The demographic history of Atlantic salmon (*Salmo salar*) across its distribution range reconstructed from approximate Bayesian computations*

Quentin Rougemont[1,2] and Louis Bernatchez[1]

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

[2]*E-mail: quentinrougemont@orange.fr*

Understanding the dual roles of demographic and selective processes in the buildup of population divergence is one of the most challenging tasks in evolutionary biology. Here, we investigated the demographic history of Atlantic salmon across the entire species range using 2035 anadromous individuals from North America and Eurasia. By combining results from admixture graphs, geo-genetic maps, and an Approximate Bayesian Computation (ABC) framework, we validated previous hypotheses pertaining to secondary contact between European and Northern American populations, but also identified secondary contacts in European populations from different glacial refugia. We further identified the major sources of admixture from the southern range of North America into more northern populations along with a strong signal of secondary gene flow between genetic regional groups. We hypothesize that these patterns reflect the spatial redistribution of ancestral variation across the entire North American range. Results also support a role for linked selection and differential introgression that likely played an underappreciated role in shaping the genomic landscape of species in the Northern hemisphere. We conclude that studies between partially isolated populations should systematically include heterogeneity in selective and introgressive effects among loci to perform more rigorous demographic inferences of the divergence process.

KEY WORDS: Approximate Bayesian computations, gene flow, heterogeneous divergence, linked selection, phylogeography, *Salmo salar*.

An understanding of demographic history, accounting for putatively alternating periods of isolation and gene flow among populations, is fundamental for accurate population genetic inferences. In particular, the genomic makeup of present day populations in the northern hemisphere is expected to be largely influenced by population splits and secondary contacts linked to climatic oscillations during the last quaternary glaciations (Bernatchez and Wilson 1998; Hewitt 2000). Yet, elucidating the degree to which the contemporary distribution of genetic variation within species reflects these historical divergence processes is challenging. Under the genic view of speciation, during allopatric phases, populations can randomly accumulate genetic Dobhzansky–Muller incompatibilities and other genetic barriers to gene flow due to genetic drift and selection (Wu 2001; Harrison and Larson 2016). Following secondary contact, gene flow is expected to partially or entirely erode past genetic differentiation outside of barrier regions. Depending on the balance between levels of gene flow following secondary contact and the number of accumulated barriers, heterogeneous landscapes of genetic divergence may arise (Wu 2001; Wolf and Ellegren 2016). Indeed, recent empirical population genomics studies have documented the near ubiquity of the heterogeneous landscape of differentiation across a continuum of increasing divergence (Seehausen et al. 2014; Wolf and Ellegren 2016).

Early studies used genome scans to identify islands of differentiation (Feder et al. 2012; Seehausen et al. 2014) and described a conceptual model of divergence with gene flow. Under such models of divergence hitchhiking gene flow keeps eroding genetic differentiation outside of selected regions while selective sweeps involved in local adaptation and linked neutral variants will show higher level of differentiation (Via and West 2008; Feder et al. 2012; Flaxman et al. 2013). However, allopatric divergence followed by secondary contact, whereby gene flow erodes past genetic differentiation in region without barrier loci or where gene flow is less effective (Barton and Bengtsson 1986), can produce the same patterns (e.g., Gagnaire et al. 2013; Burri et al. 2015; Roux et al. 2016; Rougemont et al. 2017; Rougeux et al. 2017). Moreover, the link between heterogeneity of differentiation and the role of gene flow during divergence is increasingly questioned. Indeed, various linked selective process such as hitchhiking of neutral alleles linked to a selective sweep, especially in regions of low recombination, can produce similar pattern (Noor and Bennett 2009; Cruickshank and Hahn 2014). Similarly, background selection (Charlesworth et al. 1993) can contribute to this heterogeneous genomic landscape of divergence. The effect of linked selection is a reduction in polymorphism at sites harboring positive as well as deleterious variants and at surrounding sites. The intensity of this process will mostly be determined by local recombination rate variation and the density of selected sites within the region (Kaplan et al. 1989; Nordborg et al. 1996; Payseur and Nachman 2002; Burri 2017). Linked selection can be seen as an increase in genetic drift and thus be modeled as a local reduction of effective population size (Ne). Under these linked selective processes, Ne will be reduced in genomic areas of low recombination relatively to regions of higher recombination (Hill and Robertson 1966; Charlesworth et al. 1993). Neglecting such selective effects can lead to biased demographic inferences (Ewing and Jensen 2016; Schrider et al. 2016). Therefore, a modeling approach that jointly allows for local genomic variation in effective population size and migration rate can improve our understanding of the demographic processes at play during population divergence. Tools such as approximate Bayesian computations (Tavaré et al. 1997; Beaumont et al. 2002) allow for the simulation of complex models of divergence history and new methods have been developed to model genome-wide heterogeneity in levels of introgression (Roux et al. 2013; Sousa et al. 2013; Tine et al. 2014), as well as local variation in effective population size (Christe et al. 2016; Roux et al. 2016; Rougeux et al. 2017). Together, these approaches provide an understanding of the joint effects of barrier loci reducing migration rate and of linked selection reducing local effective population size and will lead to more rigorous demographic inferences of population divergence processes.

Atlantic salmon (*Salmo salar*) is a particularly relevant model to study the interactions of contemporary and historical factors leading to heterogeneous landscapes of divergence. Distributed throughout the North Atlantic, both in Eastern North America and Europe, it undergoes long anadromous migrations to feed at sea at the adult stage before returning to natal rivers for spawning (Quinn 1993). Such homing behavior results in reduced gene flow at both local and regional scales that in turn translates into fine-scale spatial structure and a pattern of isolation by distance, which may also facilitate the establishment of local adaptation (Taylor 1991; Dionne et al. 2008; Perrier et al. 2011; Primmer 2011). Accordingly, a large body of literature has been devoted to documenting both the population genetic structure of this species (e.g., Vasemägi et al. 2005; Palstra et al. 2007; Perrier et al. 2011; Bourret et al. 2013; Moore et al. 2014) and effective population sizes (e.g., Palstra et al. 2009; Ferchaud et al. 2016).

Previous studies have also revealed a pronounced continental divergence between European populations and Northern American populations using various molecular markers (Cutler et al. 1991; Bourke 1997; King et al. 2007; Bourret et al. 2013). These studies suggested that continental divergence likely started 600,000–700,000 YBP (Nilsson et al. 2001; King et al. 2007). However, European alleles from mtDNA and allozymes have been found to segregate at low frequencies among American populations from Newfoundland and Labrador (Verspoor et al. 2005; King et al. 2007), suggesting subsequent intercontinental genetic exchanges. At the continental scale, European populations likely split following the Last Glacial Maxima (LGM) around 18,000 YPB, following expansion from putative southern refugia in the Iberian peninsula and other nonglaciated areas further north (King et al. 2007). Recent studies (King et al. 2007; Bourret et al. 2013; Bradbury et al. 2015) also support such a historical scenario whereby salmon populations may have diverged in multiple refugia within Europe, and whether these populations have come into secondary contact has not yet been resolved. In contrast to the pronounced regional level of population genetic structure among European populations, studies on North American populations have revealed comparatively lower genetic diversity and lower levels of genetic differentiation (King et al. 2007; Bourret et al. 2013).

Despite the plethora of population genetics studies mentioned above, none have performed demographic inferences to explicitly test and compare alternative scenarios of divergence. Nevertheless, these inferences are key to understanding the process of divergence and should ideally take into account among-loci variability in genetic drift and effective migration (Roux et al. 2016). Here, we explicitly tackled this issue and reconstructed the divergence history of Atlantic salmon across its whole native range. We took advantage of two previously published datasets focused on describing patterns of population genetic structure (Bourret et al. 2013; Moore et al. 2014) and performed more in-depth analysis of demographic inference that was not possible before.

In addition to classical population genetic analyses, we used an Approximate Bayesian Computation (ABC) framework and compared alternative models of population divergence that integrated genome-wide variation in both effective population size and migration rate to jointly account for the effect of linked selection and barriers to gene flow. The use of ABC was justified for the following reasons. First, coalescent approaches, as implemented in an ABC framework, and diffusion approximation methods (e.g., Gutenkunst et al. 2009) are, to the best of our knowledge, the only way to perform explicit demographic modeling. Second, inferences that explicitly take into account among-loci heterogeneity in migration rate and drift are only implemented in an ABC framework (Roux et al. 2016) or in *dadi* (e.g., Rougemont et al. 2017; Rougeux et al. 2017). However, diffusion approximation methods may fail to correctly explore the parameter space as the complexity of tested demographic scenarios increases and this approach also requires bootstrap methods to estimates confidence intervals around estimated parameters, making it no faster than coalescent-based ABC methods. Finally, the flexibility of ABC further allowed us to account for ascertainment bias present in our dataset and, more generally, it allows users to account for various ascertainment schemes.

Methods

SAMPLING AND GENOTYPING
A total of 2080 individuals representing 77 locations across the whole native range of Atlantic salmon were merged from two studies (Bourret et al. 2013; Moore et al. 2014; Fig. S1, Table S1) that resulted in a dataset with a total of 5087 loci from a SNP array previously developed at the Centre for Integrative Genetics (CIGENE) in Norway. All details from the SNP discovery as well as quality control and ascertainment bias in North American populations are fully described in Bourret et al. (2013) and detailed below in the ABC section on ascertainment bias. Genotypes were further filtered by excluding individuals with more than 5% missing genotypes (45 individuals removed) as well as SNPs with a lower than 95% genotyping rate resulting in a final dataset of 2035 individuals and 5034 SNPs. Basic summary statistics (i.e., Ho, He, pairwise F_{ST} were computed globally and for each locus using custom R scripts and the Hierfstat package (Goudet 2005). Finally, an analysis of molecular variance quantifying the hierarchical distribution of genetic variability between continent, among populations within continent and within populations was performed with Arlequin 3.5.1.2 (Excoffier and Lischer 2010).

PATTERNS OF BROAD POPULATION GENETIC STRUCTURE AND ADMIXTURE
We first visualized population genetic relationships among individuals using a principal component analysis using the R package ade4 (Dray and Dufour 2007) based on the whole dataset. Second, we inferred levels of ancestry and admixture proportions of individuals using the *snmf* function implemented in the R package LEA (Frichot and François 2015). *K*-values ranging between one and 60 were tested and cross-validations were obtained using the cross-entropy criterion with 5% of masked genotypes. The default value for the regularization parameter was used to avoid forcing individuals into groups and hence underestimating admixture. Maps of ancestry coefficients were drawn in R following recommendations available elsewhere (Caye et al. 2016).

DEMOGRAPHIC INFERENCE
Inferring population splits and sources of admixture
We combined *TreeMix*v1.12 (Pickrell and Pritchard 2012), *f3-test* (Reich et al. 2009), and *Spacemix* (Bradburd et al. 2016) to identify population splits and admixture as well as the most likely source(s) of admixture. We first mapped the SNPs on to the Atlantic salmon reference genome, as detailed in Note S1. *Treemix* analyses were performed using regional genetic groups defined in Bourret et al. (2013) as well as Moore et al. (2014) and complemented by our own clustering analysis (Table S1). This pooling strategy further allowed to reduce the noise due to ongoing gene flow between neighboring rivers and to simplify the visualization of the tree to test up to 20 migration events. Second, we used the *f3-s*tatistical tests with bootstraps in blocks of 100 SNPs to more formally test for admixture (Reich et al. 2009). Finally, we used a spatial framework for inference of the geogenetic positions of populations with their most probable source and direction of admixture implemented in the R package *Spacemix* as a complement to *Treemix*. We estimated together the population locations as well as the extent of admixture. Parameter settings strictly followed those implemented by the authors for the study of human populations (Bradburg et al. 2016). Analyses were performed using the covariance-matrix on the full dataset. The convergence of the runs was checked using functions provided in the Spacemix package.

Explicit modeling: Coalescent analyses and approximate Bayesian computation
A recently developed ABC framework (Roux et al. 2013, 2016) was used to perform explicit modeling of divergence history while incorporating the effects of selection at linked sites affecting *Ne* and differential introgression (*m*). This framework was modified to test the effect of ascertainment bias on model choice.

Demographic scenarios were tested on a subset of populations representative of the major genetic groups identified in Bourret et al. (2013) for Europe, Moore et al. (2014) for North America and in this study (Table S1) and pairwise comparison
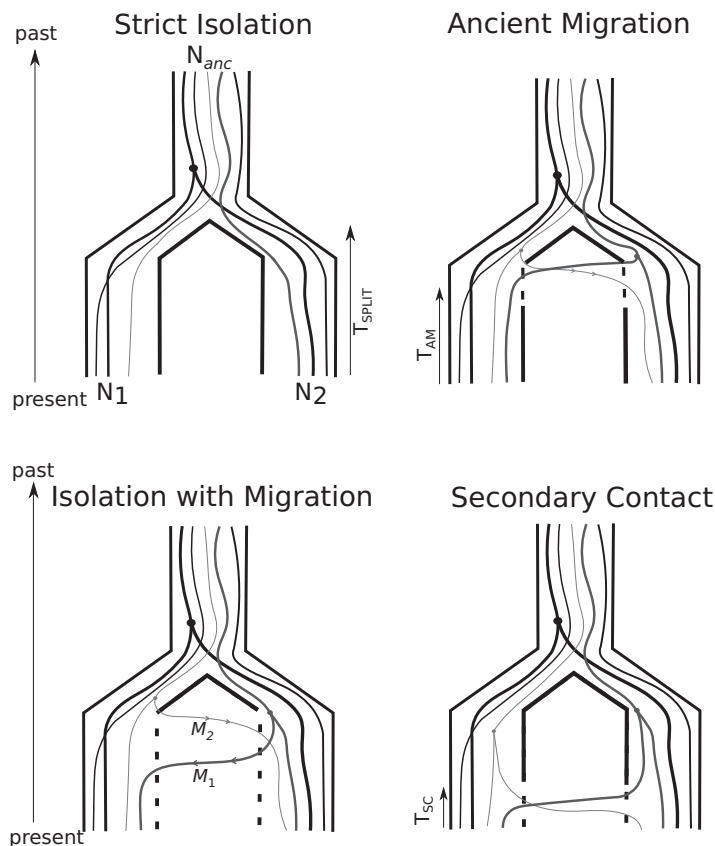
Figure 1. Representation of the demographic scenarios compared in this study: Strict Isolation (SI), Isolation with constant Migration (IM) Ancient Migration (AM) and Secondary Contact (SC). All models share the following parameters: *Tsplit*: number of generation of divergence (backwards in time). *N*anc, *N₁*, *N₂*: effective population size of the ancestral population, of the first and second daughter population compared. M₁ and M₂ represent the effective migration rates per generation with *m* the proportion of population made of migrants from the other populations. *Tam* is the number of generations since the two populations have diverged without gene flow. *Tsc* is the number of generations since the populations have started exchanging alleles (secondary contact) after a period of isolation.

of divergence scenarios was performed as detailed below. Four scenarios characterized by a set of demographic parameters and assumptions were compared (Fig. 1). These include a model of strict isolation (SI), a model of divergence with migration during the first generations of divergence or ancestral migration (AM), a model of isolation with migration (IM) and a model of secondary contact (SC). The SI model assumes an instantaneous split at T_{split} of an ancestral population of size N_{ANC} into two daughter populations of constant size N_{pop1} and N_{pop2} and no gene flow. The AM model assumes an instantaneous split of the ancestral population followed by a period of gene flow from T_{split} to T_{am} and then a period without gene flow until the present date. The IM model assumes that after T_{split}, the two daughter populations continuously exchange genes at a constant rate each generation. The SC model assumes an initial period of strict isolation without gene flow after T_{split} and is followed by a period of secondary contact T_{sc} generations ago that is still ongoing. The IM, AM, and SC models included migration as $M = 4\,N_0.m$, with $M_{1 \leftarrow 2}$ being the number

of migrants from population 2 to population 1 and $M_{2 \leftarrow 1}$ being the reciprocal. Then heterogeneity of migration and of effective population size was modeled using beta distributions as hyper-prior on each of these two parameters. For each model, 1×10^6 simulations of datasets matching the sample size of each locus in each pair-wise dataset was performed using msnsam (Ross-Ibarra et al. 2008) a modified version of the coalescent simulator ms (Hudson 2002) under an infinite-site mutation model. Large and uninformative priors were drawn as follows: priors for the divergence time ($T_{\text{split}}/4Ne$) were uniformly sampled on the interval [0–30], so that T_{am}, T_{sc}, and T_{si} were constrained to be chosen within this interval. For homogeneous Ne, N_{ANC}, N_{Pop1}, and N_{Pop2} were sampled independently and uniformly on the interval [0–20]. For the homogeneous version of the migration rate, M was sampled on the interval [0–40] independently for each direction of migrations. Heterogeneity in introgression rate was modeled using a beta distribution as a hyper prior shaped by two parameters; α randomly sampled on the interval [0–20] and β

randomly sampled on the interval [0–200]. A value was then independently assigned to each locus. For heterogeneity of Ne, α percent of loci shared a common value uniformly sampled on the interval [0–1] and represents the proportion of loci assumed to evolve neutrally. Then a proportion $1–\alpha$ was assumed to evolve nonneutrally. Their Ne values were thus drawn from a beta distribution defined on the interval [0–20], as defined on the homogeneous version. Ne values were independently drawn for N_{ANC}, N_1, and N_2 but shared the same shape parameters (Roux et al. 2016). Priors were generated using a Python version of *priorgen* software (Ross-Ibarra et al. 2008) and a panel of commonly used summary statistics (Fagundes et al. 2007; Roux et al. 2013) was computed using mscalc (Ross-Ibarra et al. 2008, 2009; Roux et al. 2011).

Accounting for ascertainment bias in model choice

The SNP array used in this study potentially possesses ascertainment bias due to (i) the unbalanced representation of European versus North American individuals in the ascertainment panel used during SNPs discovery, and (ii) the use of a minor allele frequency (MAF) cutoff that is expected to bias clustering algorithms (Guillot and Foll 2009). To account for this potential bias, we generated a pipeline that consisted of intentionally biasing the distribution of the site frequency spectrum (SFS) from each simulated dataset (described in detail in Note S1). This procedure allowed us to produce simulated datasets with a distribution of allele frequency resembling those observed in each of our empirical datasets. To ultimately address the potential confounding effect of ascertainment bias, we also performed demographic inference using only the subset of SNPs derived from both European and North American samples.

ABC: Model selection

A total of 14 models were tested: four alternative models for IM, AM, SC hereafter called *NhomoMhomo* (for Homogeneous Ne and Homogeneous migration along the genome); *NhomoMhetero* (for Homogeneous Ne and Heterogeneous migration along the genome); *NheteroMhomo* (for Heterogeneous Ne and heterogeneous migration along the genome) and *NheteroMhetero* (for Heterogeneous Ne and heterogeneous migration along the genome); and two alternative models for SI that is *Nhomo* or *Nhetero*. First, the 14 models were all compared together at the same time using ABC based on a set of 19 summary statistics (described in Note S1). Posterior probabilities were computed by retaining the 3500 "best" simulations (out of 14 × 1 million) during the rejection step, those were weighted by an Epanechnikov kernel peaking when $S_{\text{obs}} = S_{\text{sim}}$ and a neural network was used for the rejection step. Fifty neural networks were used for training with 15 hidden layers. Ten replicates were

performed to compute an average posterior probability for each model. Finally, we compared the best secondary contact models against the two remaining alternatives PSC and BSC. Each time, the same model selection procedure was repeated. The R package "abc" (Csilléry et al. 2012) was used for the model choice procedure.

Robustness

The robustness, which is defined as the probability P of correctly classifying a model M given a posterior probability threshold X, was assessed using pseudo-observed dataset (PODS) and computed using the following formula from (Roux et al. 2016): $R = P(m > X \mid M)/[\sum_1^m P(Pm > X \mid i)]$ where $P(Pm > X \mid i)$ is defined as the probability that a given dataset from the model m is indeed classified as belonging to M above the threshold X by the abc classifier.

A total of 10,000 PODS from each model were simulated with parameters drawn from the same prior distribution. Then, the ABC model choice procedure was run again but using the PODS to compute their posterior probability. For the dataset generated with ascertainment bias a total of 5000 PODS × 14 models was simulated.

Parameter estimation and posterior predictive checks

For every robustly inferred model, parameter estimation was performed using a logit transformation of the parameters and a tolerance of 0.001. The posterior probabilities of parameter values were then estimated using the neural network procedure with nonlinear regressions of the parameters on the summary statistics using 50 feed-forwards neural networks and 15 hidden layers. We then performed posterior predictive check (PPC) to assess the goodness-of-fit of the best-supported model (Rougemont et al. 2016). We ran 2000 simulations for each locus using the joint posterior parameter estimates in each comparison to generate summary statistics and compare them to those empirically observed in each dataset. The simulation pipeline was used again to compute the P-value for each summary statistic to estimate the fit.

Results
POPULATION GENOME-WIDE DIVERSITY AND DIVERGENCE

Heterozygosity among Northern American populations (Ho = 0.15) was almost half that observed in Europe (Ho = 0.26; P-value <0.0001; Table S2). Global differentiation across all populations was pronounced with a mean F_{ST} value of 0.379 (range 0.015–0.971, Fig. S1) and as expected, the mean F_{ST} among European populations was lower ($F_{\text{ST}} = 0.140$,

range 0.000–0.680) than that observed between continents. The level of differentiation among Northern American populations was almost half that observed in Europe (mean $F_{ST} = 0.0880$, range: 0.000–1.000, Table S2). Based on the AMOVA, 46% of variance was observed between continents ($P > 0.001$), 47% of variance was observed within population and 7% of variance was observed among populations within continents.

INDIVIDUAL CLUSTERING

The PCA (Fig. 2A) revealed a separation between each continent along the first axis explaining 31.68% of the variance, a pattern consistent with the strong mean intercontinental F_{ST} value. While all North American populations clustered together, the second axis (2.30% of variance) separated European populations into three major groups corresponding to an "Atlantic" group, a group from the Northern Europe (Barents-White Sea) and a Baltic group. Plots of axis 1 and 3 and axis 1 and 4 produced similar results as most differentiation was driven by continental divergence (Fig. S2). Projection of the axis 2 and 3 of the PCA captured 4% of variance; (Fig. 2B) and separated the Baltic Sea populations from all others while North American populations fell at the center of the axes and the Atlantic populations clustered along the third axis with a tendency toward a North–South organization. The separation among the major European groups was best viewed when plotted separately from North America (Fig. S3, 3d-PCA) that revealed the closer proximity of the Baltic to the Barents-White Sea group relative to the Atlantic group (axis 1 and axis 3), while the Barents-White Sea was introgressed by the Atlantic group (axis 1) and the isolation of the Baltic from other groups was apparent on axis 2.

Results of population structure showed that estimating a single fixed K was difficult (Fig. S4) as the cross-entropy criterion decreases without apparent breaks indicating that several clustering values could fit the data well. We therefore plotted a diversity of solutions to avoid overinterpretation (Falush et al. 2016) with a focus on the major groups reflecting historical grouping. A first change in the minimal cross-entropy criterion was observed for $K = 5$ and $K = 7$ (Fig. S4). $K = 5$ revealed the same clustering as the PCA but with two groups inferred in North America (Fig. S5). At $K = 7$ (Fig. 2C and bar plot Fig. 2D), 42% of individuals showed a q-value <0.9 value (Fig. 2D). The seven groups inferred corresponded to the "Baltic" group (mean q-value = 0.99), the "Icelandic" group (mean q-value = 0.99), the "Spanish group" (mean q-value = 0.97), the "Atlantic" group (mean q-value = 0.82), the "Barents-White sea" group (mean q-value = 0.86) a "Southern" group in America (mean q-value = 0.82) and a Northern Group in America (mean q-value = 0.85).

AMERICAN POPULATIONS WERE FOUNDED BY MULTIPLE EUROPEAN SOURCES

Treemix analysis indicated a steady increase in the percentage of variance of the covariance matrix explained as migration events were added to the tree (Fig. S6). While 99.8% of variance in the covariance matrix was already explained without migration (Fig. 3A), this percentage increases for different number of migration edges. The P-value for each migration event was significant ($P < 0.001$) for up to 10 migrations events, demonstrating that even if most of the history of divergence can be summarized without gene flow, a more complex history of split and subsequent gene flow provided better fit to the data (Fig. 3B). This illustrated potentially multiple migration events both between continents as well as within continent. (See Fig. S7 and Fig. S8 for residuals).

While the fit of *Treemix* tree was improved with migration, we avoided direct interpretation of migration edges and further used the *f3*-statistic to formally test for admixture. Out of 504 statistically significant tests, eight regional genetic groups from North America and one from the Barents Sea (Tuloma River) appeared significantly admixed with genetic background from other groups (Table S3). The Baltic Sea was a source of admixture towards the Barents Sea in 80% of the significant tests. In North America, the most statistically significant sources of admixture were the two southernmost North American population groups of Nova Scotia (NVS, 151) and Narraguagus River (Nar, 153). Significant *f3*-tests of admixture into North American populations always involved one European and one American population, that is, no test involving two European populations or two North American populations was significant.

Spacemix results showed that geo-genetic maps without admixture largely recovered results expected from a simple PCA (Fig. 2A vs Fig. 3C), however a model incorporating both admixture and migration provided additional information regarding the demographic history of populations (Fig. 3D). A plot of all possible admixture events would indicate that most populations displayed limited amounts of admixture and occupy a different geo-genetic position. Nevertheless, the visualization of the geo-genetic map indicates that the admixture source of several European populations (Suma, Pongoma, Tuloma, Tana, Yapoma, Lebyazhya, Vindelaiven, Gaula, and Lardaiselva stems from North American populations. Similarly, many North American populations displayed their admixture source located far away from their geo-genetic location. Figure 3E illustrates the variance in admixture estimates in each sampled location where European and North American populations displayed almost identical admixture proportions (mean Europe = 0.090 [95%CI = 0.082–0.111], mean North America = 0.101 [95%CI = 0.087–0.117].
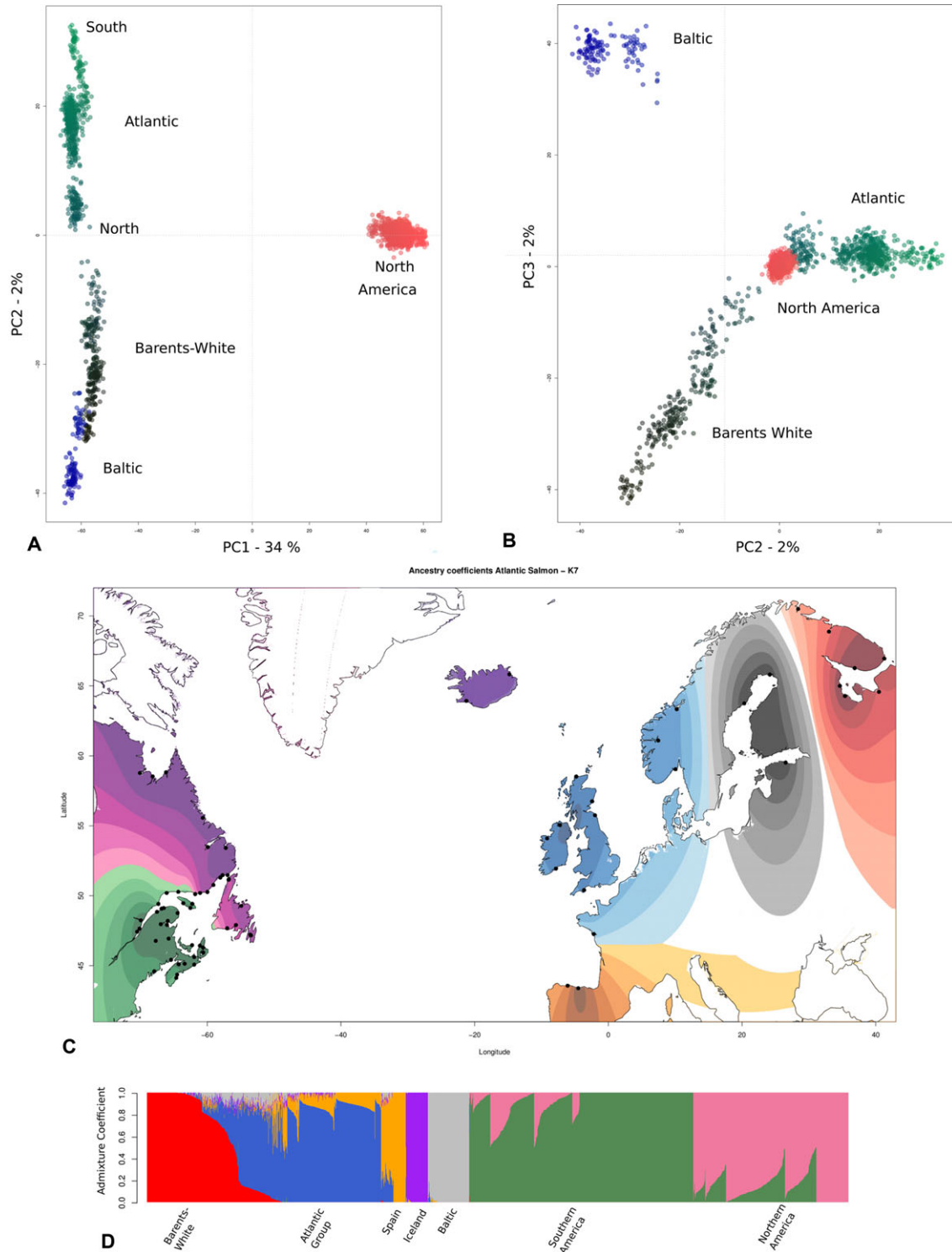
Figure 2. Patterns of population genetic structure among all 2035 Atlantic salmon analyzed in this study. (A) Principal components analysis with the first two axes plotted; (B) PCA of the axis 2 (horizontal) and 3 (vertical); (C) Spatial interpolation of population structure inferred for $K = 7$ clusters. Each black dot represents a sample points. Colors depicts ancestry coefficient of the major groups. Spatial interpolation of ancestry coefficient was obtained using a kriging method following Jay et al. (2012). We mapped only cluster with the maximum local contribution to the observed ancestry on each point. Interpolation outside of the species range is not relevant for interpretation. The associated barplot is provided in (D) where each vertical line represents an individual and each color represents a genetic cluster. Individuals are sorted according to their membership probability. Major geographic areas are provided below the plot. Results for additional K-values are provided in Figs. S4 and S5.
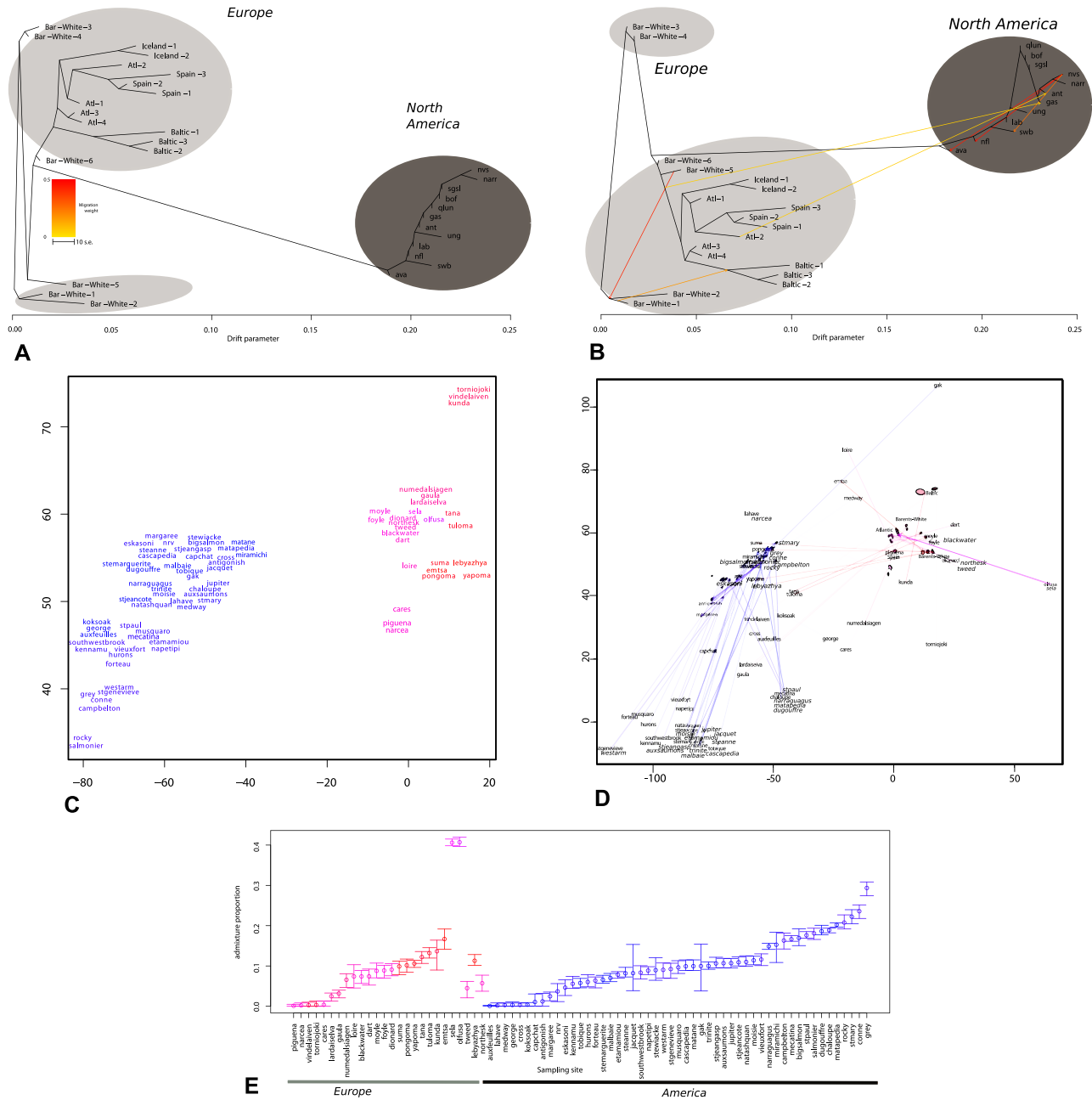
Figure 3. Population admixture and sources of gene flow. (A) Admixture graph inferred with Treemix without migration and (B) *m = 9* migration events. The *x*-axis represents the amount of genetic drift, proportional to *Ne.* Arrows depict admixture events and color gradients indicate the strength of admixture. Details of population pooling are provided in Table S1 (C) Spacemix map inferred, without admixture corresponding to a model of pure isolation by distance. Each name represents a sample site and the color gradient is made according to each sample geographic position. The map broadly recovers the results of the PCA with American sample clustered on the left side (blue gradients) and European samples clustered on the right side (red gradients). (D) Spacemix inferred geogenetic map with admixture. Source of admixture are represented in italics and the original position of each sample points are in bold. For America, original positions were removed for ease of interpretations and replaced by the "major groups" from which they belong. Arrows joins the location of admixture source (in italics) to the current location of the sample, opacity of the arrow is a function of the amount of admixture of the sample. Ellipses around each point represent the 95% CI of geogenetic locations. (E) Admixture proportions inferred by Spacemix for each sampling locations with 95% CIs. Color scheme is the same as before with color gradient based upon geographic position.

DEMOGRAPHIC HISTORY AND DIVERGENCE

Approximate Bayesian Computation was performed in 163 pairwise comparisons among populations. Demographic inference was performed pairwise for the following reasons: (i) classic scenario of divergence are generally compared between pairs of populations; (ii) the number of parameters involved in a three or four populations models including heterogeneous migration and effective population size becomes too large for the amount of data that was available here. Thus, a total of 90 comparisons were performed between the European and North American continents, 37 comparisons involved North American populations and 36 comparisons involved European populations from major groups identified in previous analyses. Due to exponential increases in computational time when accounting for ascertainment bias, we tested the effect on a random subset of 45 pairwise comparisons only.

MODEL SELECTION AND ROBUSTNESS

The most striking pattern emerging from the model selection procedure was the high posterior probability for the secondary contact model when averaged over the four declinations (*NhomoMhomo*, *NheteroMhetero*, *NheteroMhomo, NhomoMhetero*). The median averaged over all comparisons was P(SC) = 0.99 (SD ± 0.091) (Table 1, Table S4). Accordingly, 84% of the comparisons displayed robustness > 0.950 and a posterior probability of SC > 0.969. Also, comparisons between the homogenous model (*NhomoMhomo*) and heterogeneous models showed that models incorporating heterogeneity of either *Ne*, *m*, or of both were largely favored, as indicated by higher posterior probabilities (Table 1).

Testing for the effect of ascertainment bias revealed this had no effect on the model choice procedure. Indeed, tests on a random set of 45 comparisons between continents and between the European and Baltic lineages, where we expected the effect to be the greatest, consistently supported a model of secondary contact with a median P(SC) = 0.986 (SD ± 0.126) (Figs. S9 and S10) when averaged over all four alternatives models with a median robustness of 0.997 indicating that our inference can be confidently considered. Finally, the abc model selection procedure performed on the subset of markers discovered using both North American and European samples reached the same conclusion on another random set of 16 intercontinental comparisons with a median P(SC) > 0.981 (SD ± 0.021; Table S5).

To test for the potential effect of contemporary admixture, we ran another set of ABC analyses (*n = 20*) on a random subset of individuals assigned to their genetic sampling site with a *q*-value > 0.90. The P(SC) averaged over all comparisons was > 0.93 (Table S6). Therefore, recent admixture due to human stocking should have minimal effects on inferences.

PARAMETER ESTIMATIONS AND GOODNESS OF FIT

Demographic parameters were estimated for a subset of models (*n* = 148) with robust inferences at a threshold posterior probabilities ≥ 0.865. Using PPC, we found that under the best model, the simulation pipeline produced accurate estimates based on the posterior distribution of 2000 newly simulated summary statistics. The PPC indicated that π tended to be significantly different from the observed value ($P < 0.001$) in 46 out 148 comparisons and the average and standard deviation of the number of fixed differences was also less accurately reproduced (*n* = 16/148, $P < 0.001$; Table S7).

Posterior parameter estimates were well differentiated from the prior for the time of secondary contact in 147 out of the 148 comparisons and indicated very recent secondary contacts as they represented on average 1.3% of the split time between continents, 1.5% of the split time between populations within North America and 0.7% between populations within Europe (Fig. 4). In contrast, posterior estimates of split times were differentiated from the priors in 13% of the comparisons only (Table S8). To refine split times estimates the following complementary approach was used. We identified for the 148 most robust comparisons the most differentiated loci according to their F_{ST} values. All loci falling outside the 95% (or 90% when all loci were fixed) percentile of the F_{ST} distribution were sampled and parameters were estimated under a model of strict isolation (SI) using the whole simulation pipeline with a new set of one million simulations. Under the genic view of speciation (Wu 2001), barrier loci harboring various genetic incompatibilities are more likely to have been resistant to introgression, and to display deeper coalescent times, therefore having high F_{ST} values. In these conditions, a less parameterized model of strict isolation appears as a parsimonious alternative to estimate split time. Model choice performed in a subset of the data (*n* = 60) revealed SI as the best model except in 14 cases between continents. However, this was not the case for within-continent comparisons, where SC always remained the best model. For within-continent comparisons, we found that posterior estimates of split times under SC were systematically not differentiated from the prior except in four cases and consequently not interpretable. Estimates of split time under SI between continents revealed that divergence times were accurately estimated in 65% of the comparisons (Table S9) with a mean continental split time of 20.84 coalescent units (Fig. S11).

Posterior estimates under the secondary contact model also indicated a reduction in effective population size of the contemporary populations compared to the ancestral population size but with large confidence intervals. Estimates of migration rate under the secondary contact models produced large confidence intervals in more than 50% of the comparisons making the interpretation of those results not relevant (Table S8). The same was true for

Table 1. Model selection results.

	All models				Best model	
	P(SI)	P(AM)	P(IM)	P(SC)	P(SC homo)	P(SC hetero)
Within America	0.000	0.000	0.013	**0.984**	0.0040	**0.9500**
Between Continent	0.000	0.000	0.005	**0.993**	0.0115	**0.9760**
Within Europe	0.000	0.003	0.024	**0.967**	0.0385	**0.8615**

Median posterior probability for all models (left panel) and for the homogeneous versus heterogeneous version under the best model. Median computed for all pairwise comparison under different geographical configuration.

estimates of the number of SNPs exhibiting reduced introgression rates and reduced population size (Table S8, Fig. 4).

Discussion

The main goal of this study was to reconstruct the demographic history of between-continent and within-continent divergence of Atlantic salmon populations throughout its geographic distribution. To this end, an ABC framework was implemented taking into account the potential effect of linked selection locally reducing Ne and that of genetic barriers locally reducing the rate of effective gene flow (m). Demographic inferences strongly supported a long period of geographic isolation, during which various processes likely shaped the heterogeneous landscape of divergence between them. Periods of isolation were subsequently followed by widespread secondary contacts between continents as well as within each continent that have contributed to the erosion of genome-wide differentiation outside of genetic barriers.

EFFECT OF ASCERTAINMENT BIAS ON POPULATION GENETIC INFERENCES

Because the SNP chip used in this study was developed using mostly European populations of salmon from the Atlantic group, the dataset generated from it potentially exhibits ascertainment bias expected to particularly affect estimates of genetic diversity in North American populations as well as in Baltic Sea populations, albeit to a lower extent. Nevertheless, the significant differences in genetic diversity observed here was also observed by Bourret et al. (2013) when using a subset of markers discovered in both European and Canadian salmon as well as in previous studies using other makers (e.g., microsatellites, mtDNA) that have also revealed lower genetic diversity among North American populations (e.g., Verspoor et al. 1999, 2005; King et al. 2001, 2007; Consuegra et al. 2002; Tonteri et al. 2007). Since neither clustering methods (Guillot and Foll 2009) nor the Treemix and Spacemix approach are designed to account for possible ascertainment bias, we sought to explore how the ascertainment bias would impact our demographic inference by taking advantage of the flexibility of the ABC approach implemented here. Here, the ABC pipeline

indicated that results of model choice accounting for ascertainment bias were the same as results that did not account for this bias.

RESOLVING THE HISTORY OF ATLANTIC SALMON POPULATION DIVERGENCE

One of the most striking patterns of our demographic inferences was the unambiguous support for a scenario of long divergence without gene flow followed by very recent episodes of secondary contacts (approximately 1% of the total divergence time) in more than 90% of our comparisons between continents as well as among European and among North American populations. These long periods of isolation followed by relatively small periods of secondary contacts provided the best conditions for robustly inferring secondary contacts (Roux et al. 2016). This is because long secondary contact periods result in the loss of the isolation signal (Barton and Hewitt 1985) and a situation of migration-drift equilibrium with a semipermeable barrier to gene flow is more likely to be attained (Endler 1977), as observed when reconstructing demography using only putatively "neutral" markers (Bierne et al. 2013; Rougemont et al. 2016).

As mentioned above, we also observed a global reduction of genetic diversity among North American populations. Accordingly, *Treemix* analyses revealed higher levels of drift among North American populations relative to European ones. These results therefore support the hypothesis that North American populations were established by ancestral European populations (King et al. 2007), which raises several questions: how long have the populations on each continent been isolated? When was the secondary contact initiated? What were the sources of colonization of North America? What is the potential impact of these small contemporary gene flow events on the genomic make up of contemporary populations?

We cannot provide direct absolute estimate of divergence time in years without formulating assumptions regarding generation time and the unknown mutation rate. Nevertheless, estimates of split time under SI, assuming a mean generation time of four years (Palstra et al. 2009) and a $Nref = 5000$ (scaling
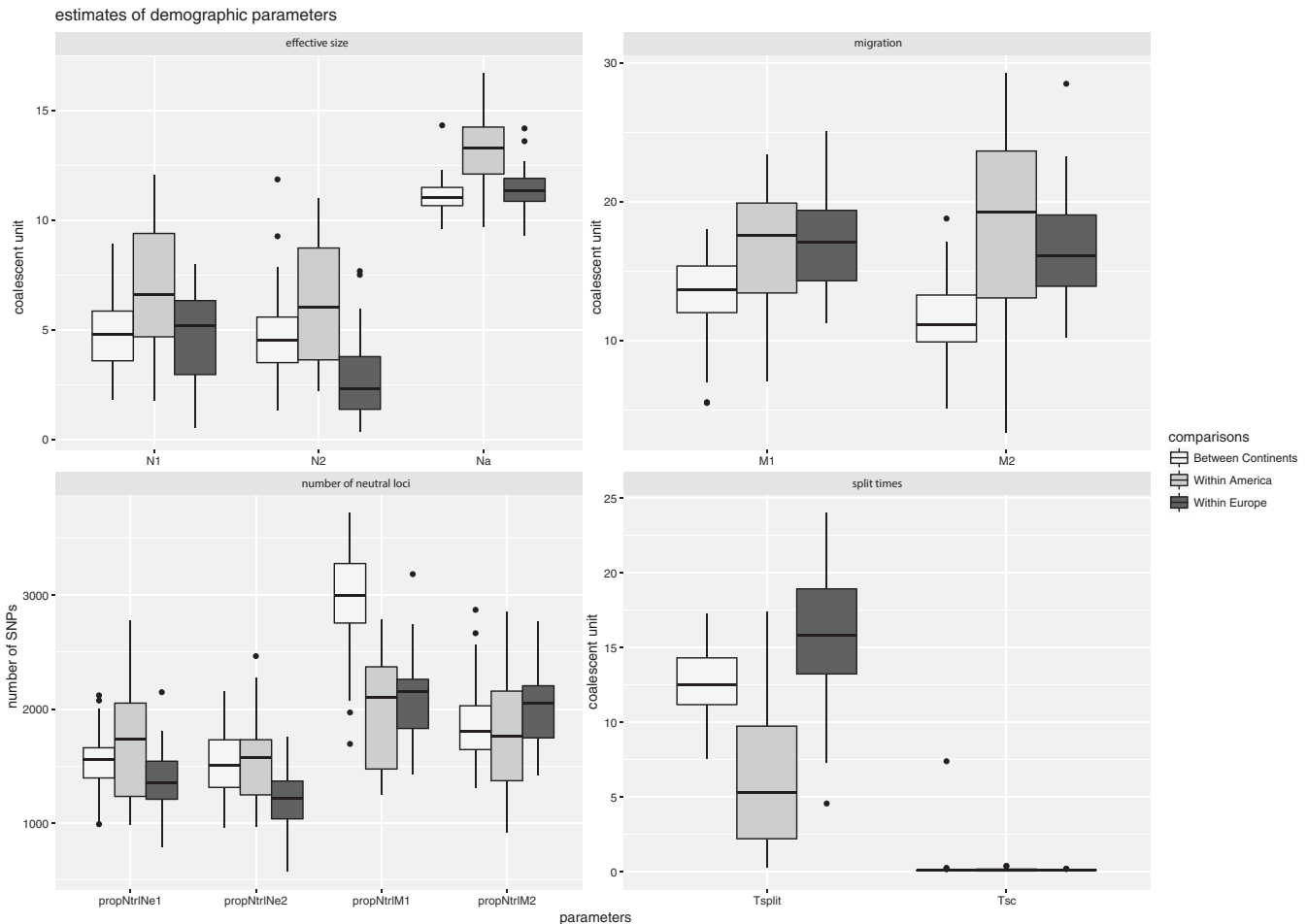
estimates of demographic parameters

Figure 4. Estimates of demographic parameters under the best secondary contact models. Mean values are provided and averaged over all comparisons between continent, within American populations and within European populations. Bottom left panel depicts the number of SNPs inferred by abc as behaving "neutrally" (i.e., number of SNPs with a nonreduced effective population size and number of SNPs freely introgressing between populations).

factor in coalescent simulations) suggests that divergence between continents was initiated ∼1,670,000 YBP[95% credible intervals = 1,564,000–1,764,000], while the secondary contact between continental populations began ∼ 13,400 years ago [95% credible intervals = 1800–41,380]. Our estimates of split times are therefore closer to those proposed by Nilsson et al. (2001 of > 1 million years than those of (King et al. 2007) who estimated the split between 600,000 and 700,000 years based on mtDNA. These differences can be explained by the fact that the authors assumed different substitution rates (1.2% per My in King et al. (2007) vs 0.5–0.9% per My to the same data in Nilsson et al. (2001)) and also used different genetic markers. Therefore, our estimates suggested that continental divergence occurred during the mid-Pleistocene that lasted from ∼2.58 MYA to ∼11,000 years ago (Gibbard et al. 2010). Accordingly, Atlantic salmon on each continent would have been in strict isolation approximately 99% of their divergence time during the Quaternary, a period

where most of the earth surface was glaciated (Hewitt 2000). This long continental isolation period has most likely facilitated the accumulation of genetic incompatibility and is supported by the observation of pronounced asymmetric outbreeding depression at the second generation of hybridization, reflecting the expression of Dobzhanskhy–Muller incompatibilities (Cauwelier et al. 2012). Such a long period of isolation also apparently resulted in karyotypic divergence between North American and European populations (Hartley 1987; Lien et al. 2016). Therefore, Atlantic salmon from each continent can be best described as partially reproductively isolated species, separated by a semipermeable barrier to gene flow, as reported in other fishes such as European Seabass *Dicentrarchus labrax* (Tine et al. 2014), the European Anchovy *Egraulis encrasicolus* (Le Moan et al. 2016), the European River and Brook Lampreys *Lampetra fluvatilis* and *L. planeri* (Rougemont et al. 2017), or the Lake Whitefish (*Coregonusclupeaformis*; Rougeux et al. 2017). Our results also suggested that

following this long phase of geographic isolation, intercontinental secondary contact would have occurred at the end of the last glacial maxima (LGM), approximately 10,000–15,000 years ago. This period was associated with major environmental changes such as melting of ice sheets in North America and Europe, accompanied by an abrupt rise in sea levels as well as changes in oceanic circulations and temperature warming (Clark et al. 2009; Negre et al. 2010). These factors may have facilitated recent gene flow over long distance and impacted the current distribution and demographic history of species (Hewitt 1996; Bernatchez and Wilson 1998).

DEMOGRAPHIC HISTORY OF EUROPEAN POPULATIONS

Bourret et al. (2013) previously documented a cline in allele frequency at the majority of outlier loci when comparing the Baltic versus Atlantic populations and the Barents-White versus Atlantic populations, and the authors hypothesized that this reflected secondary contact between these European genetic groups (see also Jeffery et al. 2017). Earlier studies have also proposed that Baltic populations must have existed in a separate refugium located either in the North Sea (Verspoor et al. 1999) or in the glacial lakes of Eastern Europe (Consuegra et al. 2002; King et al. 2007), and another refugium was suggested to have existed in Northern France (Finnegan et al. 2013). These observations raise the question of the number of historical refugia involved in the evolutionary history of Atlantic salmon. At first glance the three major clusters, together with our ABC results suggest that three refugia existed. However, our PCA using only European samples (Fig. S3) revealed three clusters but with a closer proximity of the Baltic and Barents-White Sea, with the latter having likely been more strongly introgressed by salmon of the Atlantic group, a pattern shared with many other fishes and marine invertebrates (Bierne et al. 2011). This lends supports for another scenario proposed by Bierne et al. (2011) who suggested that the Baltic and Barents-White Sea clusters might have shared a common history in the past and have been subdivided into two groups by the northward colonization of a southern lineage. Under this scenario, only two refugia would have existed rather than three and our data also lend support for this scenario in which a global signal of postglacial secondary contact is still retained.

These results have important implications for interpreting genetic-environmental associations in this species. In particular, while endogenous barriers are most easily accumulated in allopatry and form tension zones upon secondary contact, their coupling with environmental barriers often stabilizes them (Barton 1979; Barton and Hewitt 1985) resulting in spurious genetic-environmental associations that may incorrectly be interpreted as local adaptation (Bierne et al. 2011). In the Baltic-Atlantic comparison, most of the differentiation observed in several species is generally attributed to adaptation to environmental gradients (e.g., Johannesson and André 2006; Gaggiotti et al. 2009; Berg et al. 2015; Guo et al. 2015) under the hypothesis that the populations have adapted after the establishment of the Baltic sea (<8000 years old). Here, our analysis indicates that this may not necessarily be the case. Without rejecting a potential role of exogeneous barriers, our results show that the null model of demographic history can well account for the observed pattern, as observed in *Drosophila melanogaster* (e.g., Flatt 2016). We therefore propose that for any species, reconstructing the demographic history occurring along environmental gradients where hybrid zones have been documented (e.g., Daguin et al. 2001; Bierne et al. 2003; Riginos and Cunningham 2005; Nikula et al. 2008) will allow constructing appropriate null model to better understand the relative role of demographic history versus environmental adaptation.

DEMOGRAPHIC HISTORY OF NORTH AMERICAN POPULATIONS

Our inferences of broad patterns of population genetic structure revealed the existence of two major North American groups exhibiting a north-south clustering. Among those two groups, 32% of all individuals displayed mixed membership probabilities. This raises the same question as for European populations: does this pattern of contemporary admixture reflect divergence with gene flow or secondary contacts? Here, the ABC analysis between the most differentiated groups provided strong support for secondary contact. However, as for European populations, it was impossible to estimate the timing of divergence of the groups as credible intervals were too large. Second, we identified several European populations as source of admixture together with several major North American sources from the southern part of the range. This suggests that throughout the species' evolutionary history, the colonization of North America by the ancestral European salmon populations has neither been established by a single colonization event nor by a single point of secondary contact. Interestingly, the Narraguagus River is the southernmost sample from our study and appears as the most extensive source of admixture in more northern populations.

Our results also suggest that several colonization events from different European populations led to intercontinental introgression into at least two major North American genetic groups. Indeed, the inference of admixture from two ancestral European branches into the Gaspésie and Anticosti areas further supports a hypothesis of ancestral, multiple colonization events. Therefore, we propose that North American populations most likely represent a mixture of multiple European lineages that varies among populations. As such, our results do not support the hypothesis of a single colonization event and single intercontinental secondary contact. Instead, a model of multiple colonization and contacts seem more plausible as was recently proposed to explain the colonization

of North America by *D. melanogaster* (Bergland et al. 2016) where North American populations would represent a mixture of European and African lineages. Admittedly, it may seem paradoxical to observe a lower genetic diversity of North American populations in spite of potential admixture. However, we argue that admixture resulting from different colonization events has mainly proceeded through a series of founding events involving only a few individuals successfully migrating to North America each time. Under such a scenario, genetic drift would have almost certainly played a major role in reducing genetic diversity, as suggested by the *Treemix* analysis. Finally, given the possible (partial) genetic incompatibility between North American and European genomes, it is plausible that the most recent (postglacial) introgression events have been selected against resulting in little change in the overall pattern of genetic diversity.

Our ABC modeling also suggests that following these colonization events, spatial redistribution of European ancestral variants into North American rivers apparently resulted in a complex signal of admixture integrating multiple signals of split and secondary contacts. Finally, as for Europe, results for North America indicate that studies focusing on interpretation of local adaptation may benefit from accounting for the possible confounding effects of admixture in generating clines in allele frequencies (Kapun et al. 2016).

EVIDENCE FOR GENOME WIDE HETEROGENEITY: A POSSIBLE ROLE FOR LINKED SELECTION?

Another salient result of our study was that models incorporating heterogeneity of either Ne, *m* or both to account for linked selection outcompete models with homogeneous gene flow (except in three out of 148 cases). It is increasingly recognized that integrating heterogeneity of introgression rate (*m*) in demographic inferences increases the accuracy of model selection (Roux et al. 2013; Le Moan et al. 2016; Leroy et al. 2017; Rougemont et al. 2017) as genetic barriers to gene flow are known to reduce the effective rate of migration and to result in heterogeneous landscape of differentiation. Roux et al. (2016) recently demonstrated the importance of integrating local genomic variation in *Ne* to more rigorously model variation in intensity of genetic hitchhiking or to background selection (Charlesworth et al. 1993). More specifically, they showed how neglecting one of these two components can lead to false inferences about intensity of gene flow, local variation in effective population size and ultimately impact on model choice. Moreover, there is mounting evidence supporting a role of linked selection in shaping heterogeneous landscape of differentiation (Burri et al. 2015; Vijay et al. 2016, 2017) and that gene flow is not always necessary to explain heterogeneous divergence along the genome. Here, however, our modeling approach suggests that gene flow associated with secondary contact has played an important role by lowering genome-wide differentia-

tion outside of barrier loci. Admittedly, as most current modeling approaches, our ABC framework does not allow distinguishing the process underlying local variation in *Ne*. It is likely that during most of the isolation period of Atlantic salmon, background selection may have played a role, as inferred recently in Sea Bass (Duranton et al. 2017). However, disentangling the relative contribution of differential introgression, selective sweeps, and background selection was beyond the scope of this study and would require more data than those available so far in Atlantic salmon.

Conclusions

Our results suggest that the demographic history of Atlantic salmon was shaped by multiple secondary contacts both between and within the North American and European continents. Multiple contact zones from European populations in North America, followed by widespread admixture and sorting of ancestral variation seems the most likely scenario for the observed patterns. While differential introgression across the genome certainly played a role in shaping at the pattern of heterogeneous differentiation along the genome, our results also point to a role for linked selection. In these conditions, identifying targets of local adaptation will be particularly challenging. Clearly, more extensive genome-wide data with information about local variation in recombination rate will be needed to address this issue. These data will also allow drawing models of divergence with more than two populations and will allow using a more appropriate null neutral model for detecting targets of selection associated with local adaptation. Clearly, disentangling real targets of adaptation from demographic and other nonadaptive processes may be more challenging than has been appreciated thus far.

LITERATURE CITED

Barton, N., and B. O. Bengtsson. 1986. The barrier to genetic exchange between hybridising populations. Heredity 57:357–376.

Barton, N. H. 1979. Gene flow past a cline. Heredity 43:333–339.

Barton, N. H., and G. M. Hewitt. 1985. Analysis of hybrid zones. Annu. Rev. Ecol. Syst. 16:113–148.

Beaumont, M. A., W. Zhang, and D. J. Balding. 2002. Approximate Bayesian computation in population genetics. Genetics 162:2025–2035.

Berg, P. R., S. Jentoft, B. Star, K. H. Ring, H. Knutsen, S. Lien, K. S. Jakobsen, and C. Andre. 2015. Adaptation to low salinity promotes genomic divergence in Atlantic cod (*Gadus morhua* L.). Genome Biol. Evol. 7:1644–1663.

Bergland, A. O., R. Tobler, J. González, P. Schmidt, and D. Petrov. 2016. Secondary contact and local adaptation contribute to genome-wide patterns of clinal variation in *Drosophila melanogaster*. Mol. Ecol. 25:1157–1174.

Bernatchez, L., and C. C. Wilson. 1998. Comparative phylogeography of Nearctic and Palearctic fishes. Mol. Ecol. 7:431–452.

Bierne, N., P. Borsa, C. Daguin, D. Jollivet, F. Viard, F. Bonhomme, and P. David. 2003. Introgression patterns in the mosaic hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. Mol. Ecol. 12:447–461.

Bierne, N., P.-A. Gagnaire, and P. David. 2013. The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. Curr. Zool. 59:72–86.

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

Bourke, E. 1997. Allozyme variation in populations of Atlantic salmon located throughout Europe: diversity that could be compromised by introductions of reared fish. ICES J. Mar. Sci. 54:974–985.

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

Bradburd, G. S., P. L. Ralph, and G. M. Coop. 2016. A spatial framework for understanding population structure and admixture. PLoS Genet. 12:e1005703.

Bradbury, I. R., L. C. Hamilton, B. Dempson, M. J. Robertson, V. Bourret, L. Bernatchez, and E. Verspoor. 2015. Transatlantic secondary contact in Atlantic salmon, comparing microsatellites, a single nucleotide polymorphism array and restriction-site associated DNA sequencing for the resolution of complex spatial structure. Mol. Ecol. 24:5130–5144.

Burri, R., A. Nater, T. Kawakami, C. F. Mugal, P. I. Olason, L. Smeds, A. Suh, L. Dutoit, S. Bureš, L. Z. Garamszegi, et al. 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. Genome Res. 25:1656–1665.

Burri, R. 2017. Linked selection, demography and the evolution of correlated genomic landscapes in birds and beyond. Mol. Ecol. 26:3853–3856.

Cauwelier, E., J. Gilbey, C. S. Jones, L. R. Noble, and E. Verspoor. 2012. Asymmetrical viability in backcrosses between highly divergent populations of Atlantic salmon (*Salmo salar*): implications for conservation. Conserv. Genet. 13:1665–1669.

Caye, K., T. M. Deist, H. Martins, O. Michel, and O. François. 2016. TESS3: fast inference of spatial population structure and genome scans for selection. Mol. Ecol. Resour. 16:540–548.

Charlesworth, B., M. T. Morgan, and D. Charlesworth. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134:1289–1303.

Christe, C., K. N. Stölting, L. Bresadola, B. Fussi, B. Heinze, D. Wegmann, and C. Lexer. 2016. Selection against recombinant hybrids maintains reproductive isolation in hybridizing *Populus* species despite F1 fertility and recurrent gene flow. Mol. Ecol. 25:2482–2498.

Clark, P. U., A. S. Dyke, J. D. Shakun, A. E. Carlson, J. Clark, B. Wohlfarth, J. X. Mitrovica, S. W. Hostetler, and A. M. McCabe. 2009. The last glacial maximum. Science 325:710–714.

Consuegra, S., C. García De Leániz, A. Serdio, M. González Morales, L. G. Straus, D. Knox, and E. Verspoor. 2002. Mitochondrial DNA variation in Pleistocene and modern Atlantic salmon from the Iberian glacial refugium. Mol. Ecol. 11:2037–2048.

Cruickshank, T. E., and M. W. Hahn. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. 23:3133–3157.

Csilléry, K., O. François, and M. G. B. Blum. 2012. ABC: an R package for approximate Bayesian computation (ABC). Methods Ecol. Evol. 3:475–479.

Cutler, M. G., S. E. Bartlett, S. E. Hartley, and W. S. Davidson. 1991. A polymorphism in the ribosomal RNA genes distinguishes Atlantic salmon (*Salmo salar*) from North America and Europe. Can. J. Fish. Aquat. Sci. 48:1655–1661.

Daguin, C., F. Bonhomme, and P. Borsa. 2001. The zone of sympatry and hybridization of *Mytilus edulis* and *M. galloprovincialis*, as described by intron length polymorphism at locus mac-1. Heredity 86:342–354.

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

Dray, S., and A.-B. Dufour. 2007. The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. 22:1–20.

Duranton, M., F. Allal, C. Fraïsse, N. Bierne, F. Bonhomme, and P.-A. Gagnaire. 2017. The origin and remolding of genomic islands of differentiation in the European sea bass. bioRxiv. https://doi.org/10.1101/223750.

Endler, J. A. 1977. Geographic variation, speciation, and clines. Princeton Univ. Press, Princeton.

Ewing, G. B., and J. D. Jensen. 2016. The consequences of not accounting for background selection in demographic inference. Mol. Ecol. 25:135–141.

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

Fagundes, N. J. R., N. Ray, M. Beaumont, S. Neuenschwander, F. M. Salzano, S. L. Bonatto, and L. Excoffier. 2007. Statistical evaluation of alternative models of human evolution. Proc. Natl. Acad. Sci. 104:17614–17619.

Falush, D., L. van Dorp, and D. Lawson. 2016. A tutorial on how (not) to over-interpret STRUCTURE/ADMIXTURE bar plots. bioRxiv. https://doi.org/10.1101/066431.

Feder, J. L., S. P. Egan, and P. Nosil. 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28:342–350.

Ferchaud, A.-L., C. Perrier, J. April, C. Hernandez, M. Dionne, and L. Bernatchez. 2016. Making sense of the relationships between Ne, Nb and Nc towards defining conservation thresholds in Atlantic salmon (*Salmo salar*). Heredity 117:268–278.

Finnegan, A. K., A. M. Griffiths, R. A. King, G. Machado-Schiaffino, J.-P. Porcher, E. Garcia-Vazquez, D. Bright, and J. R. Stevens. 2013. Use of multiple markers demonstrates a cryptic western refugium and post-glacial colonisation routes of Atlantic salmon (*Salmo salar* L.) in northwest Europe. Heredity 111:34–43.

Flatt, T. 2016. Genomics of clinal variation in *Drosophila*: disentangling the interactions of selection and demography. Mol. Ecol. 25:1023–1026.

Flaxman, S. M., J. L. Feder, and P. Nosil. 2013. Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. Evolution 67:2577–2591.

Frichot, E., and O. François. 2015. LEA: an R package for landscape and ecological association studies. Methods Ecol. Evol. 6:925–929.

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

Gagnaire, P.-A., S. A. Pavey, E. Normandeau, and L. Bernatchez. 2013. The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake white fish species pairs assessed by RAD sequencing. Evolution 67:2483–2497.

Gibbard, P. L., M. J. Head, M. J. C. Walker, and the Subcommission on Quaternary Stratigraphy. 2010. Formal ratification of the Quaternary System/Period and the Pleistocene Series/Epoch with a base at 2.58 Ma. J. Quat. Sci. 25:96–102.

Goudet, J. 2005. HIERFSTAT, a package for r to compute and test hierarchical F-statistics. Mol. Ecol. Notes 5:184–186.

Guillot, G., and M. Foll. 2009. Correcting for ascertainment bias in the inference of population structure. Bioinformatics 25:552–554.

Guo, B., J. DeFaveri, G. Sotelo, A. Nair, and J. Merilä. 2015. Population genomic evidence for adaptive differentiation in Baltic Sea three-spined sticklebacks. BMC Biol. 13:1–18.

Harrison, R. G., and E. L. Larson. 2016. Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. Mol. Ecol. 25:2454–2466.

Hartley, S. E. 1987. The chromosomes of salmonids fishes. Biol. Rev. 62:197–214.

Hewitt, G. 2000. The genetic legacy of the quaternary ice ages. Nature 405:907–913.

———. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. Biol. J. Linn. Soc. 58:247–276.

Hill, W. G., and A. Robertson. 1966. The effect of linkage on limits to artificial selection. Genet. Res. 8:269–294.

Hudson, R. R. 2002. Generating samples under a Wright–Fisher neutral model of genetic variation. Bioinforma. Oxf. Engl. 18:337–338.

Jeffery, N. W., R. R. E. Stanley, B. F. Wringe, J. Guijarro-Sabaniel, V. Bourret, L. Bernatchez, P. Bentzen, R. G. Beiko, J. Gilbey, M. Clément, et al. 2017. Range-wide parallel climate-associated genomic clines in Atlantic salmon. R. Soc. Open Sci. 4:171394.

Johannesson, K., and C. André. 2006. Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea: genetics of marginal Baltic Sea populations. Mol. Ecol. 15:2013–2029.

Kaplan, N. L., R. R. Hudson, and C. H. Langley. 1989. The "hitchhiking effect" revisited. Genetics 123:887–899.

Kapun, M., D. K. Fabian, J. Goudet, and T. Flatt. 2016. Genomic evidence for adaptive inversion clines in *Drosophila melanogaster*. Mol. Biol. Evol. 33:1317–1336.

King, T. L., S. T. Kalinowski, W. B. Schill, A. P. Spidle, and B. A. Lubinski. 2001. Population structure of Atlantic salmon (*Salmo salar* L.): a range-wide perspective from microsatellite DNA variation. Mol. Ecol. 10:807–821.

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

Le Moan, A., P.-A. Gagnaire, and F. Bonhomme. 2016. Parallel genetic divergence among coastal-marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. Mol. Ecol. 25:3187–3202.

Leroy, T., C. Roux, L. Villate, C. Bodénès, J. Romiguier, J. A. P. Paiva, C. Dossat, J.-M. Aury, C. Plomion, and A. Kremer. 2017. Extensive recent secondary contacts between four European white oak species. New Phytol. 214:865–878.

Lien, S., B. F. Koop, S. R. Sandve, J. R. Miller, M. P. Kent, T. Nome, T. R. Hvidsten, J. S. Leong, D. R. Minkley, A. Zimin, et al. 2016. The Atlantic salmon genome provides insights into rediploidization. Nature 533:200–205.

Moore, J.-S., V. Bourret, M. Dionne, I. Bradbury, P. O'Reilly, M. Kent, G. Chaput, and L. Bernatchez. 2014. Conservation genomics of anadromous Atlantic salmon across its North American range: outlier loci identify the same patterns of population structure as neutral loci. Mol. Ecol. 23:5680–5697.

Negre, C., R. Zahn, A. L. Thomas, P. Masqué, G. M. Henderson, G. Martínez-Méndez, I. R. Hall, and J. L. Mas. 2010. Reversed flow of Atlantic deep water during the Last Glacial Maximum. Nature 468:84–88.

Nikula, R., P. Strelkov, and R. VäInöLä. 2008. A broad transition zone between an inner Baltic hybrid swarm and a pure North Sea subspecies of *Macoma balthica* (Mollusca, Bivalvia). Mol. Ecol. 17:1505–1522.

Nilsson, J., R. Gross, T. Asplund, O. Dove, H. Jansson, J. Kelloniemi, K. Kohlmann, A. Löytynoja, E. E. Nielsen, T. Paaver, et al. 2001. Matrilinear phylogeography of Atlantic salmon (*Salmo salar* L.) in Europe and postglacial colonization of the Baltic Sea area. Mol. Ecol. 10:89–102.

Nordborg, M., B. Charlesworth, and D. Charlesworth. 1996. The effect of recombination on background selection. Genet. Res. 67:159–174.

noor, m. a. F., and S. M. Bennett. 2009. Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. Heredity 103:439–444.

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

———. 2009. Age structure, changing demography and effective population size in Atlantic salmon (*Salmo salar*). Genetics 182:1233–1249.

Payseur, B. A., and M. W. Nachman. 2002. Natural selection at linked sites in humans. Gene 300:31–42.

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

Pickrell, J. K., and J. K. Pritchard. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8:e1002967.

Primmer, C. R. 2011. Genetics of local adaptation in salmonid fishes. Heredity 106:401–403.

Quinn, T. P. 1993. A review of homing and straying of wild and hatchery-produced salmon. Fish. Res. 18:29–44.

Reich, D., K. Thangaraj, N. Patterson, A. L. Price, and L. Singh. 2009. Reconstructing Indian population history. Nature 461:489–494.

Riginos, C., and C. W. Cunningham. 2005. Local adaptation and species segregation in two mussel (*Mytilus edulis* × *Mytilustrossulus*) hybrid zones. Mol. Ecol. 14:381–400.

Ross-Ibarra, J., M. Tenaillon, and B. S. Gaut. 2009. Historical divergence and gene flow in the genus *Zea*. Genetics 181:1399–1413.

Ross-Ibarra, J., S. I. Wright, J. P. Foxe, A. Kawabe, L. DeRose-Wilson, G. Gos, D. Charlesworth, and B. S. Gaut. 2008. Patterns of polymorphism

and demographic history in natural populations of *Arabidopsis lyrata*. PLoS ONE 3:e2411.

Rougemont, Q., P.-A. Gagnaire, C. Perrier, C. Genthon, A.-L. Besnard, S. Launey, and G. Evanno. 2017. Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. Mol. Ecol. 26:142–162.

Rougemont, Q., C. Roux, S. Neuenschwander, J. Goudet, S. Launey, and G. Evanno. 2016. Reconstructing the demographic history of divergence between European river and brook lampreys using approximate Bayesian computations. PeerJ 4:e1910.

Rougeux, C., L. Bernatchez, and P.-A. Gagnaire. 2017. Modeling the multiple facets of speciation-with-gene-flow toward inferring the divergence history of lake white fish species pairs (*Coregonus clupeaformis*). Genome Biol. Evol. 9:2057–2074.

Roux, C., V. Castric, M. Pauwels, S. I. Wright, P. Saumitou-Laprade, and X. Vekemans. 2011. Does speciation between *Arabidopsis halleri* and *Arabidopsis lyrata* coincide with major changes in a molecular target of adaptation? PLoS ONE 6:e26872.

Roux, C., C. Fraïsse, J. Romiguier, Y. Anciaux, N. Galtier, and N. Bierne. 2016. Shedding light on the grey zone of speciation along a continuum of genomic divergence. PLoS Biol. 14:e2000234.

Roux, C., G. Tsagkogeorga, N. Bierne, and N. Galtier. 2013. Crossing the species barrier: genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. Mol. Biol. Evol. 30:1574–1587.

Schrider, D. R., A. G. Shanku, and A. D. Kern. 2016. Effects of linked selective sweeps on demographic inference and model selection. Genetics 204:1207–1223.

Seehausen, O., R. K. Butlin, I. Keller, C. E. Wagner, J. W. Boughman, P. A. Hohenlohe, C. L. Peichel, G.-P. Saetre, C. Bank, Å. Brännström, et al. 2014. Genomics and the origin of species. Nat. Rev. Genet. 15:176–192.

Sousa, V. C., M. Carneiro, N. Ferrand, and J. Hey. 2013. Identifying loci under selection against gene flow in isolation-with-migration models. Genetics 194:211–233.

Tavaré, S., D. J. Balding, R. C. Griffiths, and P. Donnelly. 1997. Inferring coalescence times from DNA sequence data. Genetics 145:505–518.

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

Tine, M., H. Kuhl, P.-A. Gagnaire, B. Louro, E. Desmarais, R. S. T. Martins, J. Hecht, F. Knaust, K. Belkhir, S. Klages, et al. 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat. Commun. 5:5770.

Tonteri, A., A. J. Veselov, S. Titov, J. Lumme, and C. R. Primmer. 2007. The effect of migratory behaviour on genetic diversity and population divergence: a comparison of anadromous and freshwater Atlantic salmon *Salmo salar*. J. Fish Biol. 70:381–398.

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

Verspoor, E., J. A. Beardmore, S. Consuegra, C. García de Leániz, K. Hindar, W. C. Jordan, M.-L. Koljonen, A. A. Mahkrov, T. Paaver, J. A. Sánchez, et al. 2005. Population structure in the Atlantic salmon: insights from 40 years of research into genetic protein variation. J. Fish Biol. 67: 3–54.

Verspoor, E., E. M. McCARTHY, D. Knox, E. A. Bourke, and T. F. Cross. 1999. The phylogeography of European Atlantic salmon (*Salmo salar* L.) based on RFLP analysis of the ND1/16sRNA region of the mtDNA. Biol. J. Linn. Soc. 68:129–146.

Via, S., and J. West. 2008. The genetic mosaic suggests a new role for hitchhiking in ecological speciation. Mol. Ecol. 17:4334–4345.

Vijay, N., C. M. Bossu, J. W. Poelstra, M. H. Weissensteiner, A. Suh, A. P. Kryukov, and J. B. W. Wolf. 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. Nat. Commun. 7:13195.

Vijay, N., M. Weissensteiner, R. Burri, T. Kawakami, H. Ellegren, and J. B. W. Wolf. 2017. Genome-wide signatures of genetic variation within and between populations—a comparative perspective. bioRxiv. https://doi.org/10.1101/104604.

Wolf, J. B. W., and H. Ellegren. 2016. Making sense of genomic islands of differentiation in light of speciation. Nat. Rev. Genet. https://doi.org/10.1038/nrg.2016.133.

Wu, C.-I. 2001. The genic view of the process of speciation. J. Evol. Biol. 14:851–865.

Associate Editor: S. Baird
Handling Editor: M. Noor

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.