Genomics and telemetry suggest a role for migration harshness in determining overwintering habitat choice, but not gene flow, in anadromous Arctic Char

Jean-Sébastien Moore[1], Les N. Harris[2], Jérémy Le Luyer[1,3], Ben J. G. Sutherland[1,4], Quentin Rougemont[1], Ross F. Tallman[2], Aaron T. Fisk[5] & Louis Bernatchez[1]

[1] Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Québec, Canada
[2] Freshwater Institute, Fisheries and Oceans Canada, Winnipeg, Manitoba, Canada
[3] Institut Français de Recherche pour l'Exploitation de la Mer, Taravao, Tahiti, Polynésie Française
[4] Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, British Columbia, Canada
[5] Great Lakes Institute of Environmental Research, University of Windsor, Windsor, Ontario, Canada

Corresponding author: jean-sebastien.moore@bio.ulaval.ca

Running head: Migration ecology of Arctic Char

Keywords: population genomics, RADseq, genotyping by sequencing, fish migration, Arctic, fishery management, conservation

Abstract

Migration is a ubiquitous life history trait with profound evolutionary and ecological consequences. Recent developments in telemetry and genomics, when combined, can bring significant insights on the migratory ecology of non-model organisms in the wild. Here, we used this integrative approach to document dispersal, gene flow and potential for local adaptation in anadromous Arctic Char from six rivers in the Canadian Arctic. Acoustic telemetry data from 124 tracked individuals indicated asymmetric dispersal, with a large proportion of fish (72%) tagged in three different rivers migrating up the same short river in the fall. Population genomics data from 6,136 SNP markers revealed weak, albeit significant, population differentiation (average pairwise $F_{ST} = 0.011$) and asymmetric dispersal was also revealed by population assignments. Approximate Bayesian Computation simulations suggested the presence of asymmetric gene flow, although in the opposite direction to that observed from the telemetry data, suggesting that dispersal does not necessarily lead to gene flow. These observations suggested that Arctic Char home to their natal river to spawn, but may overwinter in rivers with the shortest migratory route to minimize the costs of migration in non-breeding years. Genome scans and genetic-environment associations identified 90 outlier markers putatively under selection, 23 of which were in or near a gene. Of these, at least four were involved in muscle and cardiac function, consistent with the hypothesis that migratory harshness could drive local adaptation. Our study illustrates the power of integrating genomics and telemetry to study migrations in non-model organisms in logistically challenging environments such as the Arctic.

Introduction

Migrations are a common feature of the life history of many animal species (Dingle 2014). Despite their obvious evolutionary and ecological consequences, migrations have been a challenging topic of study. Quantifying migratory phenotypes is difficult because they can occur over vast distances. Furthermore, while some migratory traits are amenable to laboratory studies and more classical quantitative genetic approaches (e.g. Berthold & Querner 1981; Roff & Fairbairn 1991), many migratory traits are impossible to recreate in experimental settings. Recent technological developments in telemetry techniques both on land (Kays *et al.* 2015) and underwater (Hussey *et al*. 2015) now allow observations of animal movement in the wild over broad spatial scales and with unprecedented levels of detail. Parallel to these developments is the exponential increase in the availability of genomic technologies for non-model organisms (Davey *et al.* 2011, Andrews *et al.* 2016). These new genomic tools make genome-wide assessments of genetic variation possible, thus offering novel ways to link genotypes to migratory phenotypes in the wild for non-model organisms (Liedvogel *et al.* 2011; Hess et al. 2014; Shafer *et al.* 2016; Franchini *et al.* 2017). The integration of telemetry and genomic datasets, therefore, provides a powerful approach to document the population level consequences of migrations and to better understand the genetic basis of migratory traits (Shafer *et al.* 2016).

One population level consequence of migration is that it can redistribute genetic variation. Indeed, while homing to reproduction sites is commonly associated with migrations (Dingle 2014), it is rarely perfect, thus leading to dispersal and gene flow. Consequently, the precision of homing and the spatial scale over which migrations take place influence the scale over which gene flow, and thus genetic structure, can be observed (e.g., Castric & Bernatchez 2004). Furthermore, inter-individual differences in migratory behaviour are common and can influence gene flow (e.g., Turgeon *et al.* 2012; Shafer *et al*. 2012; Delmore *et al*. 2015). Therefore, the study of how migratory behaviour, dispersal, and gene flow interact to determine the genetic structure of populations and their capacity to locally adapt can greatly benefit from integrating genomic and telemetry data.

The diversity of migratory life histories in salmonids makes them excellent model systems for the study of migration ecology (Hendry *et al*. 2004; Quinn 2005). Anadromy, a trait shared by many salmonid species, refers to a migratory life cycle whereby individuals are born in freshwater, feed and grow in saltwater, and return to freshwater to reproduce. In most salmonids, return migrations to freshwater occur exclusively for the purpose of reproduction, but some species, notably in the genus *Salvelinus*, must also return to freshwater to overwinter (e.g., Johnson 1980; Moore *et al.* 2013; Bond *et al*. 2015). The Arctic Char (*Salvelinus alpinus*) is a facultatively anadromous salmonid with a circumboreal distribution (Johnson 1980; Klemetsen *et al*. 2003; Reist *et al*. 2013). Anadromous individuals undergo annual feeding migrations to the ocean and must return to freshwater in the fall because winter conditions in the Arctic Ocean are not favorable (Johnson 1980; Klemetsen *et al*. 2003; see Jensen & Rikardsen 2012 for an exception). The summer feeding and growth period is thus limited by the ice-free period on the rivers used for migrations, which can be as short as a month at higher latitudes (Johnson 1980). This restricted feeding period limits energy gains, and results in skipped reproduction such that in most populations, spawning occurs once every two to four years (Dutil 1986). The rate of iteroparity is among the highest reported for anadromous salmonids, with 32-50% of individuals observed breeding more than once in some populations (Fleming 1998). The migratory behaviour of anadromous Arctic Char is therefore best understood as three distinct migrations: (1) spring feeding migrations to the ocean; (2) fall spawning migrations to freshwater spawning sites in headwater lakes; and (3) fall overwintering migrations to freshwater overwintering sites in headwater lakes when the individual is not in breeding condition (Fig. 1). Given that optimal spawning habitats likely differ from optimal overwintering habitats, we can predict that individuals might select different habitats in different years depending on their maturity status. Accordingly, there is evidence that Arctic Char home to their natal sites to spawn, but often utilize non-natal overwintering sites in years when they are not in breeding condition (Johnson 1980; Gyselman 1994; Moore *et al*. 2013; Gilbert *et al*. 2016). Natal homing to reproduction sites combined with the use of alternative overwintering sites may lead to temporary mixing of different stocks in freshwater, making both population-specific sampling and fisheries management challenging. The integration of telemetry

and genomic datasets is therefore particularly promising for studying Arctic Char migrations.

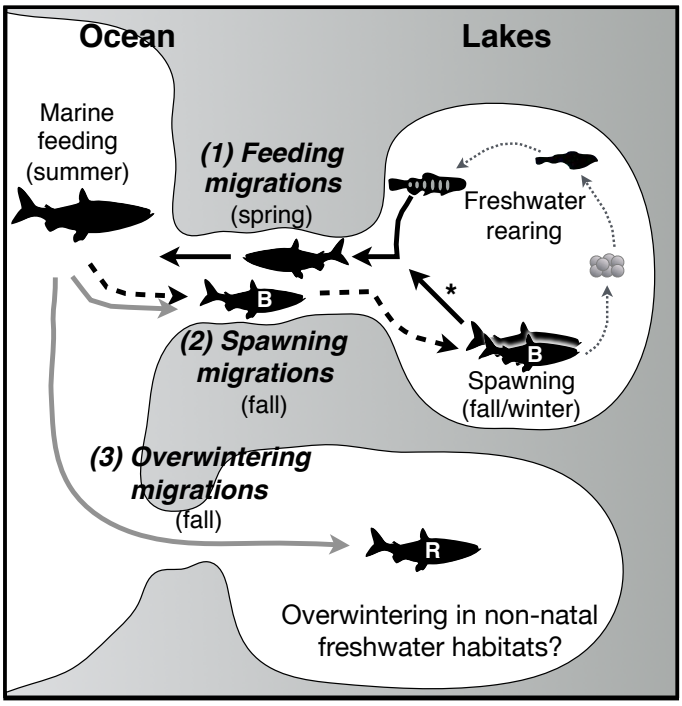

Figure 1. The anadromous Arctic Char life cycle highlighting three distinct migrations. (1) Feeding migrations (black arrows): after hatching and rearing in freshwater for several years (grey dotted arrows), individuals smoltify and begin to migrate to saltwater to feed during the summer when the rivers and ocean are free of ice. (2) Spawning migrations (dashed black arrows): individuals in breeding condition (B) return to their natal freshwater habitat to spawn. Because individuals are not in breeding condition every year, spawning migrations occur only once every 2-4 years. The arrow with the asterisk indicates that Arctic Char are iteroparous and can resume feeding migrations the following spring after spawning (3) Overwintering migrations (full grey arrows): in the years when individuals are not in breeding condition (resting (R)), they must still return to freshwater to overwinter to avoid lower temperatures and increased salinity in the marine environment. Because optimal conditions for spawning and overwintering may differ, use of non-natal overwintering sites might be more frequent than use of non-natal spawning sites. We hypothesized that individuals would favor less harsh migrations for overwintering, as indicated by the shorter river. Silhouette of adult fish from PhyloPic.org.demographic

A key variable driving the evolution of migrations and associated traits is migration distance. Migration distance greatly influences the costs of migrations, and strategies that minimize these costs will tend to be favored (Roff 2002). An example is the rapid evolution of a shorter migratory route in the warbler from continental Europe, the blackcap (*Sylvia atricapilla*), to take advantage of milder conditions in Britain (Berthold *et al*. 1992). Freshwater migration lengths and elevation gain (i.e. their

'harshness') are known to have a strong effect on the bioenergetic costs of migrations in anadromous fishes, including salmonids (Bernatchez & Dodson 1987). Migratory harshness can drive local adaptation of life history (Schaffer & Elson 1975), morphological (Crossin et al. 2004), and physiological traits (Eliason *et al*. 2011). While studies from Arctic Char have shown the influence of habitat accessibility on freshwater migrations (Gyselman 1994; Gilbert et al. 2016), the extent to which migration harshness drives habitat choice (particularly during overwintering migrations) and local adaptation remains largely unknown.

We combined double-digest RADseq (Andrews et al. 2016) and acoustic telemetry to study the migrations of anadromous Arctic Char from southern Victoria Island in the Canadian Arctic Archipelago (Fig. 2). The largest commercial fishery for Arctic Char in Canada has been operating in this region for over five decades (Day & Harris 2013) and the local Inuit people rely on this resource for economic, subsistence, and cultural purposes (Kristofferson & Berkes 2005). At least seven rivers in this region support populations of Arctic Char, five of which are commercially fished. Previous studies in the region have demonstrated that many of the stocks mix at sea (Dempson & Kristofferson 1987; Moore *et al.* 2016), and that genetic differentiation estimated from microsatellite markers is low (Harris *et al*. 2016). We first tested the hypothesis that Arctic char in the region home to natal rivers to spawn, but may overwinter in non-natal freshwater systems. This hypothesis predicts that individuals would use different freshwater systems in different years, but that this movement to non-natal streams would not necessarily result in gene flow. We used acoustic telemetry data to show that many individuals used different freshwater habitats in different years, and used population genomic data to infer the natal origins of migrating individuals, thus confirming that dispersal to non-natal habitats occurs. We used an Approximate Bayesian Computation (ABC) framework (Csilléry *et al.* 2010) to determine that an observed asymmetry in dispersal towards the Ekalluk River was not associated with increased gene flow in that direction. This further supports the hypothesis that dispersal does not necessarily lead to gene flow. We second hypothesized that overwintering migrations to non-natal freshwater systems would predominantly occur towards the system offering the least costly migration, here the Ekalluk River (Fig. 2; Table 1), given the central role of

migratory harshness in driving survival and local adaptation in other salmonids. Consistent with our hypothesis, both the acoustic telemetry data and the population genomic data suggested a highly asymmetric pattern of dispersal towards the Ekalluk River. Finally, exploratory analyses using genome scans and genotype-environment associations identified outlier loci potentially associated with local adaptation. These markers were subsequently characterized using available genomic information from related salmonids. Together, our analyses provided new insights on patterns and consequences of migrations in a species of ecological, economic, and cultural importance.

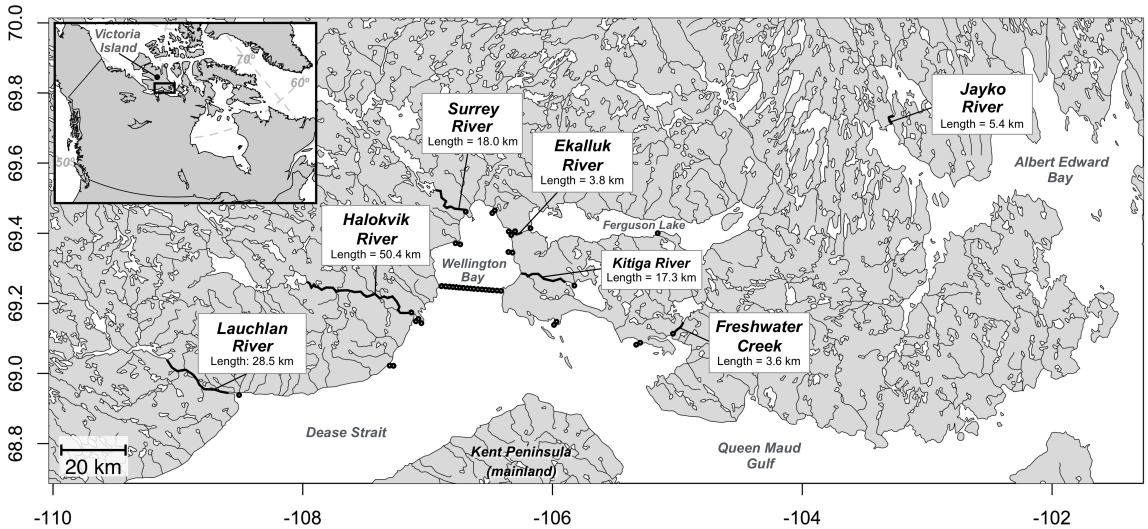

Figure 2. Map of the study area highlighting the names and lengths of the rivers sampled. The black circles with a white center represent the locations of the moored acoustic receivers used to track the movements of Arctic Char surgically implanted with transmitters at the Ekalluk, Surrey and Halokvik rivers.

Materials and Methods

Telemetry data and sampling

Detailed methods for the acoustic telemetry data collection can be found in Moore *et al*. (2016). In short, 42 moored Vemco VR2W passive acoustic receivers were deployed between July 13 and September 8, 2013 (Fig. 2) along the southern shore of Victoria Island, Nunavut, near the community of Cambridge Bay. Detections were continually recorded until receivers were last retrieved between July 27 and August 31 2016. There are at least seven watersheds in the region that support runs of anadromous Arctic Char, and these watersheds are drained by rivers that vary in length and elevation gain (Fig. 2 & Table 1). We placed a receiver in six of these seven rivers to detect

migrating fish (the Jayko River was excluded from the telemetry study because of the large distance between this site and the others). A total of 126 adult Arctic Char (>500 mm) were surgically implanted with Vemco V16 acoustic transmitters at three tagging locations (Table 1; only 124 were detected on the array and have telemetry data available). In all cases, the fish were collected at the river mouth near the ocean, and for the Ekalluk and Surrey River, tagging occurred during the downstream migration (early July), whereas tagging at the Halokvik River targeted upstream migrants (mid/late-August). A fin biopsy was taken from each tagged fish for genomic analysis. Additional baseline samples for genetic analysis were taken from sampling conducted during the upstream migrations to monitor commercial stocks or were collected from commercial fishery catches (Table 1). Samples from the Kitiga River (N69.29° W106.24°), which harbors an Arctic Char population, were also obtained, but DNA extractions for these samples failed. All of the sampling targeted adults but no information was available about the breeding status of sampled individuals. In the absence of this information, we refer to the use of non-natal freshwater systems as 'dispersal' throughout the manuscript, while we use the term 'homing' to refer to the use of natal habitat regardless of the purpose of the migration. The natal river of migratory adults cannot be determined with absolute confidence and is instead inferred from a combination of information on the sampling location, the telemetry data, and/or the genomic data. We also refer to all migrations back to freshwater in the fall as 'return migrations' regardless of whether an individual is dispersing or homing.

Library preparation, sequencing, and SNP calling

A salt-extraction protocol adapted from Aljanabi & Martinez (1997) was used to extract genomic DNA. Sample quality and concentration were checked on 1% agarose gels and a NanoDrop 2000 spectrophotometer (Thermo Scientific). Concentration of DNA was normalized to 20 ng/µl (volume used = 10 µl; 200 ng total) based on PicoGreen assays (Fluoroskan Ascent FL, Thermo Labsystems). Libraries were constructed and sequenced on the Ion Torrent Proton platform following a double-digest RAD (restriction site-associated DNA sequencing; Andrews *et al.* 2016) protocol modified from Mascher *et al*. (2013). In short, genomic DNA was digested with two restriction enzymes (*PstI* and

MspI) by incubating at 37°C for two hours followed by enzyme inactivation by incubation at 65°C for 20 minutes. Sequencing adaptors and a unique individual barcode were ligated to each sample using a ligation master mix including T4 ligase. The ligation reaction was completed at 22°C for 2 hours followed by 65°C for 20 minutes to inactivate the enzymes. Samples were pooled in multiplexes of 48 individuals, insuring that individuals from each sampling location were sequenced as part of at least 6 different multiplexes to avoid pool effects. One randomly chosen replicate per sampling location was also run on a different multiplex to evaluate sequencing errors (Mastreta-Yanes *et al.* 2015). Libraries were size-selected using a BluePippin prep (Sage Science), amplified by PCR and sequenced on the Ion Torrent Proton P1v2 chip. The obtained reads were aligned to the closely related Rainbow Trout (*Oncorhynchus mykiss*) genome (Berthelot *et al.* 2014) using GSNAP v2016-06-09 (Wu *et al.* 2016), then from these alignments STACKS v.1.40 (Catchen *et al.* 2013) was used for SNP identification, SNP filtering and genotyping (see online supplementary materials).

Genomic data: basic statistics, population structure, and identification of putative dispersers

Observed (H_O) and expected heterozygosity (H_E), and the inbreeding coefficient (G_{IS}) were estimated using GenoDive v2.0b27 (Meirmans and Van Tienderen 2004). Effective population size (N_E) was estimated for the baseline samples using the linkage disequilibrium method in NeEstimator V2.01 (Do *et al.* 2013) with a critical value for rare alleles of 0.05. Pairwise population differentiation was quantified using F_{ST} (Weir & Cockerman 1984) and their significance values estimated with 1000 permutations also in GenoDive. A Discriminant Analysis of Principal Components (DAPC; Jombart *et al*. 2010) in the R package adegenet (Jombart 2008) was used to describe population structure. First, the 'find.clusters' function was used with the number of clusters *K* varying from 1 to 12. PCA scores of individuals were then plotted and the 'compoplot' function was used to calculate their proportion of membership to the genetic clusters identified. ADMIXTURE (Alexander *et al.* 2009) was also run varying the number of clusters *K* between 1 and 12. Putative dispersers were identified on the basis of cluster membership probabilities in DAPC and ADMIXTURE. Putative dispersers were defined

as having a >75% probability of membership to a different cluster than the majority of other sampled individuals at that location. These putative dispersers, as well as individuals that had less than 75% probability of membership to any genetic clusters (i.e. putatively admixed individuals) were removed from some analyses (when noted) to avoid biases. Note that both of these thresholds are purposefully low to ensure all putative dispersers are removed.

Population assignments

The software *gsi_sim* (Anderson *et al*. 2008) implemented in the R package *AssigneR* (Gosselin *et al*. 2016) was used to perform population assignments and test the power of assignment tests. We first simulated the assignment power of the baseline samples (see Table 1) and evaluated the impact of the number of markers used for assignment on power. Markers were ranked based on their F_{ST} values computed from a 'training' set of individuals comprising 50% of the total sample size per population and assignment success was computed for the remaining 'hold-out' individuals to avoid high-grading bias (Anderson 2010). The analysis was repeated for ten randomly generated training sets, and with ten randomly generated subsets of 30 individuals per population to eliminate any biases that could arise due to unequal sample sizes. We used the package *RandomForestSRC* (Ishwaran & Kogalur 2015) implemented in *AssigneR* to impute missing data based on allele frequencies per population using 100 random trees and 10 iterations and compared assignment success with and without imputations. We repeated these analyses on a dataset from which we removed the putative dispersers identified with the clustering analyses (see previous section for details). Next, we treated the acoustically tagged individuals as samples of unknown origin and assigned them back to the baseline samples using the function *assignment_mixture* in *AssigneR*. To avoid biasing the accuracy of the assignment with the missing data imputation, for each individual we only used the allele frequencies of the loci for which genotypes were present – the number of markers used for the assignments in this case therefore varied among individuals (min = 4,633; max = 5,966).

Demographic simulations

To better understand the relative contributions of recent divergence and current gene flow on observed population structure, and to verify whether the observed patterns of dispersion translated into realized gene flow, an ABC pipeline using coalescent simulations was used (Csilléry et al. 2010). Following other approaches (e.g., Illera *et al*. 2014), three demographic models were compared and demographic parameters for the best model were estimated. The analyses focused on comparing the Ekalluk and Lauchlan samples and then the Ekalluk and Halokvik samples to limit computational demands. The datasets from which putative dispersers were removed were used for this analysis. Furthermore, these comparisons are the most relevant to test our hypotheses as they are representative of the eastern (Ekalluk) and western (Lauchlan and Halokvik) group of populations identified as the first level of hierarchical structure (see Results). We tested three models: (1) a null hypothesis of panmixia; (2) a model of population divergence with gene flow (i.e. the isolation with migration model, IM); and (3) a model of population divergence without gene flow (strict isolation, SI). The simplest model of panmixia was controlled by a single parameter, $\theta = 4\ N_1\mu$, where N_1 is the effective size of the panmictic population and μ is the mutation rate per generation. The SI model is characterized by six demographic parameters, namely $\theta_1=4N_1\mu$, $\theta_2=4N_2\mu$, and $\theta_A=4N_A\mu$, with N_1 and N_2 the effective population size of the two daughter populations and N_A the effective size of the ancestral population, respectively. All these parameters are scaled by $\theta_1=4N_{ref}\mu$ where N_{ref} is the effective size of a reference population. The two daughter populations diverged from the ancestral T_{split} generations ago with $\tau = T_{split}/4N_{ref}$. The IM model was further characterized by the migration rates $M_1 = 4Nm$ (migration into population 1 from population 2) and $M_2 = 4Nm$ (migration into population 2 from population 1) sampled independently and where Nm is the effective migration rate. Large prior distributions following those commonly applied were used and drawn from uniform distributions (e.g., Sousa & Hey 2013). Coalescent simulations ($n = 10^6$) were performed under each model using msnam (Ross-Ibarra *et al*. 2008), which is a modified version of the widely used ms simulator (Hudson 2002), under the infinite site model of DNA mutation. The pipeline of Roux *et al*. (2013) was used with modifications to test for panmixia, and priors were computed with a Python version of priorgen (Ross-Ibarra *et al*.

2008). Details of the summary statistics used, the model selection procedure, analysis of robustness, parameter estimations and posterior predictive checks can be found in the online supplementary materials. Values were not transformed into biological units since the mutation rate (μ) was unknown. Instead coalescent units were kept, and only ratios of θ and M_1/M_2 were interpreted.

Genome scans and functional annotation of outliers

Genome scans using different methods to identify outlier loci often only detect partially overlapping sets of loci (Gagnaire *et al*. 2015). A common practice to partially circumvent these problems is to combine genome scans with genetic-environment association (GEA) methods as a way to more reliably identify the most likely targets of selection (De Villemereuil *et al.* 2014). Therefore, we identified markers putatively under selection using two genome scan methods and a GEA method. First, Bayescan v1.2 (Foll & Gaggiotti 2008) was used to detect outlier loci with elevated F_{ST} among the baseline populations (Table 1) with 5,000 iterations, and a burn-in length of 100,000. To increase our chances of finding outliers while also limiting type-two errors, we tested values of 100, 1,000 and 5,000 for the prior odds (Lotterhos & Whitlock 2014). This analysis was run on the dataset from which putative dispersers were removed to avoid biases (n = 273 after dispersers removed). Second, we used PCAdapt (Luu *et al*. 2017), which uses PCA to control for population structure without *a priori* population definition, a particularly useful feature in the current study where we expect significant admixture among populations. Because of this feature, we ran the analysis on the full dataset including the putative dispersers, which also increased the analytical power associated with increased sample size (n = 318). Simulations have also shown that PCAdapt is less prone to type-2 errors than Bayescan, and to be less affected by the presence of admixed individuals in the samples (Luu *et al*. 2017). The optimal number of principal components necessary to describe population structure was determined using the graphical method described in Luu *et al.* (2017) and varying *K* from 1 to 10. Finally, we also used latent factor mixed models (LFMM; Frichot *et al*. 2013) implemented in the R package LEA to test directly for correlations between allele frequencies at specific loci and migratory difficulty. We also varied K from 1 to 10 to identify the optimal number of latent factors required to

describe population structure. Migration harshness was measured as 'work' (Crossin *et al*. 2004), that is the product of river length and altitude gain, standardized so that the mean over all populations was of zero with a standard deviation of one. Because the environmental data had to be assigned to a sampling location, we used the dataset with the putative dispersers removed for this analysis (n = 273). As recommended by the authors in both user manuals, for PCAdapt and LFMM, the missing data were imputed as before using package *RandomForestSRC* (100 random trees and 10 iterations). All three approaches controlled for false discovery rates (FDR) with an α of 0.05.

An Arctic Char reference genome is in preparation (Macqueen et al. 2017), but it was not available at the time these analyses were conducted. Given the absence of a reference genome, we identified approximate locations for the markers by using the recently developed high-density linkage map for Arctic Char (Nugent *et al*. 2017). To increase the number of markers being positioned, the sex-averaged consensus map was used, including female- and male-specific markers. Although male salmonids have a low recombination rate in general (Sakamoto *et al.* 2000) and thus markers on the linkage map may not be as well positioned within linkage groups due to less frequent recombination, here we included them since we were interested in approximate general positions and the linkage group to which each marker belonged. Pairing of anonymous markers onto the linkage map was done using MapComp run iteratively, as previously described (Sutherland *et al.* 2016; Narum *et al.* 2017). In brief, MapComp aligns the flanking sequence of RAD markers of two genetic maps (here the Arctic Char genetic map and the anonymous Arctic Char markers) to a reference genome of a closely related species. As per Sutherland *et al.* (2016), here we used the Atlantic salmon genome as the closely related reference genome. Markers from each input comparison that are closest to each other in nucleotide position on the reference genome (and within a set distance) are then paired. The position of the paired marker is then given to the anonymous marker. This pairing was conducted with a max distance of 1 Mb and done in 10 iterations, each time removing paired anonymous Arctic Char markers then rerunning the pairing with the Arctic Char genetic map to permit pairing of more than one anonymous marker with a single genetic map marker, as previously described (Narum *et al.* 2017).

All significant outliers were also used in a BLAST query against the Atlantic Salmon genome, which had the best annotation available at the time these analyses were conducted (Lien et al. 2016; NCBI Genome ICSASG_v2 reference Annotation Release 100; GCA_000233375.4). Outlier markers that were found by more than one method, but did not BLAST unambiguously to the Atlantic Salmon genome, were also checked against the Rainbow Trout (*Oncorhynchus mykiss*) genome (Berthelot *et al.* 2014).

Results

Telemetry data

The receivers located at the mouth of each river allowed the inference of likely freshwater overwintering/spawning sites of 98 acoustically tagged Arctic Char for up to three consecutive years (Table S1 in online supplementary materials). The telemetry data suggested a pattern of asymmetric dispersal, where many fish from all three tagging locations were detected one or more years in the Ekalluk River (Fig. 3). The Ekalluk River is the shortest and least harsh (i.e. least vertical and horizontal distance) river around Wellington Bay (Fig. 1 and Table 1), and one of the shortest in the entire area. Of the 47 fish tagged in both 2013 and 2014 at the Ekalluk River, 44 returned to this river each year (94%). Of the 21 fish tagged at the Surrey River, 16 returned to the Ekalluk River each year they were detected (76%), while only one fish returned to the Surrey River (5%). We therefore hypothesized that the fish tagged at the Surrey River were likely Ekalluk River fish that we intercepted moving north and west, an interpretation consistent with observations that many Ekalluk River fish visit the Surrey River estuary shortly after their out-migrations to the ocean (Moore *et al*. 2016) and also consistent with the population genomic data presented below. Of the 30 fish tagged at the Halokvik River, 28 returned to Halokvik River at least once (93%), and six returned to the Ekalluk River at least once (20%). Furthermore, of the 12 fish with more than one year of data that were inferred to have returned to the Halokvik River at least once, seven (58%) migrated to the Ekalluk River during other years, demonstrating that an individual might readily utilize different freshwater sites during its lifetime. In summary, a majority of fish from all three tagging locations (72%) migrating to the Ekalluk River at least once.

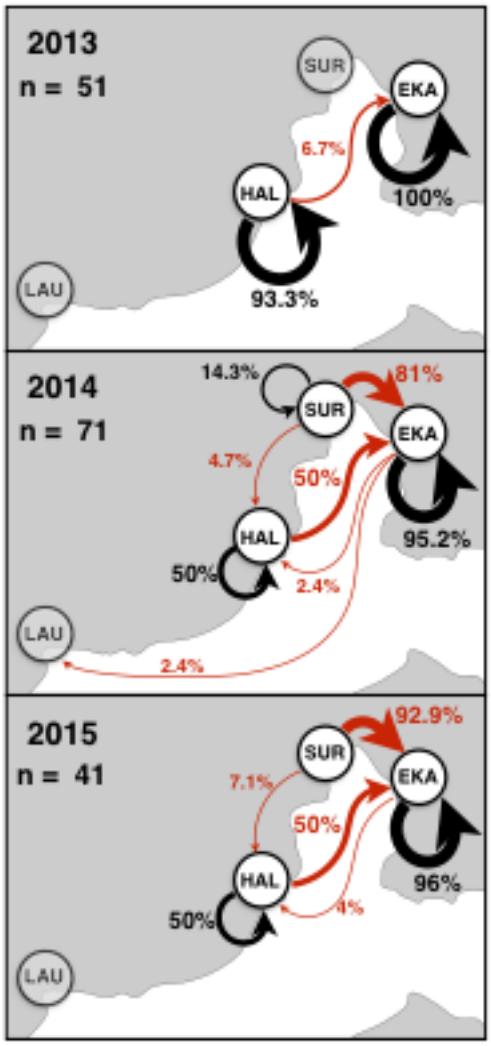

Figure 3. Choice of overwintering/spawning habitat by Arctic Char from three tagging locations inferred from acoustic telemetry data collected over three summers (2013, 2014, and 2015). The proportion of individuals returning to the same location where they were tagged (black) or dispersing to a different location from that where they were tagged (red) are indicated. The thickness of the arrows is roughly proportional to the proportion of individuals observed migrating to this location. The map indicates the approximate locations of Lauchlan (LAU), Halokvik (HAL), Surrey (SUR), and Ekalluk (EKA) rivers (shaded labels indicate no individuals were tagged there that year). The numbers of individuals that were unambiguously assigned to an overwintering location each year (n) are indicated.

Sequencing and SNP calling

After cleaning and demultiplexing, a total of 1.9 billion reads were left with an average of 4.0 million reads per individual (coefficient of variation: 30.3%). The assembly resulted in a catalog containing 2,963,980 loci and a pre-filter total of 13,568,653 SNPs (over 1,272,427 polymorphic loci) after the population module of STACKS (Table S3 in supplementary materials). Individuals with more than 25% missing genotypes were removed from all analyses (Table 1), and high-quality SNPs were retained after filters (Table S3). In addition, 94 putatively sex-linked markers were removed (see Fig. S4 & S5 in supplementary materials and Benestan et al. 2017 for details) leaving a total of 6,136 SNPs used in all subsequent analyses (Table S3). Replicate individuals (n = 11) sequenced twice had identical genotypes at 92.0-97.1% of the markers (mean 94.6%), which was comparable to error rates reported in Mastreta-Yanes et al. (2015).

Genomic data: basic statistics, population structure, and identification of dispersers

Levels of heterozygosity were comparable among populations (H_O: 0.099-0.105; H_E: 0.103-0.107), and although all G_{IS} values were significantly positive, indicating an excess

of homozygotes, the values were all small (0.001-0.037) (Table 1). Effective population size estimates varied between 335 (Lauchlan River; 95% c.i.: 318-353) and 1081 (Ekalluk River; 95% c.i.: 935-1280). A weir enumeration study conducted in 1979-1983 found the Ekalluk River to be the most abundant population in the region, and the Lauchlan River the least abundant, but the estimates of census size were much larger than the N_E estimates reported here (183,203 for Ekalluk River and 10,850 for Lauchlan; McGowan 1990). Population differentiation among rivers was weak albeit significant, with an overall F_{ST} of 0.011 (95% c.i. 0.010-0.012; Fig 4a). The F_{ST} value between Ekalluk and Surrey rivers was substantially lower than the other values (F_{ST} = 0.001; 95% c.i. 0.0006-0.0016), which was in agreement with our telemetry observations that most fish captured at Surrey migrated to the Ekalluk River in the fall. It is thus likely that sampling at the Surrey River resulted in the interception of Ekalluk River fish migrating through the area, and those two sampling locations were combined for subsequent analyses unless otherwise noted. When the putative dispersers were removed (*see below*), the overall F_{ST} increased to 0.014 (95% c.i. 0.013-0.015; Fig. 4a). The presence of population structure between most sampling locations was also supported by the PCA analysis, although there was some important overlap among some sampling sites (Fig. 4b). The sampling locations that were most separated in the PCA (Jayko and Lauchlan rivers) were also the most geographically distant sampling locations. The PCA analysis was also repeated on a data set from which putative outlier loci were removed and returned essentially identical results (Fig. S8 online supplementary materials).

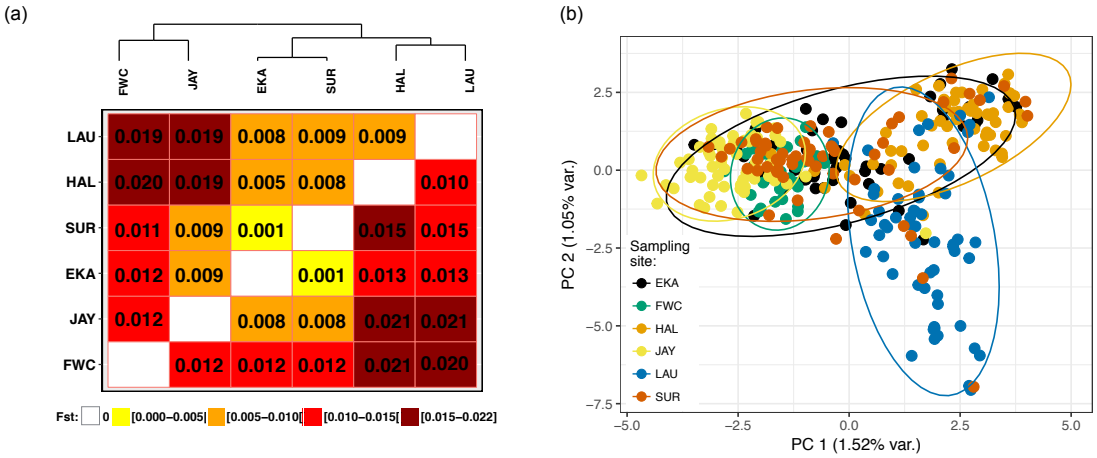

Figure 4. Description of population structure among the Arctic Char sampling sites: (a) heatmap of pairwise F_{ST} values before (above the diagonal) and after (below the diagonal) putative dispersers were removed from the samples (see text for details). All values are significantly different from zero ($p < 0.05$). (b) Results of the principal components analysis showing PC scores of each individual along the first two principal axes (1.52% and 1.05% of the total variance explained respectively). Individuals and the 95% confidence ellipses are colour-coded by sampling location.

The Bayesian Information Criterion in the DAPC analysis best supported the presence of two genetic clusters (Fig. S6 online supplementary materials) differentiating the two westernmost populations (Lauchlan and Halokvik) from all others (Fig. 5). Based on cluster membership, several putative dispersers and admixed individuals could be identified (Fig. 5). A majority of putative dispersers (38/45; 85.4%) were individuals belonging to the western genetic cluster, but sampled in the Ekalluk or Surrey River. In contrast, many fewer dispersers from the eastern group were identified in the western sampling locations (three in Lauchlan and three in Halokvik; i.e., 13.3% of putative dispersers). In other words, the DAPC analysis supported the conclusion of asymmetric dispersal from the western sampling populations towards the shorter and less harsh Ekalluk River. Unlike the DAPC analysis, the cross-validation errors in the ADMIXTURE analysis did not support the presence of two genetic groupings (K = 1 had the lowest cross-validation error). Nonetheless, at $K = 2$, the individual Q-values were consistent with the results of the DAPC, and all putative dispersers identified with the DAPC had >50% probability of membership to the alternative genetic cluster in ADMIXTURE as well. Furthermore, at K=5 the ADMIXTURE results suggested that genetic differentiation among rivers was present, albeit weak.

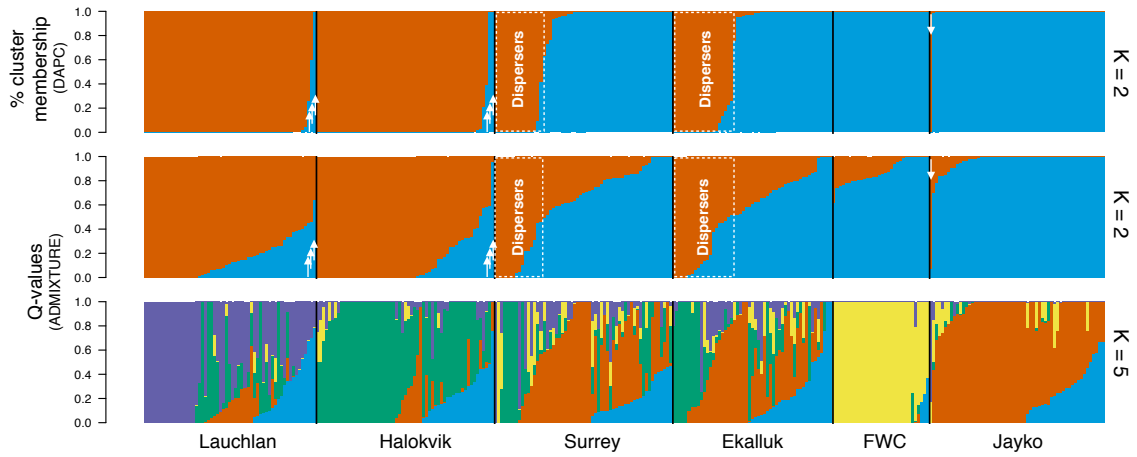

Figure 5. Results of two clustering analyses used to identify putative dispersers. The top panel is a compoplot generated from a DAPC using two genetic clusters showing the percent cluster membership of each individual in the analysis. The bottom two panels are the individual Q-values from the ADMIXTURE analyses with two (middle) and five (bottom) clusters. The putative dispersers are individuals with Q-values or percent cluster membership greater than 0.75 but assigned to a cluster different from that of most other individuals sampled at the same sampling location and are identified with a white arrow or a white dotted rectangle. Note that the order of individuals varies among panels, but the identity of putative dispersers was the same in the two analyses.

Population assignments

Overall, the results of the assignment tests performed on the baseline samples using all 6,136 markers revealed that there was sufficient power to infer the origin of individuals with high accuracy (lowest estimate including putative dispersers and without imputation: 83.1%; highest estimate excluding putative dispersers and with imputation: 92.7%; Fig. 6). When putative dispersers were included in the dataset, assignment success was lower overall (means over all sampling locations when all markers are included: 83.1% without imputations; 88.6% with imputations; range per sampling location: 66.0-100%; Fig 6a), and many individuals from Halokvik and Lauchlan River were assigned to Ekalluk and Surrey River, and vice-versa (Fig. 6b). As expected, assignment success was improved when the putative dispersers were removed from the dataset (means over all sampling locations when all markers are included: 88.2% without imputations; 92.7% with imputations; range per sampling location 78.5-100%; Fig. 6c), and most of the mis-assignments involved fish caught at Lauchlan River but assigned to Halokvik River, or fish caught at Jayko River assigned to Ekalluk and Surrey River (Fig. 6d).

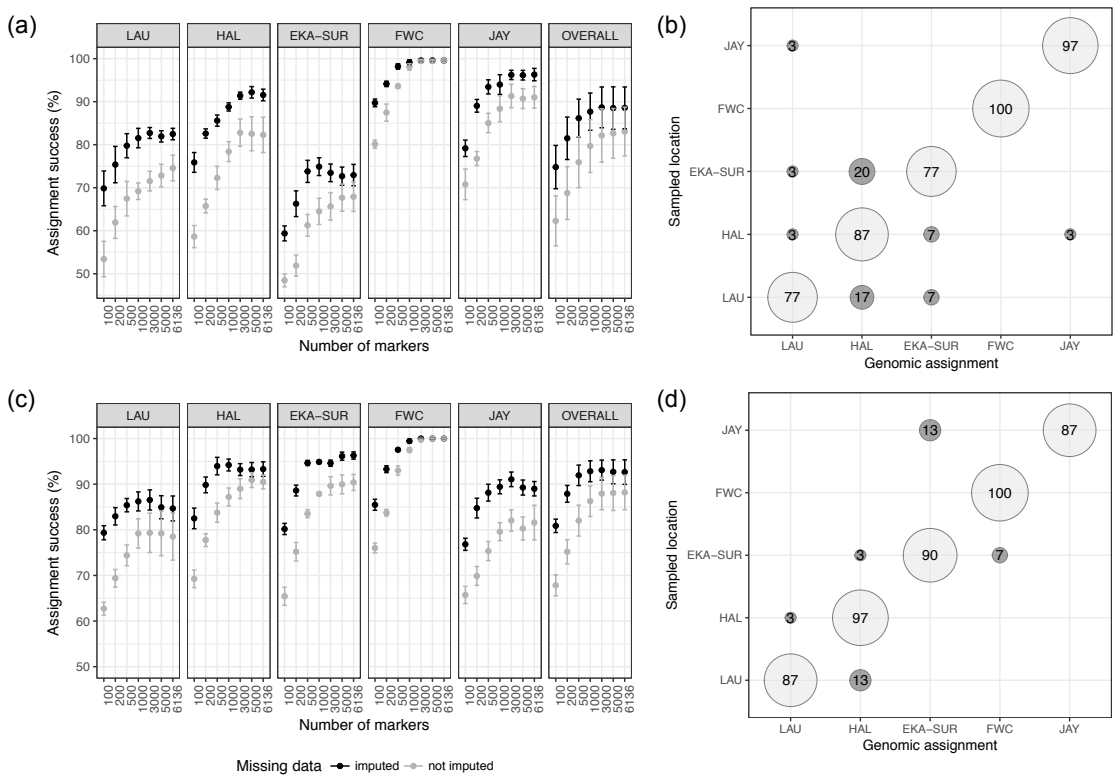

Figure 6. Results of assignment tests with all individuals (a & b) or with putative dispersers removed (c & d). The two left panels (a & c) show how assignment success per sampling location varied according to the number of markers used for the assignment, and on whether missing data were imputed using Random Forest (RF) or not. The two right panels (b & d) show results of the assignments with all markers with missing data imputed. The light grey circles on the diagonal display successful assignments, and the dark grey circles display mis-assignments (i.e., the sample was assigned to a different river from that where it was sampled). Circle size is proportional to the number of individuals in each category.

After confirming that there was sufficient assignment power, we used genomic data using all markers to assign the tagged individuals to their most likely river of origin (Table S1 online supplementary materials). Eleven of the 57 fish (19%) tagged at the Ekalluk River in 2013 and 2014 were assigned to the Halokvik River and two to the Lauchlan River (4%). Only two of the 11 putative Halokvik-origin fish were detected overwintering/spawning in the Halokvik River at least once, and all others were detected only in the Ekalluk system one or more years. In contrast, all but one (29/30) of the fish tagged at the Halokvik River (97%) were assigned to the Halokvik River itself. Six of those 29 individuals were detected overwintering/spawning in both Ekalluk and Halokvik River in different years. Notably, all but one fish (32/33) detected at least one year at Halokvik River were assigned to the Halokvik River. Finally, fish tagged at Surrey River

were assigned to all possible rivers (except Freshwater Creek), but a majority (19/31) were assigned to the grouped Ekalluk-Surrey sample. In short, population assignments also support the prevalence of asymmetric dispersal towards the Ekalluk River.

Demographic simulations

The model selection procedure unambiguously indicated that the isolation with migrations (IM) model was the best with P(IM) ~ 1, while both the panmixia and the strict isolation models had posterior probabilities close to zero (Table S5 in online supplementary materials). This inference was highly significant with a robustness of 1 (Fig. S9 in online supplementary materials). Estimates of demographic parameters from coalescent simulations produced confidence intervals of various widths, and differed most from the posterior in the Ekalluk-Halokvik comparison than in the Ekalluk-Lauchlan comparison. These estimates were mostly informative with regards to estimates of effective populations size, divergence times (except in the Ekalluk-Lauchlan comparisons), and intensity of migration rates, pointing especially to highly asymmetric gene flow (Fig. S10 & S11 in online supplementary materials). This asymmetric gene flow, however, was in the opposite direction of the asymmetric dispersal inferred from both the telemetry and the genetic data. Indeed, gene flow was higher from Ekalluk to Lauchlan, with a ratio of $M_2/M_1 = 3.08$, and from Ekalluk to Halokvik, with a ratio of $M_2/M_1 = 4.06$ (Table 2). Ratios of effective population size (θ_1/θ_A) tended to indicate an expansion of the Ekalluk population relative to the ancestral population in both comparisons (ratios of 9.68 and 3.45). Posterior distributions also indicated an expansion of the Halokvik population ($\theta_2/\theta_A = 2.58$), but the Lauchlan population size was inferred to be smaller than the ancestral population size ($\theta_2/\theta_A = 0.082$), although this should be interpreted cautiously since the estimate of the ancestral population size was not highly accurate. Posterior predictive checks in the Ekalluk-Lauchlan indicated that we were able to accurately reproduce the summary statistics. In the Ekalluk-Halokvik comparison, however, three statistics were significantly different from the observed data, namely the averaged number of shared polymorphic sites and the averaged and standard deviations of the net divergence (Da). In summary, the coalescent simulations and ABC approach provided robust estimates of demographic parameters, and suggested that the asymmetric

dispersal towards Ekalluk River observed with both the telemetry data and the population assignment results did not translate into pronounced gene flow in that direction.

Genome scans and functional annotation of outliers

The three different methods of outlier detection identified a total of 90 markers putatively under selection. The Bayescan analysis identified 30 outlier loci when the prior odds were set at 100 (Fig. S12 online supplementary materials), 10 loci with prior odds of 1000, and 5 loci with prior odds of 5000, all with an FDR of 0.05. The PCAdapt analysis was run assuming two genetic clusters after graphical evaluation of the eigenvalues according to Luu et al. (2017). Consistent with the DAPC results, the first two PC axes explained most variation and only those were retained to account for population structure (Fig. S13 online supplementary materials). Six outliers were identified with FDR = 0.05 (Fig S14 online supplementary materials). Of those six, four were also identified by the Bayescan run with prior odds of 100 and two with the run at prior odds of 5000. For the LFMM analysis, cross-entropy was also minimal at K = 2 (Fig. S15 online supplementary materials), and two latent factors were thus retained to control for population structure. With an FDR = 0.05, 58 markers were identified as outliers by the LFMM analysis, none of which overlapped with the markers identified by Bayescan or PCAdapt.

A total of 1405 of the 6136 markers were positioned onto the Arctic Char high-density linkage map (Nugent et al. 2017) (Fig. 7). Outliers were found on several different linkage groups throughout the genome. Of the 90 outlier markers, 34 provided unambiguous BLAST results against the Atlantic Salmon genome (see Table S6 for full list of successful BLAST results). Of those, 23 were found to be in a gene, while nine were found within 100kb of a gene. Only one of the four outliers that were significant in both the PCAdapt and the Bayescan had unambiguous BLAST results on the Atlantic Salmon genome, but three were located on the same 914Kb scaffold (*scaffold_363*) on the Rainbow Trout genome. The marker with the BLAST result was within the gene *FAM179B*, involved in the structure and function of primary cilia, an organelle usually associated with the reception of extracellular mechanical and chemical stimuli (Das *et al*. 2015). Interestingly, several genes close to outliers were involved in muscle or cardiac muscle function and development (*Myocyte-specific enhancer factor 2C*, *nebulette*,

glycogen phosphorylase muscle form) or in gluconeogenesis (*fructose-1,6-bisphosphatase 1-like*), but a Gene Ontology analysis (*not shown*) failed to identify functions that were significantly enriched in the targets.

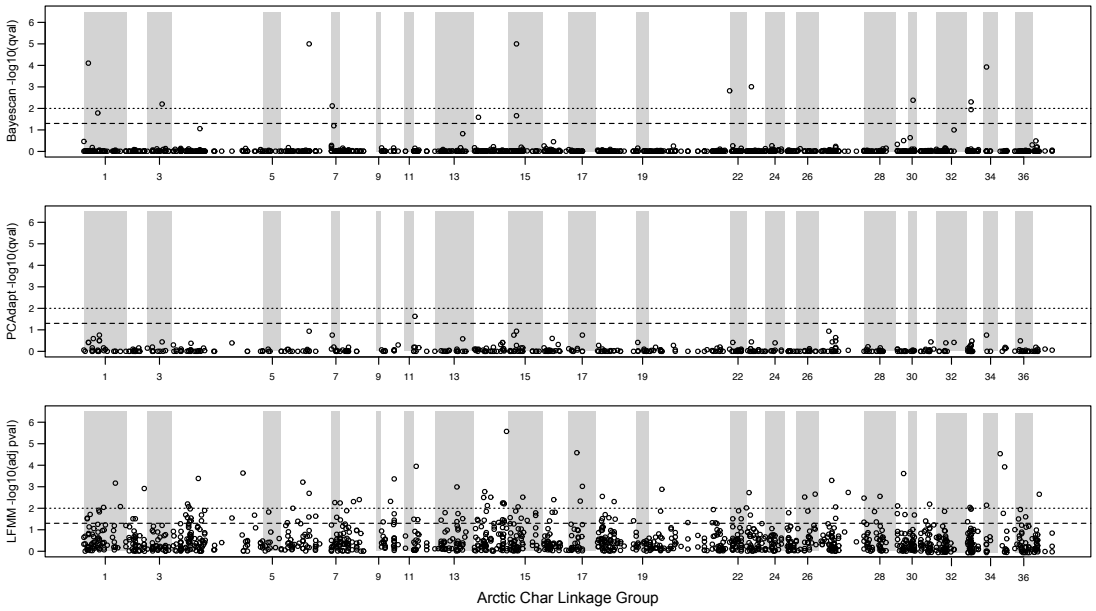

Figure 7. Manhattan plots showing the approximate locations of the 1405 markers successfully positioned on the Arctic Char genetic map along with multiple test corrected test statistics from two genome scan analyses (top: Bayescan and middle: PCAdapt) and a genetic environment association analysis correlating allele frequencies with migration harshness measured as 'work' (bottom: LFMM). Linkage groups are delineated by grey shading and x-axis labels. Dashed lines indicate 0.05 and 0.01 significance thresholds.

Discussion

Concurrent developments in both genomic and telemetry technologies offer new and powerful ways to study the migratory ecology of animals behaving in their natural environments (e.g., Hess *et al*. 2014; Schaffer *et al*. 2016; Franchini *et al*. 2017). The present study illustrates how this integrative approach can be especially fruitful for animals displaying complex migratory behaviours and inhabiting logistically challenging regions. Here telemetry data suggested the presence of asymmetric dispersal between high Arctic populations of anadromous Arctic Char, and population genomic data helped to infer the river of origin of putative dispersers. Dispersal was predominantly from western sampling locations towards the Ekalluk River, the shortest and least harsh river in the Wellington Bay area. Coalescent simulations using an ABC approach also revealed an asymmetry in gene flow, but in the opposite direction, indicating that the observed

asymmetric dispersal does not necessarily lead to gene flow. Instead, the coalescent simulations suggested that Ekalluk River, the most abundant Arctic char population in the region (McGowan 1990), was a source of gene flow to surrounding areas. Together, our observations suggested that Arctic Char home to their natal river to spawn, but may overwinter in rivers with the least harsh migratory route, thus potentially minimizing the costs of migrations in non-breeding years. Another finding facilitated by the integration of telemetry and genomic data was the identification of many outlier markers linked to genes demonstrated to have a role in muscle or cardiac function. Related physiological traits have been demonstrated to evolve in response to migration harshness in other anadromous salmonids (Eliason *et al*. 2011), thus providing a list of candidate genes that could be used as targets of future functional studies of migratory physiology in these Arctic Char populations.

Asymmetric dispersal: overwintering habitat choice minimizes cost of migrations

Unlike most other anadromous salmonids, which return to freshwater exclusively to reproduce (Fleming 1998; Quinn 2005), Arctic Char and other *Salvelinus* species must return yearly to freshwater to overwinter (Armstrong 1974; Johnson 1980). Given that optimal habitats for reproduction and for overwintering likely differ, we might expect individuals to migrate to different locations depending on the purpose of the migration (Dingle 2014). Precise homing to reproduction sites is the norm in salmonids, and it has been suggested that local adaptation to nesting and rearing environments is an important driver of philopatry (Hendry *et al*. 2004). Requirements to home, however, might be relaxed when an individual is returning to freshwater to overwinter instead of to reproduce (i.e., when there are no genetic or evolutionary consequences). Accordingly, there is evidence that Arctic Char (and the closely related Dolly Varden; *Salvelinus malma*) home to their natal habitats to spawn, but can use non-natal habitats to overwinter (Armstrong 1974; Gyselman 1994; Moore *et al*. 2013; Gilbert *et al*. 2016). Here, we documented the use of alternative freshwater habitats by the same individuals in consecutive years using acoustic telemetry data. In addition, we used population genomic data to infer the natal origins of individuals, and confirmed that the asymmetric dispersal observed with the telemetry data was indeed due to fish of western origin (i.e., from

Halokvik and Lauchlan) dispersing to the Ekalluk River. This was the case with the Ekalluk and Surrey baseline samples, which clustering analyses suggested contained a large proportion of western-origin fish. In addition, population assignment of the tagged fish showed that all but one fish detected at least once overwintering/spawning in Halokvik River were from Halokvik River, while many of the fish migrating to the Ekalluk River were from Halokvik River. The inference of asymmetric dispersal, therefore, is strengthened by the combination of independent and complementary sources of evidence from telemetry and genomic data.

The population genomic data could also be used to infer whether the observed dispersal resulted in high levels of gene flow. While dispersal was highly asymmetric towards Ekalluk River, the ABC modeling suggested that it did not necessarily translate into realized gene flow. Indeed, the best supported ABC model was the isolation with migration model with asymmetric gene flow, but the direction of the asymmetry was opposite to that observed with direct dispersal. This asymmetry in gene flow, instead, was from the most abundant population (Ekalluk) towards the least abundant populations (Halokvik and Lauchlan), as suggested by our N_E estimates (both from LDNE and ABC) and from field measures of abundance (McGowan 1990), and may thus simply reflect differences in effective gene flow. In summary, the observed asymmetry in dispersal does not translate into realized gene flow and suggests that dispersers from the Halokvik and Lauchlan rivers do not regularly reproduce in the Ekalluk River system.

Taken together, the observations of asymmetric dispersal with both genomic and telemetry data, but the lack of realized gene flow in the same direction, are consistent with the hypothesis that individuals in breeding condition home to spawn but may select overwintering habitats that minimize the costs of migrations in the years when they are not in breeding condition. Indeed, the Ekalluk River is one of the most easily accessible rivers in the region (measured as work; Table 1). While other rivers are also easily accessible (e.g., Freshwater Creek, Jayko River), they are further away from Wellington Bay, and comparatively fewer fish from the Wellington Bay tagging locations travel there (Moore *et al*. 2016). Given the lack of replication of this environmental axis in the current study, it is impossible to conclude without doubt that migration harshness is the main driver of overwintering habitat selection. Other features of the freshwater habitats

may also explain the observed habitat choice. For example, the Ekalluk River drains the largest lake on Victoria Island, Ferguson Lake, which could offer more abundant or better overwintering habitats, perhaps resulting in lower overwintering mortality. Nonetheless, other studies of Arctic Char have also found spawning habitat accessibility to be a major constraint to migrations (Gyselman 1994; Gilbert *et al.* 2016), and the easier migrations afforded by the short Ekalluk River constitutes a plausible explanation for the observed patterns of dispersal. This hypothesis, however, assumes the ability of Arctic Char to assess which river provides the most readily accessible overwintering habitat. One possibility is that individuals explore a variety of freshwater habitats during their summer migrations and assess their accessibility. Observations of back and forth movement during homing in other salmonids (Quinn 2005), and the regular use of different estuaries by individuals during the summer documented by our telemetry work (Moore *et al.* 2016) suggest this is plausible. Another possibility is that the decision of where to overwinter relies in part on collective navigation (i.e., increased ability to navigate through social interactions), a possibility that has been discussed in relation to salmon homing (Berdahl *et al*. 2014). Given that the Ekalluk River has the most abundant population of Arctic Char in the region (McGowan 1990; Day & Harris 2013), it is also possible that collective navigation biases dispersal towards the river with the greater number of individuals. Continued collection of telemetry data will allow future tests of some of these hypotheses and may help in specifically testing the idea suggested by our observations that migration harshness drives overwintering habitat choice.

In summary, the integration of telemetry and genomic data allowed nuanced insights into the complex migratory ecology of anadromous Arctic Char. Notably, it allowed circumventing logistical challenges posed by the sampling of baseline samples. In most salmonids, sampling in the freshwater is used to ensure the origin of sampled fish for baseline collections, but here the use of different freshwater habitats by adults within their lifetime precludes any certainty in inferring origins of sampled adults even if collected in freshwater. Sampling juveniles prior to their first marine migration (e.g., Moore *et al*. 2013) would also be an option, but attempts at collecting juveniles in 2010 and 2015 failed: seine nets, minnow traps, and electrofishing were unsuccessful, (the latter possibly affected by the low conductivity of Arctic waters), the watersheds under

study are vast and inaccessible, and the location of juvenile rearing habitats are unknown. Given the complex migratory behaviour of Arctic Char and the logistical challenges associated with sampling, the integration of telemetry and genomic data provided a powerful tool for understanding the ecology of this species.

Evidence for local adaptation in migratory traits

Anadromous salmonids display tremendous diversity in migratory traits both among and within species, and their study has contributed to our understanding of migration ecology (Hendry *et al*. 2004; Dingle 2014). The application of molecular tools has provided important insights to the study of the genetic basis of migratory traits in salmonids. For example, O'Malley & Banks (2008) used a candidate gene approach to link variation in two circadian rhythm genes, *OtsClock1a* and *OtsClock1b*, to variation in run timing in anadromous Chinook salmon (*O. tshawytscha*). Other studies have identified SNP markers linked to migration timing in both Chinook salmon (Brieuc *et al.* 2015) and steelhead trout (*O. mykiss*; Hess *et al*. 2016). Finally, Hecht *et al*. (2015) linked variation in RAD markers with 24 environmental correlates using redundancy analysis in Chinook salmon, finding that freshwater migration distance was the environmental variable explaining most of the genomic variation. This last result parallels our own observations suggesting the importance of migratory harshness, and is also corroborated by many classic studies demonstrating the importance of migratory difficulty as a selective factor driving adaptation in morphological, physiological, and life-history traits (Shaffer & Elson 1975; Bernatchez & Dodson 1987; Crossin *et al*. 2004; Eliason *et al*. 2011).

In the present study, we used genome scans (Bayescan and PCAdapt) and genetic-environment correlations (LFMM) with migratory harshness to identify markers putatively under divergent selection. These approaches have been used extensively to identify putative targets of selection, but also have several limitations, which have been amply discussed in the literature (Lotterhos & Whitlock 2014; Haasl & Payseur 2015; Bernatchez 2016; Hoban *et al*. 2016). For example, recent demographic history and patterns of isolation by distance have been shown to lead to high levels of false positives (Lotterhos & Whitlock 2014; Hoban *et al*. 2016). The geographical proximity of the

populations under study, however, precludes important differences in terms of demographic history since the re-colonization of the area from a single source population in the last 6,500 years (Moore *et al*. 2015). In addition, Lotterhos & Whitlock (2015) concluded that maximum power in genome scan studies could be achieved by examining geographically proximate and genetically similar populations that experience contrasting environments, criteria that our system fulfilled. Nevertheless, few of the outlier markers identified by the three methods overlapped, a common finding (de Villemereuil *et al*. 2014; Gagnaire *et al*. 2015).

In accordance with our *a priori* expectation that migration harshness might be a selective factor differing among rivers, several of the outlier SNPs were linked to muscle and cardiac functions and development. One of the outlier SNPs was located within 75 kb (5' side) of *myocyte enhancer factor 2* (MEF2), a transcription factor acting as an important regulator of vertebrate skeletal muscle differentiation and heart development (Potthof & Olson 2007). Supporting the relevance of this protein for migrating salmonids, cardiac mRNA levels of MEF2 were elevated in Atlantic salmon following 10 weeks of experimental exercise training (Castro *et al*. 2013). An outlier was also located within the gene coding for *nebulette*, a cardiac-specific actin-binding protein essential in the structure of the sarcomere Z-disc (Bonzo *et al*. 2008). Another marker was within the gene coding for *muscle glycogen phosphorylase* (PYGM), the enzyme responsible for breaking up glycogen into glucose subunits to power muscle cells (Kitaoka 2014). Finally, another marker was within the gene coding for *fructose-1,6-biphosphatase*, which is involved in the conversion of glycerol into glucose (i.e., gluconeogenesis; Lamont *et al*. 2006). The abundance of putative targets of selection identified linked to muscle and cardiac functions and development is interesting in the context of variation in migration harshness. A Gene Ontology (GO) analysis, however, did not identify any functions that were significantly enriched. Nevertheless, the list of outliers identified provides interesting candidates for future functional studies on the physiology of migrating Arctic char that might help solidify the causal link between the migratory environment and genomic variation. This work will be further aided by the upcoming availability of an Arctic Char reference genome (Macqueen *et al*. 2017).

Conclusions

Here we used an integrative approach to link genome-wide data with telemetry and concluded that extensive dispersal to overwintering habitats is not necessarily associated with high gene flow. These movements were biased towards a short river suggesting that migration harshness might drive overwintering habitat choice in anadromous Arctic Char. Genome scans for outliers and genetic-environment association analysis identified several markers in genes associated with muscle and cardiac functions, further suggesting the importance of migratory harshness in driving local adaptation. Future work in this system will build on these results to describe physiological differences among populations in relation to migration harshness. These will be particularly relevant in the context of adaptation to a changing Arctic, where migratory environments will probably exert an important selective pressure on populations (e.g., Eliason *et al*. 2011). Many of our conclusions would have been difficult, or impossible, to reach with genomic or telemetry data alone, and our study therefore illustrated the synergies made possible by combining the two types of data (Shafer *et al*. 2016). Such integrative approaches will continue to increase our understanding of how migratory behaviour interacts with gene flow to influence the spatial scale at which local adaptation can evolve, and will advance the study of the genetic basis of migratory traits of species behaving in their wild habitats.

Acknowledgements

We are very grateful for the support of the Ekaluktutiak Hunters and Trappers Organization and of the residents of Cambridge Bay, who made this work possible. Koana! We also acknowledge Kitikmeot Food Ltd. for logistical support and their precious collaboration with the plant sampling program. We thank B. Boyle and G. Légaré at the IBIS sequencing platform for their help and advice regarding library preparation and sequencing. A.-L. Ferchaud, L. Benestan, T. Gosselin, S. Bernatchez, A. Perreault-Payette, & C. Perrier shared scripts and provided advice on data analysis and interpretation. The telemetry work was supported by the Ocean Tracking Network (OTN) through a network project grant (NETGP No. 375118-08) from Natural Sciences and Engineering Research Council of Canada (NSERC) with additional support from the

Canadian Foundation for Innovation (CFI, Project No. 13011), by the Polar Continental Shelf Project of Natural Resources Canada (grant #107-15), and by in kind logistical support from the Arctic Research Foundation. J.-S. Moore was supported by fellowships from the Fonds Québécois de Recherche sur la Nature et les Technologies and the W. Garfield Weston Foundation. J. Le Luyer, B. Sutherland and Q. Rougemont were supported from various grants from the Canada Research Chair in Genomics and Conservation of Aquatic Resources led by L. Bernatchez.

Author contribution: J.-S.M. and L.N.H. designed this study and conducted the fieldwork. J-SM performed part of the lab work, most of the analyses, and wrote the manuscript. J.L. identified the SNPs using the STACKS workflow. B.J.S. mapped the markers using MapComp and contributed to the functional annotations. Q.R. performed the ABC analyses. R.F.T., A.T.F., and L.B. provided guidance and supervision at various stages of the study. All authors contributed significantly to the writing of the article.

Data accessibility: The .vcf files, the R code, the sequences of the RAD tags and the results of the genome scans and GEA analyses used in the construction fo the Manhattan plot can be found on Dryad (http://dx.doi.org/10.5061/dryad.f3sm9.2). The raw reads will be deposited to NCBI Short Read Archive. The pipeline used for the ABC simulations will be available at https://github.com/QuentinRougemont/abc_inferences. The pipeline for collecting and preparing the Arctic Char genetic map, running MapComp iteratively, and plotting the results are available at:
https://github.com/bensutherland/salp_anon_to_salp

References

Alexander, D.H., Novembre, J., and Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655–1664.

Aljanabi, S.M., and Martinez, I. (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research* **25**, 4692–4693.

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

Anderson, E.C., Waples, R.S., and Kalinowski, S.T. (2008) An improved method for predicting the accuracy of genetic stock identification. *Canadian Journal of Fisheries and Aquatic Sciences*, **65**, 1475–1486.

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., and Hohenlohe, P.A. (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17,** 81–92.

Armstrong RH (1974) Migration of anadromous Dolly Varden (Salvelinus malma) in southeastern Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, **31**, 435–444.

Benestan, L., J.-S. Moore, B.J. Sutherland, J. LeLuyer, N. Rycroft, J. Attema, L.N. Harris, R.F. Tallman, C. Rougeux, E. Normandeau, L. Bernatchez. (2016) Sex matters in Massive Parallel Sequencing: evidence for biases in genetic parameter estimation and investigation of sex determination systems. bioRxiv https://doi.org/10.1101/096065

Berdahl A, Westley PAH, Levin SA, Couzin ID, Quinn TP (2014) A collective navigation hypothesis for homeward migration in anadromous salmonids. *Fish and Fisheries*, **17**, 525–542.

Bernatchez L, JJ. Dodson (1987) Relationship between bioenergetics and behaviour in anadromous fish migrations. *Canadian Journal of Fisheries and Aquatic Sciences*, **44**, 399-407.

Bernatchez L (2016) On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. *Journal of Fish Biology*, **89**, 2519–2556.

Berthold, P., and Querner, U. (1981) Genetic basis of migratory behavior in European warblers, **212**, 77–79.

Berthold P, Helbig AJ, Mohr G, Querner U (1992) Rapid microevolution of migratory behaviour in a wild bird species. *Nature*, **360**, 668–670.

Bond, M.H., Miller, J.A., and Quinn, T.P. (2015) Beyond dichotomous life histories in partially migrating populations: cessation of anadromy in a long‐lived fish. *Ecology*, **96**, 1899–1910.

Bonzo JR, Norris AA, Esham M, Moncman CL (2008) The nebulette repeat domain is necessary for proper maintenance of tropomyosin with the cardiac sarcomere. *Experimental cell research*, **314**, 3519–3530.

Brieuc MSO, Ono K, Drinan DP, Naish KA (2015) Integration of Random Forest with population‐based outlier analyses provides insight on the genomic basis and evolution of run timing in Chinook salmon (*Oncorhynchus tshawytscha*). *Molecular Ecology*, **24**, 2729–2746.

Castric, V., and Bernatchez, L. (2004) Individual assignment test reveals differential restriction to dispersal between two salmonids despite no increase of genetic differences with distance. *Molecular Ecology,* **13**, 1299–1312.

Castro V, Grisdale-Helland B, Helland SJ *et al.* (2013) Cardiac molecular-acclimation mechanisms in response to swimming-induced exercise in Atlantic salmon. *PLoS ONE*, **8**, e55056.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., and Cresko, W.A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.

Crossin, G.T., Hinch, S.G., Farrell, A.P., Higgs, D.A., Lotto, A.G., Oakes, J.D., and Healey, M.C. (2004) Energetics and morphology of sockeye salmon: effects of upriver migratory distance and elevation. *Journal of Fish Biology,* **65**, 788–810.

Csilléry, K., Blum, M.G.B., Gaggiotti, O.E., and François, O. (2010) Approximate Bayesian Computation (ABC) in practice. Trends in Ecology and Evolution **25**, 410–418.

Das A, Dickinson DJ, Wood CC, Goldstein B, Slep KC (2015) Crescerin uses a TOG domain array to regulate microtubules in the primary cilium. *Molecular biology of the cell*, **26**, 4248–4264.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., and Blaxter, M.L. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics,* **12**, 499–510.

Day, A.C. and Harris, L.N. (2013) Information to support an updated stock status of commercially harvested Arctic Char (*Salvelinus alpinus*) in the Cambridge Bay region of Nunavut, 1960−2009. DFO Canadian Science Advisory Secretariat Research Document 2013/068. v + 30 p.

Delmore, K.E., Hübner, S., Kane, N.C., Schuster, R., Andrew, R.L., Câmara, F., Guigó, R., and Irwin, D.E. (2015) Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Molecular Ecology,* **24**, 1873–1888.

Dempson, J.B., and Kristofferson, A.H. (1987) Spatial and temporal aspects of the ocean migration of anadromous Arctic Char. *In* Common Strategies of Anadromous and Catadromous Fishes, American Fisheries Society Symposium. *Edited by* M.J. Dadswell, R.J. Klauda, C.M. Moffitt, R.L. Saunders, R.A. Rulifson, and J.E. Cooper. pp. 340–357.

de Villemereuil P, Frichot E, Bazin E, François O, Gaggiotti OE (2014) Genome scan methods against more complex models: when and how much should we trust them? *Molecular Ecology*, **23**, 2006–2019.

Dingle, H. (2014) Migration: the biology of life on the move. 2nd edition. Oxford University Press, Oxford UK. 306 pp.

Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J., and Ovenden, J.R. (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources,* **14**, 209–214.

Dutil, J.-D. (1986) Energetic constraints and spawning interval in the anadromous Arctic Charr (*Salvelinus alpinus*). *Copeia*, **4**, 945–955.

Eliason, E.J., Clark, T.D., Hague, M.J., Hanson, L.M., Gallagher, Z.S., Jeffries, K.M., Gale, M.K., Patterson, D.A., Hinch, S.G., and Farrell, A.P. (2011) Differences in thermal tolerance among sockeye salmon populations. *Science,* **332**, 109–112.

Fleming IA (1998) Pattern and variability in the breeding system of Atlantic salmon (Salmo salar), with comparisons to other salmonids. *Canadian Journal of Fisheries and Aquatic Sciences*, **55(Suppl. 1)**, 59–76.

Foll, M., and 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.

Franchini P, Irisarri I, Fudickar A *et al.* (2017) Animal tracking meets migration genomics: transcriptomic analysis of a partially migratory bird species. *Molecular Ecology*, DOI: 10.1111/mec.14108

Frichot, E., Schoville, S.D., Bouchard, G., and François, O. (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution,* **30**, 1687–1699.

Gagnaire P-A, Broquet T, Aurelle D *et al.* (2015) Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary Applications*, **8**, 769–786.

Gilbert, M., Donadt, C.R., Swanson, H.K., and Tierney, K.B. (2016) Low Annual Fidelity and Early Upstream Migration of Anadromous Arctic Char in a Variable Environment. *Transactions of the American Fisheries Society,* **145**, 931–942.

Gosselin, T, Anderson, E. C., Bradbury, I. (2016). assigner: Assignment Analysis with GBS/RAD Data using R. Rpackage version 0.4.1. https://github.com/thierrygosselin/assigner. doi : 10.5281/zenodo.51453

Gross MR, Coleman RM, McDowall RM (1988) Aquatic productivity and the evolution of diadromous fish migration. *Science*, **239**, 1291–1293.

Gyselman EC (1994) Fidelity of anadromous Arctic char (*Salvelinus alpinus*) to Nauyuk Lake, NWT, Canada. *Canadian Journal of Fisheries and Aquatic Sciences,* 51(9): 1927-1934.

Haasl RJ, Payseur BA (2016) Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication. *Molecular Ecology*, **25**, 5–23.

Harris, L.N., Moore, J.S., Bajno, R., and Tallman, R.F. (2016) Genetic Stock Structure of Anadromous Arctic Char in Canada's Central Arctic: Potential Implications for the Management of Canada's Largest Arctic Char Commercial Fishery. *North American Journal of Fisheries Management*, **36**, 1473–1488.

Hecht BC, Matala AP, Hess JE, Narum SR (2015) Environmental adaptation in Chinook salmon (*Oncorhynchus tshawytscha*) throughout their North American range. *Molecular Ecology*, **24**: 5573–5595.

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, Caudill CC, Keefer ML *et al.* (2014) Genes predict long distance migration and large body size in a migratory fish, Pacific lamprey. *Evolutionary Applications*, **7**, 1192–1208.

Hess JE, Zendt JS, Matala AR, Narum SR (2016) Genetic basis of adult migration timing in anadromous steelhead discovered through multivariate association testing. *Proceedings of the Royal Society B: Biological Sciences*, **283**, 20153064.

Hoban S, Kelley JL, Lotterhos KE *et al.* (2016) Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist*, **188**, 379–397.

Hudson, R.R. (2002) Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics,* **18**, 337–338.

Hussey, N.E., Kessel, S.T., Aarestrup, K., Cooke, S.J., Cowley, P.D., Fisk, A.T., Harcourt, R.G., Holland, K.N., Iverson, S.J., Kocik, J.F., Mills Flemming, J.E., and Whoriskey, F.G. (2015) Aquatic animal telemetry: A panoramic window into the underwater world. *Science*, **348**, 1255642–1255642.

Illera JC, Palmero AM, Laiolo P *et al.* (2014) Genetic, morphological, and acoustic evidence reveals lack of diversification in the colonization process in an island bird. *Evolution*, **68**, 2259–2274.

Ishwaran H. and Kogalur U.B. (2015). Random Forests for Survival, Regression and Classification (RF-SRC), R package version 1.6.1.

Jensen, J., and Rikardsen, A.H. (2012) Archival tags reveal that Arctic Charr Salvelinus alpinus and brown trout Salmo trutta can use estuarine and marine waters during winter. *Journal of Fish Biology*, **81**, 735–749.

Johnson, L. (1980) The Arctic Charr. *In* Charrs: salmonid fishes of the genus Salvelinus. *Edited by* E.K. Balon. Dr. W. Junk bv Publishers, The Hague. pp. 15–98.

Jombart, T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart, T., Devillard, S., and Balloux, F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics,* **11**, 94.

Kays, R., Crofoot, M.C., Jetz, W., and Wikelski, M. (2015) Terrestrial animal tracking as an eye on life and planet. *Science* **348**, aaa2478.

Kitaoka Y (2014) McArdle disease and exercise physiology. *Biology*, **3**, 157–166.

Klemetsen, A., Amundsen, P.A., Dempson, J.B., Jonsson, B., Jonsson, N., O'Connell, M.F., and Mortensen, E. (2003) Atlantic salmon *Salmo salar* L., brown trout *Salmo trutta* L. and Arctic Charr *Salvelinus alpinus* L.: a review of aspects of their life histories. *Ecology of Freshwater Fish,* **12**, 1–59.

Kristofferson AH, Berkes F (2015) Adaptive co-management of Arctic char in Nunavut territory. In: *Breaking ice renewable resource and ocean management in Canada's North* (eds Berkes F, Huebert R, Fast H, Manseau M, Diduck A), pp. 249–268. University of Calgary Press, Calgary.

Lamont BJ, Visinoni S, Fam BC *et al.* (2006) Expression of human fructose-1,6-bisphosphatase in the liver of transgenic mice results in increased glycerol gluconeogenesis. *Endocrinology*, **147**, 2764–2772.

Liedvogel, M., Akesson, S., and Bensch, S. (2011) The genetics of migration on the move. *Trends in Ecology & Evolution* **26**, 561–569.

Lien, S., Koop, B.F., Sandve, S.R., Miller, J.R., Kent, M.P., Nome, T., Hvidsten, T.R., Leong, J.S., Minkley, D.R., Zimin, A., Grammes, F., Grove, H., Gjuvsland, A., Walenz, B., Hermansen, R.A., Schalburg, von, K., Rondeau, E.B., Di Genova, A., Samy, J.K.A., Olav Vik, J., Vigeland, M.D., Caler, L., Grimholt, U., Jentoft, S., Inge Våge, D., de Jong, P., Moen, T., Baranski, M., Palti, Y., Smith, D.R., Yorke, J.A., Nederbragt, A.J., Tooming-Klunderud, A., Jakobsen, K.S., Jiang, X., Fan, D., Hu, Y., Liberles, D.A., Vidal, R., Iturra, P., Jones, S.J.M., Jonassen, I., Maass, A., Omholt, S.W., and Davidson, W.S. (2016) The Atlantic salmon genome provides insights into rediploidization. *Nature,* **533**, 200–207.

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

Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, **24**, 1031–1046.

Luu, K., Bazin, E., and Blum, M.G.B. (2017) PCAdapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources,* **17**, 67–77.

Macqueen DJ, Primmer CR, Houston RD *et al.* (2017) Functional Annotation of All Salmonid Genomes (FAASG): an international initiative supporting future salmonid research, conservation and aquaculture. *BMC Genomics*, **18**, 484.

Mascher, M., Wu, S., Amand, P.S., Stein, N., and Poland, J. (2013) Application of genotyping-by-sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley. *PLoS ONE,* **8**, e76925.

Mastretta-Yanes A, Arrigo N, Alvarez N *et al.* (2015) Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, **15**, 28–41.

McGowan DK (1990) Enumeration and Biological Data from the Upstream Migration of Arctic Charr, *Salvelinus alpinus* (L.), in the Cambridge Bay Area, Northwest Territories, 1979-1983. Canadian Data Report of Fisheries and Aquatic Sciences 811: iv + 27 p.

McIsaac, D. O., and Quinn, T. P. (1988) Evidence for a hereditary component in homing behavior of chinook salmon (*Oncorhynchus tshawytscha*). Can. J. Fish. Aquat. Sci. **45**: 2201-2205.

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

Moore, J.-S., Harris, L.N., Tallman, R.F., and Taylor, E.B. (2013) The interplay between dispersal and gene flow in anadromous Arctic Char (*Salvelinus alpinus*): implications for potential for local adaptation. *Canadian Journal of Fisheries and Aquatic Sciences*, **70**, 1327–1338.

Moore JS, Bajno R, Reist JD, Taylor EB (2015) Post‐glacial recolonization of the North American Arctic by Arctic Char (*Salvelinus alpinus*): genetic evidence of multiple northern refugia and hybridization between glacial lineages. *Journal of Biogeography* **42**, 2089–2100.

Moore, J.S., Harris, L.N., Kessel, S.T., Bernatchez, L., Tallman, R.F., and Fisk, A.T. (2016) Preference for near-shore and estuarine habitats in anadromous Arctic Char (*Salvelinus alpinus*) from the Canadian high Arctic (Victoria Island, NU) revealed by acoustic telemetry. *Canadian Journal of Fisheries and Aquatic Sciences,* **73**, 1434–1445.

Narum SR, Gallardo P, Correa C, Matala A, Hasselman D, Sutherland BJG and Bernatchez L. (2017) Genomic patterns of diversity and divergence of two introduced salmonid species in Patagonia, South America. *Evolutionary Applications*, **10**, 402–416.

Nugent C. M., Easton A. A., Norman J. D., Ferguson M. M., Danzmann R. G., 2017 A SNP based linkage map of the Arctic Charr (*Salvelinus alpinus*) genome provides insights into the diploidization process after whole genome duplication. *G3 - Genes|Genomes|Genetics*, **7**, 543-556.

O'Malley KG, Banks MA (2008) A latitudinal cline in the Chinook salmon (*Onchorhynchus tshawytscha*) Clock gene: evidence for selection on PolyQ length variants. *Proc. R. Soc. Lond. B*, **275**, 2813–2821.

Potthoff MJ, Olson EN (2007) MEF2: a central regulator of diverse developmental programs. *Development*, **134**, 4131–4140.

Quinn, T. (2005) The Behavior and Ecology of Pacific Salmon & Trout. University of Washington Press.

Reist, J.D., Power, M., and Dempson, J.B. (2013) Arctic Charr (*Salvelinus alpinus*): a case study of the importance of understanding biodiversity and taxonomic issues in northern fishes. *Biodiversity,* **14**, 45–56.

Roff DA (2002) *Life History Evolution*. Sinauer Associates Inc., Sunderland, MA, USA.

Roff, D.A., and Fairbairn, D.J. (1991) Wing dimorphisms and the evolution of migratory polymorphisms among the Insecta. *American Zoologist,* **31**, 243–251.

Ross-Ibarra J, Wright SI, Foxe JP *et al.* (2008) Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. *PLoS ONE*, **3**, e2411.

Roux C, Tsagkogeorga G, Bierne N, Galtier N (2013) Crossing the species barrier: genomic hotspots of introgression between two highly divergent *Ciona intestinalis* species. *Molecular Biology and Evolution*, **30**, 1574–1587.

Sakamoto T., Danzmann R. G., Gharbi K., Howard P., Ozaki A., Khoo S. K., Woram R. A., Okamoto N., Ferguson M. M., Holm L. E., Guyomard R., Hoyheim B., 2000 A microsatellite linkage map of Rainbow Trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genetics* **155,** 1331–1345.

Schaffer, W.M., and Elson, P.F. (1975) The adaptive significance of variations in life history among local populations of Atlantic salmon in North America. *Ecology*, **56**, 577–590.

Shafer, A.B.A., Northrup, J.M., White, K.S., Boyce, M.S., Côté, S.D., and Coltman, D.W. (2012) Habitat selection predicts genetic relatedness in an alpine ungulate. *Ecology,* **103**, 1317–1329.

Shafer, A.B.A., Northrup, J.M., Wikelski, M., Wittemyer, G., and Wolf, J.B.W. (2016) Forecasting Ecological Genomics: High-Tech Animal Instrumentation Meets High-Throughput Sequencing. *PLoS Biology,* **14**, e1002350.

Sousa, V., and Hey, J. (2013) Understanding the origin of species with genome-scale data: modelling gene flow. *Nature Reviews Genetics*, **14**, 404–414.

Sutherland B, Gosselin T (2016) Salmonid chromosome evolution as revealed by a novel method for comparing RADseq linkage maps. *Genome Biology and Evolution*, **12**, 3600-3617.

Turgeon, J., Duchesne, P., Colbeck, G.J., Postma, L.D., and Hammill, M.O. (2012) Spatiotemporal segregation among summer stocks of beluga (*Delphinapterus leucas*) despite nuclear gene flow: implication for the endangered belugas in eastern Hudson Bay (Canada). *Conservation genetics,* **13**, 419–433.

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

Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ (2016) GMAP and GSNAP for Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and Functionality. In: *Statistical Genomics* Methods in Molecular Biology. pp. 283–334. Springer New York, New York, NY.

Table 1. Sample information, river characteristics, and basic genetic diversity estimates for the six sampling locations used in the present study. Sample type refers to whether the individuals were collected specifically for DNA analysis ('Baseline') or were individuals that were equipped with an acoustic tag as part of the telemetry study ('Tagged'). The sampling period refers to whether sampling targeted downstream migrating fish ('Spring') or upstream migrating fish ('Fall'). N_{tot} is the total number of individuals sampled, $N_{<0.25\ NAs}$ the number of individuals with less than 25% missing data retained for all analyses, H_O the observed heterozygosity, H_S the expected heterozygosity, and G_{IS} is the inbreeding coefficient. All G_{IS} values are significant (<0.001).

Sampling location	Coordinates	River length (km)	Elevation gain (m)	Sampling year	Sample type	Sampling period	N_{tot} / $N_{<0.25\ NAs}$	H_O	H_S	G_{IS}	N_e (95% CI)
Lauchlan River	N68.94° W108.52°	28.5	64	2012	Baseline	Fall	57/57	0.102	0.105	0.03	335 (318 – 353)
Halokvik River	N69.16° W107.09°	50.4	118	2013	Baseline	Fall	60/59	0.099	0.103	0.037	823 (736 - 934)
				2013	Tagged	Fall	34/30				-
Surrey River	N69.45° W106.68°	18	35	2013	Baseline	Spring	60/59	0.102	0.105	0.023	892 (798 – 1009)
				2014	Tagged	Spring	31/31				-
Ekalluk River	N69.41° W106.31°	3.8	15	2013	Baseline	Fall	60/53	0.104	0.106	0.017	1081 (935 – 1280)
				2013	Tagged	Spring	30/26				-
				2014	Tagged	Spring	31/31				-
Kitiga River	N69.29° W106.24°	17.3	38	NA	NA	NA	NA	NA	NA	NA	NA
Freshwater Creek	N69.12° W105.00°	3.6	7	1992	Baseline	Unknown	32/32	0.105	0.105	0.001	389 (351 – 436)
Jayko River	N69.72° W103.28°	5.4	3	2014	Baseline	Fall	60/58	0.104	0.107	0.022	368 (351 – 387)

Table 2. Parameter estimates for the isolation with migration (IM) model which was determined to have the highest posterior probability among the three models tested (P(IM)≈ 1) for two independent runs comparing Ekalluk and Halovik rivers (EKA-HAL) and Ekalluk and Lauchlan rivers (EKA-LAU). The parameters are θ_A/θ_{ref} for the ancestral population; θ_1/θ_{ref} and θ_2/θ_{ref} for the two daughter populations (θ_1 = Ekalluk in both cases and θ_2 = Halokvik and Lauchlan depending on the analysis), with $\theta_{ref} = 4N_{ref}\mu$ where N_{ref} is the effective size of a reference population and μ is the mutation rate per generation, M_1 is the migration rate into population 1 (Ekalluk) from population 2 (Halokvik and Lanchlan) and vice-versa for M_2, and $\tau = T_{split}/4N_{ref}$ where T_{split} is the time (in generations) since population 1 and 2 diverged from the ancestral population.

Parameter	Prior	EKA-HAL Median [95% Credible Intervals]	EKA-LAU Median [95% Credible Intervals]
θ_A/θ_{ref}	Uniform[0-20]	1.21 [0.20 – 2.38]	5.75 [1.40 – 12.83]
θ_1/θ_{ref}	Uniform[0-20]	11.71 [7.77 – 18.05]	19.86 [19.65 – 19.96]
θ_2/θ_{ref}	Uniform[0-20]	3.10 [1.91 – 6.05]	0.47 [0.18 – 2.75]
M_1	Uniform[0-40]	7.91 [3.93 – 16.28]	12.31 [6.18 -28.39]
M_2	Uniform[0-40]	32.15 [28.15 – 36.32]	37.59 [32.03 – 39.82]
τ	Uniform[0-30]	3.80 [2.30 – 5.30]	11.24 [5.96 – 25.65]