to highlight differences with the microsatellite-based analysis). Asterisks denote clusters that contain only a single sampling location (i.e. singletons). The number of clusters inferred from the BAPS analysis is indicated in the upper-right corner of each panel as well as the number of singleton populations.

2001; Dionne *et al.* 2008; Bourret *et al.* 2013a; Bradbury *et al.* 2014a). Contrary to the analysis of Dionne *et al.* (2008) for the Québec populations, our initial analysis on the entire data set clustered all St. Lawrence North Shore sampling locations together and grouped the Anticosti locations with the other Gaspésie locations. The differences, however, were resolved when the hierarchical population structure analysis was conducted and unambiguously recovered these groupings, as well as when SNP loci were used for clustering. Multiple levels of hierarchical population structure are common in anadromous Atlantic salmon (Vähä *et al.* 2007; Perrier *et al.* 2011). Note also that the three populations from Anticosti Island are less genetically differentiated than other comparably distant populations, adding confi-

dence in the reality of this genetic grouping. Our analysis was also consistent with that of Bradbury *et al.* (2014a) and Palstra *et al.* (2007) in showing that the Labrador and Newfoundland populations were distinct, and that within Newfoundland, the Avalon Peninsula sampling locations were also distinct. Several hypotheses were suggested by Palstra *et al.* (2007) to explain the distinctiveness of the Avalon Peninsula samples, including smaller population sizes, differences in run timing and increased prevalence of mature male parr (Myers *et al.* 1986). Both studies, however, found evidence for further genetic regional groupings within Newfoundland (Palstra *et al.* 2007; Bradbury *et al.* 2014a), something our microsatellite analysis, including the hierarchical population structure analysis, failed to
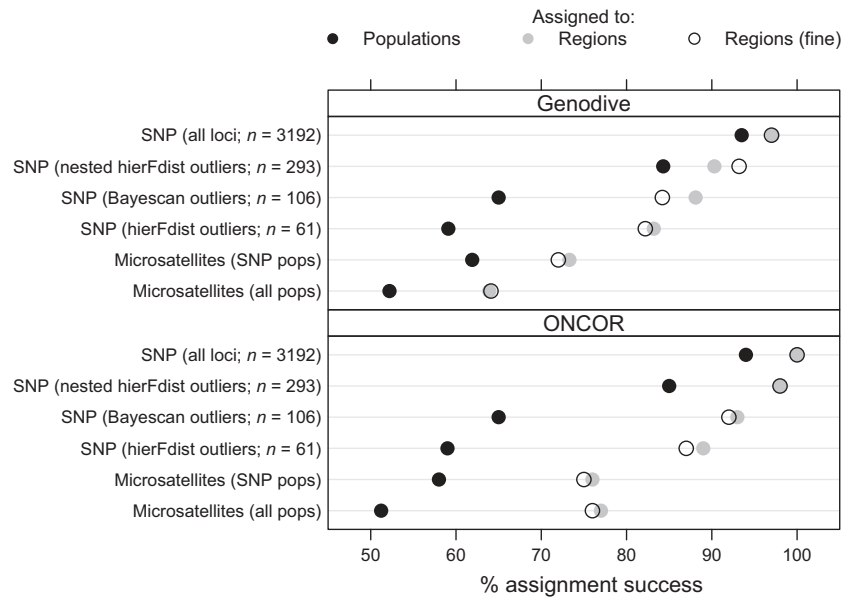
Fig. 5 Dot plots showing the results of the assignment tests performed in GENODIVE (top) and ONCOR (bottom) for different sets of markers. The x-axis represents the proportion of individuals that has the highest likelihood of being assigned to the sampling location where they were sampled. Assignment tests were performed at the population level (filled black symbols), and the regional level (filled grey symbols) and at a finer scale of regional differentiation (hollow symbols; see methods for details). Assignment tests were performed using the entire microsatellite data set ('all pops'), a subset of the microsatellite data set including only populations for which we also had SNP data ('SNP pops'), a data set including all the available SNP markers ('all loci'; $n = 3192$), the Bayescan outliers ($n = 106$), the hierarchical Fdist outliers ($n = 61$), and the nested Fdist outliers ($n = 293$).
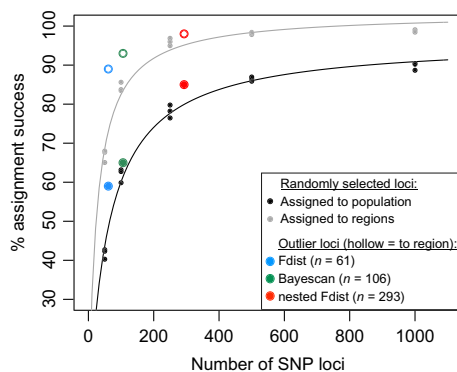
Fig. 6 Effect of the number of randomly selected SNP loci on assignment success at the population (small black circles) and regional levels (grey dots). Assignment success for the three sets of outlier loci previously used is also shown for comparison, with the full symbols representing assignment to the population, and the hollow symbols representing assignment to the regional: Fdist outliers (blue, $n = 61$); Bayescan outliers (green, $n = 106$), nested Fdist outliers (red, $n = 293$). Michaelis–Menten curves were fitted to the data from randomly selected markers to help visualization.

uncover. Another notable result is the increased prevalence of singletons in Newfoundland from the BAPS analysis, which appears consistent with previous conclusions that drift was one of the most important forces

structuring populations in Newfoundland (Bradbury *et al.* 2014a). Our analysis was also consistent with previous studies in suggesting the inner Bay of Fundy populations were evolutionarily distinct, most likely due to a different recolonization history following the last glaciation (McConnell *et al.* 1997; Verspoor *et al.* 2002; Vandersteen Tymchuk *et al.* 2010). While the populations did not cluster together in the BAPS analysis, they were clearly related in the NJ tree (Fig. 2). The BAPS analysis, in fact, suggested that two clusters were present, a result consistent with previous assessments that found differentiation between Chignecto Bay (NS-PWF and NS-BSR) and Minas Basin (NS-GRV and NS-STW) populations (Verspoor *et al.* 2002; Vandersteen Tymchuk *et al.* 2010). Other sampling locations clustered as singletons, which was perhaps explained by high drift due to small population sizes. Effective numbers of breeders of 12 anadromous Atlantic salmon populations from Québec have been found to vary between 48 and 876 (mean = 252) (Perrier *et al.* 2013). If typical, such small effective sizes could result in important genetic drift. The Prince Edward Island (PEI) rivers mainly clustered with the other Gulf of St. Lawrence populations, which is expected given that many of these rivers have been stocked heavily with Miramichi River fish (Cairns *et al.* 2010). The fact that two rivers clustered apart in the BAPS analysis, however, suggest that some locations

may retain ancestral allele frequencies and that PEI may have constituted a different genetic group prior to stocking. Finally, our analysis included only a single population from the United States, but its clustering in a group by itself was consistent with previous results suggesting US populations (including the Penobscot River) were genetically differentiated from Canadian populations (Spidle *et al.* 2003).

In short, the analysis of the microsatellite data set supported the definition of 12 genetic regional groups for anadromous Atlantic salmon in North America: (1) upper North Shore of St. Lawrence River, (2) lower North Shore of St. Lawrence River, (3) Gaspésie Peninsula, (4) Anticosti Island, (5) Gulf of St. Lawrence (i.e. New Brunswick, Nova Scotia, Prince Edward Island), (6) Nova Scotia east coast, (7) Newfoundland (excluding Avalon Peninsula), (8) Labrador, (9) Bay of Ungava, (10) inner Bay of Fundy, (11) Avalon Peninsula and southeast Newfoundland, (12) United States (Penobscot River) (Fig. 1). Detailed justifications for these conclusions can be found in the Supporting information. These genetic regional groups differ in some notable ways from the designatable units (DUs) currently recognized for Atlantic salmon in Canada by COSEWIC (2011). There are currently four recognized DUs in Newfoundland, but the microsatellite analysis suggested only two conservation units for the island (including the Avalon Peninsula, which was not previously recognized as distinct from the other southeast Newfoundland populations). The Gaspé Peninsula and the Gulf populations were previously grouped in a single DU, but our analyses suggested that they are in fact distinct. The eastern Cape Breton Island populations were recognized as a DU, but our analysis suggested that they are part of the Gulf of St. Lawrence conservation unit. Three DUs were recognized on the St. Lawrence North Shore, which is partly consistent with the recognition of a separate group formed of a pair of populations at the boundary between our upper North Shore and Labrador groups. We concluded that these populations should be grouped with the Labrador populations, but we recognize that this grouping is somewhat ambiguous. Note that some of these DUs were defined on the basis of ecological or demographic information and may therefore represent useful conservation units despite the lack of concordance with the genetic data.

The BAPS analysis on the SNP data sets containing all loci and only neutral loci only differed in a minor way from the analysis on microsatellites. The SNP data sets uncovered all the geographical clusters identified with the microsatellites, including the subregional groupings in the St. Lawrence North Shore and Gaspésie/Anticosti that required hierarchical analyses

to be evident with the microsatellite data set. Main differences were the identification of a new grouping along the North Shore close to the Labrador grouping, as identified by Dionne *et al.* (2008) (therein called 'Higher North Shore'), and a grouping on the Northern Peninsula sampling locations in Newfoundland. This latter grouping was consistent with geography, and to a certain extent with the analysis of Bradbury *et al.* (2014a), where the Northern Peninsula rivers appeared distinct in a NJ tree, but less so in a STRUCTURE analysis.

Defining conservation units with adaptive genetic variation

Analysis of population structure on the outlier loci from the Bayescan ($n = 106$) and Fdist ($n = 61$) analyses led to roughly the same results as that conducted on the putatively neutral loci ($n = 1673$). This contrasted with previous results obtained from other species where putatively adaptive markers uncovered population structure at a finer scale (Hohenlohe *et al.* 2010; André *et al.* 2011; Bradbury *et al.* 2013; Hemmer-Hansen *et al.* 2013; Hess *et al.* 2013). The failure of putatively adaptive markers to uncover further levels of population structure in North American anadromous Atlantic salmon may be due to both technical and biological factors. First, studies have shown that results of genome scans should be interpreted with caution and that false positives are likely to be common (Narum & Hess 2011; Mita *et al.* 2013), especially for species that recently recolonized their range and are thus violating assumptions of equilibrium (Lotterhos & Whitlock 2014). If false positives were a random subset of the entire pool of loci, they would not be expected to lead to differences in population structure compared to neutral markers. The fact that we found limited overlap between the identities of outlier loci generated with Bayescan and hierarchical Fdist provided some support for this explanation. Second, outlier loci could be associated with several environmental variables that vary along different geographical axes. Their joint consideration as outlier loci (instead of considering alleles that are associated with specific environmental variables) would then blur potential population structure along a single axis of variation. This could be especially important for North American Atlantic salmon given the complexity of the coastline. Third, there is evidence of potentially polygenic adaptive traits in Atlantic salmon which are underlain by many genes of small effects (Bourret *et al.* 2014), thus making them difficult to identify with a traditional genome scan approach. The even distribution of outlier loci identified in the genome scan analyses presented here lends support to this

explanation. This contrasted with cases from marine fishes where genomic islands of divergence were identified (Hemmer-Hansen *et al.* 2013). Fourth, neutral markers could have been sufficient to represent the scale of local adaptation even if they are not themselves under selection. Previous work on Atlantic salmon concluded that regional groupings identified with neutral markers were associated with environmental differences (Dionne *et al.* 2007; Bourret *et al.* 2013a; Bradbury *et al.* 2014a). This could lead to selection against migrants from different regional groups (Nosil *et al.* 2005). It has also been proposed that straying is more likely to occur among rivers with similar characteristics, for example among rivers with similar chemical cues used for olfaction during homing (Dittman & Quinn 1996; Perrier *et al.* 2011; Bourret *et al.* 2013a). It is possible that selection effectively limits gene flow among regional groups, therefore resulting in similar patterns of allelic variation at both neutral and adaptive markers (Thibert-Plante & Hendry 2010). Finally, many of the studies that found drastically different population structure with outlier vs. neutral markers focused on species with high gene flow and large population sizes (i.e. low genetic drift) such as marine fishes (André *et al.* 2011; Bradbury *et al.* 2013; Hemmer-Hansen *et al.* 2013). In such cases, alleles conferring an adaptive advantage to certain environmental conditions would be expected to (i) increase in frequency rapidly and (ii) spread rapidly to all parts of the range experiencing these conditions. In the case of Atlantic salmon, however, low gene flow and small population sizes (i.e. high drift) could result in much more idiosyncratic patterns of adaptation at the molecular level, perhaps making genome scan approaches less likely to identify the loci underlying adaptation. As many species of conservation concern are likely to be species with small population size and restricted gene flow (e.g. through habitat fragmentation), this could be an important factor to consider for a broad range of species.

In summary, the use of SNPs for delineating conservation units did not change substantially the conservation units defined on the basis of the microsatellite data. This is likely to be a common feature of such data sets and genome scans may not offer the hoped-for solution to the problem of incorporating adaptive information in the definition of conservation units. This also means that the framework proposed by Funk *et al.* (2012) may only be applicable in a minority of cases, unless better methods for identifying adaptive variation from genomic data are developed. In the meantime, the use of genomic data in conjunction with other types of data may offer more potential. Future work could, for instance, identify outlier loci from correlations with environmental data (e.g. Bourret *et al.* 2013a)

instead of from genome scans to alleviate some of the problems discussed above. Alternatively, markers under selection and involved in adaptation may often correspond more to markers that covary but involve only subtle differences in allele frequencies and are thus not detectable by genome scans (Le Corre & Kremer 2012; Bourret *et al.* 2014).

Individual assignments: SNPs vs microsatellites

An important management and conservation concern with Atlantic salmon is the operation of mixed-stock fisheries that target migratory individuals on their feeding grounds far from their natal rivers (Gauthier-Ouellet *et al.* 2009; Chaput 2012). Determining the origin of harvested salmon is therefore an important management priority. There are two main approaches utilizing genetic tools that can be used by managers to address this question: genetic mixture analysis and assignment tests (Manel *et al.* 2005). Bradbury *et al.* (2014b) have already evaluated the power of the microsatellite database in a mixed-stock fishery context and found an average mixture analysis accuracy of 97.2% at the regional level. In some cases (e.g. management actions required at the scale of a single river, forensics), however, a mixture approach evaluating the proportional representation of regional groups may be insufficient. We therefore focused on the problem of assigning individuals to their region and river of origin. We were aided in this by intensive sampling of all possible regions of origin of Atlantic salmon in North America.

The use of a large number of SNP loci greatly increased assignment success in our study compared to microsatellites. Because microsatellites have far more alleles per locus than SNPs, previous studies based on a small number of SNPs rarely outperformed microsatellites in the context of population or parental assignment (Narum *et al.* 2008). Our use of >3000 SNP loci definitively showed that numerous SNPs could outperform microsatellites, probably a result of both an increase in the number of independent loci and in the total number of alleles (Bernatchez & Duchesne 2000). Very few studies have used large SNP data sets in assignment tests. In one example, Larson *et al.* (2014) were able to assign individual Chinook salmon to their river of origin with 91% to 100% success using 10 944 RAD-generated SNPs. The study, however, only included five populations from a restricted portion of the Chinook salmon's range. In the present study, we demonstrate comparable success rate in assignment with a far greater number of populations distributed throughout the geographical range of Atlantic salmon in North America.

Despite high assignment success with the entire SNP data set, assignment success based on outlier SNPs was only marginally better than that based on a similar number of randomly selected SNPs. At first glance, this result contrasted with several existing studies that have demonstrated increased assignment success with outlier SNPs compared to neutral markers (O'Malley *et al.* 2007; Ackerman *et al.* 2011; Freamo *et al.* 2011; Russello *et al.* 2011). Most of these studies, however, used a small number of loci (<100). For instance, Freamo *et al.* (2011) found that 14 outlier SNPs were more accurate than 67 neutral SNPs at assigning Bay of Fundy Atlantic salmon to their group of origin (outer vs. inner Bay of Fundy). At a small number of loci, we also find that using outlier SNPs could perhaps improve assignment success: assignment to the population was ~20% higher when using the 61 Fdist outlier SNPs than when using a random subset of 50 SNPs. Few studies using a comparable number of SNPs for assignment exist to compare our results. Glover *et al.* (2010) used 300 SNPs in Atlantic salmon and found that the use of outlier SNPs did not result in great increases in assignment success when the overall number of SNPs used was >100. Although it has been shown that some methods used to select markers for population assignment may result in upwardly biased estimates (Anderson 2010), the method we used here seems unlikely to play a major role in determining assignment success. Indeed, the fact that using outliers obtained from different genome scan methods led to similarly marginal increase in assignment success over random loci suggested that it had little bearing on the end result. Other methods exist for selecting markers that provide increased assignment power. For instance, software BELS (Bromaghin 2008) appeared to select loci that increased assignment success in Atlantic salmon, but performed poorly with more than 100 markers (Glover *et al.* 2010).

In summary, managers wishing to decrease costs by selecting a small subset of SNPs can carefully select highly differentiated markers to increase assignment success relative to a comparatively sized random subset of SNPs, but they are unlikely to attain the assignment success possible with a large number of SNPs. The use of the present medium-density SNP chip for assignment of a large number of individuals will, however, entail significant financial costs. If the goal is to assign individuals to their river of origin, our analysis suggest that a large number of SNPs are required (>1000), and we suggest that using low-cost genotyping by sequencing technology (Davey *et al.* 2011) is probably the better strategy to lower costs. If the goal is to assign individuals to their region of origin, our analysis showed that using ~100 carefully selected SNPs allows assignment

success of >90%. Investing in the development of a lower-density SNP array, such as the Fluidigm array used extensively in Pacific salmon (Ackerman *et al.* 2011; Larson *et al.* 2013), could therefore be more advantageous in this case.

Conclusions

The combination of a microsatellite database with high geographic resolution with data from thousands of SNPs allowed us to perform the most complete assessment of population structure of anadromous Atlantic salmon in North America to date. While the results were largely consistent with previous regional assessments, some important differences emerged from the range-wide analysis. Notably, our analysis suggested that some currently recognized conservation units did not coincide with genetic boundaries, something managers will want to take into consideration in future conservation efforts. Contrary to expectations, the definition of conservation units on the basis of microsatellites and SNPs, even those putatively under selection, did not differ all that much. This suggested that microsatellites are still a useful tool for defining conservation units, and that neutral markers may satisfyingly reflect the scale of local adaptation in anadromous salmon even if they are not under selection (e.g. via selection against migrants). Nevertheless, there are still several advantages to using a large number of SNPs. First, SNPs provide a more repeatable platform when genotyping is performed across laboratories (standardization across laboratories for the present article required large efforts; Table S2, Supporting information). Second, they provided increased confidence in the conservation units defined on the basis of microsatellites and were able to recover regional groupings that required hierarchical analyses to identify on the basis of microsatellites alone. Third, the equivalency of the outlier SNPs and the microsatellites for defining conservation units may stem from methodological issues related to the identification of putatively selected SNPs on the basis of genome scans alone. Future work may benefit from using alternative ways of identifying loci under selection, for example with methods that directly correlate allelic frequencies with specific environmental variables (Bourret *et al.* 2013a). There was no doubt, however, that using thousands of SNP loci for assignment of individuals to their population of origin provided greater assignment success compared to analyses based on microsatellites. The geographically extensive microsatellite and SNP databases presented here will therefore provide a diverse and powerful set of tools for managers attempting to achieve various research and management goals.

Acknowledgements

References

Ackerman MW, Habicht C, Seeb LW (2011) SNPs under diversifying selection provide increased accuracy and precision in mixed stock analyses of sockeye salmon from Copper River, Alaska. *Transactions of the American Fisheries Society*, **140**, 865–881.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews: Genetics*, **11**, 697–709.

Anderson EC (2010) Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. *Molecular Ecology Resources*, **10**, 701–710.

André C, Larsson LC, Laikre L *et al.* (2011) Detecting population structure in a high gene-flow species, Atlantic herring (*Clupea harengus*): direct, simultaneous evaluation of neutral vs putatively selected loci. *Heredity*, **106**, 270–280.

Avise JC (2009) Perspective: conservation genetics enters the genomics era. *Conservation Genetics*, **11**, 665–669.

Bernatchez L, Duchesne P (2000) Individual-based genotype analysis in studies of parentage and population assignment: how many loci, how many alleles? *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 1–12.

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

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

Bourret V, Kent MP, Primmer CR, Vasemägi A (2013b) SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Molecular Ecology*, **22**, 532–551.

Bourret V, Dionne M, Bernatchez L (2014) Detecting genotypic changes associated with selective mortality at sea in Atlantic salmon: polygenic multilocus analysis surpasses genome scan. *Molecular Ecology*, **23**, 4444–4457.

Bradbury IR, Hubert S, Higgins B *et al.* (2013) Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. *Evolutionary Applications*, **6**, 450–461.

Bradbury IR, Hamilton LC, Robertson MJ *et al.* (2014a) Landscape structure and climatic variation determine Atlantic salmon genetic connectivity in the Northwest Atlantic. *Canadian Journal of Fisheries and Aquatic Science*, **71**, 246–258.

Bradbury IR, Hamilton LC, Rafferty S *et al.* (2014) Genetic evidence of local exploitation of Atlantic salmon in a coastal subsistence fishery in the Northwest Atlantic. *Canadian Journal of Fisheries and Aquatic Sciences*, doi:10.1139/cjfas-2014-0058.

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.

Bromaghin JF (2008) BELS: backward elimination locus selection for studies of mixture composition or individual assignment. *Molecular Ecology Resources*, **8**, 568–571.

Cairns DK, Guignon DL, Dupuis T, MacFarlane RE (2010) Stocking history, biological characteristics, and status of Atlantic salmon (*Salmo salar*) on Prince Edward Island. *DFO Canadian Science Advisory Secretariat Research Document*, **2010/104**, iv–50p.

Carr JW, Whoriskey F, O'Reilly P (2004) Efficacy of releasing captive reared broodstock into an imperilled wild Atlantic salmon population as a recovery strategy. *Journal of Fish Biology*, **65**, 38–54.

Chaput G (2012) Overview of the status of Atlantic salmon (*Salmo salar*) in the North Atlantic and trends in marine mortality. *ICES Journal of Marine Science*, **69**, 1538–1548.

Chaput G, Legault CM, Reddin DG, Caron F, Amiro PG (2005) Provision of catch advice taking account of non-stationarity in productivity of Atlantic salmon (*Salmo salar* L.) in the Northwest Atlantic. *ICES Journal of Marine Science*, **62**, 131–143.

Corander J, Waldmann P, Sillanpää MJ (2003) Bayesian analysis of genetic differentiation between populations. *Genetics*, **163**, 367–374.

Corander J, Waldmann P, Marttinen P, Sillanpaa MJ (2004) BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics*, **20**, 2363–2369.

Corander J, Marttinen P, Sirén J, Tang J (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics*, **9**, 539.

COSEWIC (2011) Assessment and status report on the Atlantic salmon *Salmo salar* in Canada. Committee on the Status of Endangered Wildlife in Canada, xlvii–136 pp.

Crandall KA, Bininda-Edmonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution*, **15**, 290–295.

Crispo E, Moore J-S, Lee-Yaw JA, Gray SM, Haller BC (2011) Broken barriers: human-induced changes to gene flow and

introgression in animals: an examination of the ways in which humans increase genetic exchange among populations and species and the consequences for biodiversity. *BioEssays*, **33**, 508–518.

Davey JW, Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews: Genetics*, **12**, 499–510.

Dionne M, Miller KM, Dodson JJ, Caron F, Bernatchez L (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, Caron F, Dodson JJ (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, Caron F, Dodson J, Bernatchez L (2009) Comparative survey of within-river genetic structure in Atlantic salmon; relevance for management and conservation. *Conservation Genetics*, **10**, 869–879.

Dittman AH, Quinn TP (1996) Timing of imprinting to natural and artificial odors by coho salmon (*Oncorhynchus kisutch*). *Canadian Journal of Fisheries and Aquatic Sciences* **53**, 434–442.

Dodson JJ, Gibson RJ, Cunjak RA *et al.* (1998) Elements in the development of conservation plans for Atlantic salmon (*Salmo salar*). *Canadian Journal of Fisheries and Aquatic Sciences*, **55**, 312–323.

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

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

Felsenstein J (1993) *PHYLIP (Phylogeny Inference Package), Version 3.5. 1*. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

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

Fraser DJ, Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology*, **10**, 2741–2752.

Fraser D, Cook AM, Eddington JD, Bentzen P, Hutchings JA (2008) Mixed evidence for reduced local adaptation in wild salmon resulting from interbreeding with escaped farmed salmon: complexities in hybrid fitness. *Evolutionary Applications*, **1**, 501–512.

Fraser DJ, Weir LK, Bernatchez L, Hansen MM, Taylor EB (2011) Extent and scale of local adaptation in salmonid fishes: review and meta-analysis. *Heredity*, **106**, 404–420.

Freamo H, O'Reilly P, Berg PR, Lien S, Boulding EG (2011) Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. *Molecular Ecology Resources*, **11**, 254–267.

Friedland KD, Chaput G, MacLean JC (2005) The emerging role of climate in post-smolt growth of Atlantic salmon. *ICES Journal of Marine Science*, **62**, 1338–1349.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, **27**, 489–496.

Garant D, Dodson JJ, Bernatchez L (2000) Ecological determinants and temporal stability of the within-river population structure in Atlantic salmon (*Salmo salar* L.). *Molecular Ecology*, **9**, 615–628.

Garcia de Leaniz C, Fleming IA, Einum S *et al.* (2007) A critical review of adaptive genetic variation in Atlantic salmon: implications for conservation. *Biological Reviews*, **82**, 173–211.

Gauthier-Ouellet M, Dionne M, Caron F, King TL, Bernatchez L (2009) Spatiotemporal dynamics of the Atlantic salmon (*Salmo salar*) Greenland fishery inferred from mixed-stock analysis. *Canadian Journal of Fisheries and Aquatic Science*, **66**, 2040–2051.

Glover KA, Hansen MM, Lien S *et al.* (2010) A comparison of SNP and STR loci for delineating population structure and performing individual genetic assignment. *BMC Genetics*, **11**, 2.

Hemmer-Hansen J, Nielsen EE, Therkildsen NO *et al.* (2013) A genomic island linked to ecotype divergence in Atlantic cod. *Molecular Ecology*, **22**, 2653–2667.

Hendry A, Castric V, Kinnison M, Quinn T (2004) The evolution of philopatry and dispersal: homing Versus Straying in Salmonids. In: *Evolution Illuminated: Salmon and Their Relatives* (eds Hendry A, Stearns S), pp. 53–91. Oxford University Press, New York, USA.

Hess JE, Campbell NR, Close DA, Docker MF, Narum SR (2013) Population genomics of Pacific lamprey: adaptive variation in a highly dispersive species. *Molecular Ecology*, **22**, 2898–2916.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.

Hohenlohe PA, Day MD, Amish SJ *et al.* (2013) Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular Ecology*, **22**, 3002–3013.

Kalinowski ST (2005) hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187–189.

Kalinowski ST, Manlove KR, Taper ML (2007) *ONCOR: A Computer Program for Genetic Stock Identification*. Department of Biology, Montana State University, Bozeman, Montana.

King TL, Kalinowski ST, Schill WB, Spidle AP, Lubinski BA (2001) Population structure of Atlantic salmon (*Salmo salar* L.): a range-wide perspective from microsatellite DNA variation. *Molecular Ecology*, **10**, 807–821.

Kohn MH, Murphy WJ, Ostrander EA, Wayne RK (2006) Genomics and conservation genetics. *Trends in Ecology & Evolution*, **21**, 629–637.

Lamaze FC, Sauvage C, Marie A, Garant D, Bernatchez L (2012) Dynamics of introgressive hybridization assessed by SNP population genomics of coding genes in stocked brook charr (*Salvelinus fontinalis*). *Molecular Ecology*, **21**, 2877–2895.

Larson WA, Utter FM, Myers KW *et al.* (2013) Single-nucleotide polymorphisms reveal distribution and migration of Chinook salmon (*Oncorhynchus tshawytscha*) in the Bering Sea and North Pacific Ocean. *Canadian Journal of Fisheries and Aquatic Science*, **70**, 128–141.

Larson WA, Seeb LW, Everett MV *et al.* (2014) Genotyping by sequencing resolves shallow population structure to inform

conservation of Chinook salmon (*Oncorhynchus tshawytscha*). *Evolutionary Applications*, **7**, 355–369.

Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, **21**, 1548–1566.

Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology*, **23**, 2178–2192.

Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology & Evolution*, **20**, 136–142.

McConnell SKJ, Ruzzante DE, O'Reilly PT, Hamilton L, Wright JM (1997) Microsatellite loci reveal highly significant genetic differentiation among Atlantic salmon (*Salmo salar* L.) stocks from the east coast of Canada. *Molecular Ecology*, **6**, 1075–1089.

McKay JK, Latta RG (2002) Adaptive population divergence: markers, QTL and Traits. *Trends in Ecology & Evolution*, **17**, 285–291.

Meirmans PG, van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792–794.

Mita S, Thuillet AC, 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.

Moritz C (1994) Defining "Evolutionarily Significant Units" for conservation. *Trends in Ecology & Evolution*, **9**, 373–375.

Myers RA, Hutchings JA, Gibson RJ (1986) Variation in male parr maturation within and among populations of Atlantic Salmon, *Salmo salar*. *Canadian Journal of Fisheries and Aquatic Sciences*, **43**, 1242–1248.

Narum SR, Hess JE (2011) Comparison of FST outlier tests for SNP loci under selection. *Molecular Ecology Resources*, **11**, 184–194.

Narum SR, Banks M, Beacham TD *et al.* (2008) Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Molecular Ecology*, **17**, 3464–3477.

Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, New York.

Nosil P, Vines TH, Funk DJ (2005) Perspective: reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution*, **59**, 705–719.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

O'Malley KG, Camara MD, Banks MA (2007) Candidate loci reveal genetic differentiation between temporally divergent migratory runs of Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology*, **16**, 4930–4941.

Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology*, **4**, 347–354.

Paetkau D, Slade R, Burdens M, Estoup A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, **13**, 55–65.

Palstra FP, O'Connel MF, Ruzzante DE (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. *Molecular Ecology*, **16**, 4504–4522.

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

Perrier C, Dionne M, Coté G, Bernatchez L (2013) Évaluation et suivi de l'état génétique des populations de saumon atlantique au Québec: années 2010 à 2012. Québec, Ministère du Développement durable, de lEnvironnement, de la Faune et des Parcs, Direction générale de lexpertise sur la faune et ses habitats, Direction de la faune aquatique, 57pp.

Primmer CR (2009) From conservation genetics to conservation genomics. *Annals of the New York Academy of Sciences*, **1162**, 357–368.

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

R Development Core Team (2006) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Raymond M, Rousset F (1995) GENEPOP (Version 1. 2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Russello MA, Kirk SL, Frazer KK, Askey PJ (2011) Detection of outlier loci and their utility for fisheries management. *Evolutionary Applications*, **5**, 39–52.

Spidle AP, Schill WB, Lubinski BA, King TL (2001) Fine-scale population structure in Atlantic salmon from Maine's Penobscot River drainage. *Conservation Genetics*, **2**, 11–24.

Spidle AP, Kalinowski ST, Lubinski BA *et al.* (2003) Population structure of Atlantic Salmon in maine with reference to populations from Atlantic Canada. *Transactions of the American Fisheries Society*, **132**, 196–209.

Stabell OB (1984) Homing and olfaction in salmonids: a critical review with special reference to the Atlantic salmon. *Biological Reviews*, **59**, 333–388.

Stapley J, Reger J, Feulner PGD *et al.* (2010) Adaptation genomics: the next generation. *Trends in Ecology & Evolution*, **25**, 705–712.

Taylor EB (1991) A review of local adaptation in Salmonidae, with particular reference to Pacific and Atlantic salmon. *Aquaculture*, **98**, 185–208.

Thibert-Plante X, Hendry AP (2010) When can ecological speciation be detected with neutral loci? *Molecular Ecology*, **19**, 2301–2314.

Tonteri A, Veselov AJVJ, Zubchenko AVZV, Lumme J, Primmer CRPR (2009) Microsatellites reveal clear genetic boundaries among Atlantic salmon (*Salmo salar*) populations from the Barents and White seas, northwest Russia. *Canadian Journal of Fisheries and Aquatic Sciences*, **66**, 717–735.

Vähä J-P, Erkinaro J, Niemelä E, Primmer CR (2007) Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. *Molecular Ecology*, **16**, 2638–2654.

Vandersteen Tymchuk W, O'Reilly P, Bittman J, Macdonald D, Schulte P (2010) Conservation genomics of Atlantic salmon: variation in gene expression between and within regions of the Bay of Fundy. *Molecular Ecology*, **19**, 1842–1859.

Verspoor E (2005) Regional differentiation of North American Atlantic salmon at allozyme loci. *Journal of Fish Biology*, **67**, 80–103.

Verspoor E, O'Sullivan M, Arnold AL, Knox D, Amiro PG (2002) Restricted matrilineal gene flow and regional differentiation among Atlantic salmon (*Salmo salar* L.) populations within the Bay of Fundy, eastern Canada. *Heredity*, **89**, 465–472.

Waples RS (1991) Pacific salmon, *Oncorhynchus* spp., and the definition of "species" under the Endangered Species Act. *Marine Fisheries Review*, **53**, 11–22.

Waples RS (1995) Evolutionarily significant units and the conservation of biological diversity under the Endangered Species Act. *American Fisheries Society Symposium*, **17**, 8–27.

Waples RS, Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419–1439.

J.-S.M. performed most of the analyses and wrote the manuscript. V.B. coordinated the sample collection and microsatellite genotyping for the database, including the standardization across laboratories, and participated in data analysis and writing. P.O. also helped coordinate sample collection and provided samples and helped with the writing. M.K. coordinated the SNP genotyping. M.D., I.B. and G.C. provided samples and help with the writing. L.B. designed and supervised the study and was involved in the analysis and writing phases.

Data accessibility

The raw microsatellite and SNP genotypes (GENEPOP format) have been deposited in Dryad: doi:10.5061/dryad.sb601. The different subsets of outlier and neutral SNPs (also in GENEPOP format), the NJ tree file and the PHYLIP infile, the Bayescan parameter file, an example STRUCTURE param file from the hierarchical structure analysis, the genetic map data for the loci, and the code to generate the random loci in Fig. 6 are also accessible on Dryad.

Supporting information

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

Table S1 Sampling locations included in the microsatellite database, along with information on the year and life stages sampled, and summary statistics for the microsatellite data.

Table S2 Standardization rules for comparison of microsatellite genotype data collected by the three laboratories and within Laval University.

Table S3 Pairwise F_{ST} values among all populations included in the microsatellite database.

Table S4 Sampling locations included in the SNP database, along with summary statistics for the SNP data.

Table S5 Pairwise F_{ST} values among all populations included in the SNP database.

Fig. S1 Maps showing the approximate locations of the mouths of the rivers where samples of Atlantic salmon were collected for the microsatellite database.

Fig. S2 Hierarchical population structure in the St. Lawrence North Shore regional group using STRUCTURE.

Fig. S3 Hierarchical population structure in the Gaspésie/Anticosti regional group using STRUCTURE.

Fig. S4 Hierarchical population structure in the Gulf of St. Lawrence regional group using STRUCTURE.

Fig. S5 Hierarchical population structure in the Nova Scotia East Coast regional group using STRUCTURE.

Fig. S6 Hierarchical population structure in the Newfoundland regional group using STRUCTURE.

Fig. S7 Hierarchical population structure in the Labrador regional group using STRUCTURE.