Migration harshness drives habitat choice and local adaptation in anadromous Arctic Char: evidence from integrating population genomics and acoustic telemetry

Jean-Sébastien Moore[1], Les N. Harris[2], Jérémy Le Luyer[1], Ben J. G. Sutherland[1], Quentin Rougemont[1], Ross F. Tallman[2], Aaron T. Fisk[3] & 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] Great Lakes Institute of Environmental Research, University of Windsor, Windsor, Ontario, Canada

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

Running head: Migratory 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 local adaptation in anadromous Arctic Char from six rivers in the Canadian Arctic. Telemetry data from 124 tracked individuals indicated asymmetric dispersal, with a large proportion (72%) of fish tagged in three rivers migrating up the shortest river in the fall. Population genomics data from 6,136 SNP markers revealed weak, albeit significant, population differentiation (F_{ST} = 0.011) and population assignments confirmed the asymmetric dispersal revealed by telemetry data. Approximate Bayesian Computation simulations suggested the presence of asymmetric gene flow but in the opposite direction than 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 least harsh migratory route to minimize the costs of migration in non-breeding years. Genome scans and genetic-environment-associations identified 90 markers putatively associated with local adaptation, 23 of which were in or near a gene. Of those, at least four were involved in muscle and cardiac function, further highlighting the potential importance of migratory harshness as a selective pressure. Our study illustrates the power of integrating genomic 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, however, migrations have been a challenging topic of study. Quantifying migratory phenotypes can be 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; 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 diversity. 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 facultative anadromous, iteroparous 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 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; and (3) fall overwintering migrations to freshwater overwintering sites 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, but 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.

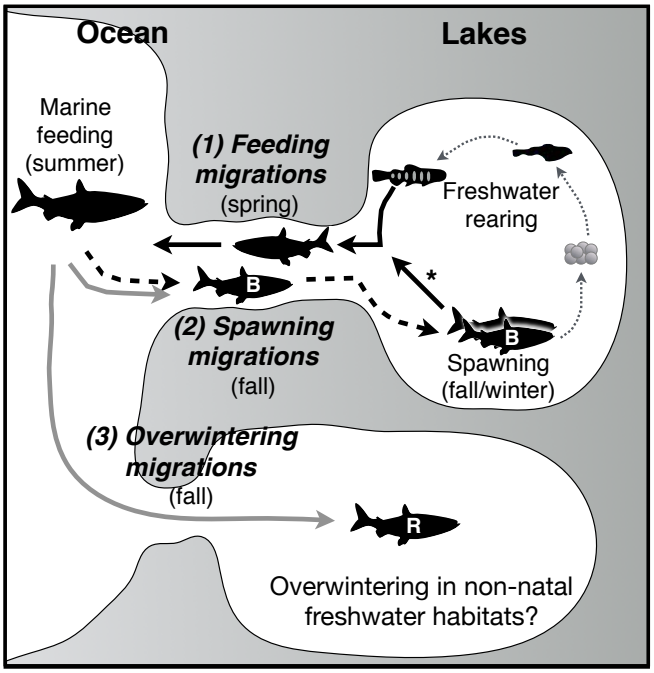


Figure 1. The anadromous Arctic Char life cycle highlighting three distinct migrations. (1) Feeding migrations (black arrows): after hatching and rearing in freshwater for several years (grey dotted arrows), individuals smoltify and begin to migrate to saltwater to feed during the summer when the rivers 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 and harshness. Indeed, 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 new migratory route in the blackcap (*Sylvia atricapilla*), a migratory warbler from continental Europe, in response to changing environmental conditions (Berthold *et al*. 1992). Prior to the 1970s, this bird migrated to Africa to overwinter, but milder conditions and the presence of bird feeders led to an increase in the frequency of individuals overwintering in Britain, thus benefiting from much shorter migrations (Berthold *et al*. 1992). Freshwater migration lengths and

131 elevation gain (i.e. their 'harshness') are also well known drivers of the bioenergetics
132 costs of migrations in anadromous fishes, including salmonids (e.g., Bernatchez &
133 Dodson 1987), and can drive local adaptation of their life history (e.g., Schaffer & Elson
134 1975), morphological (e.g., Crossin et al. 2004), and physiological traits (e.g., Eliason *et*
135 *al*. 2011). The extent to which migration harshness drives habitat choice (particularly
136 during overwintering migrations) and local adaptation in anadromous Arctic Char,
137 however, remains largely unknown (but see Gilbert *et al*. 2016).
138 In order to investigate the possible role of migration harshness in driving habitat
139 choice and local adaptation, we combined genotyping-by-sequencing (GBS) and acoustic
140 telemetry data to study the migrations of anadromous Arctic Char from southern Victoria
141 Island in the Canadian Arctic Archipelago (Fig. 2). The largest commercial fishery for
142 Arctic Char in Canada has been operating in this region for over five decades (Day &
143 Harris 2013) and the local Inuit people rely on this resource for economic, subsistence,
144 and cultural purposes (Kristofferson & Berkes 2005). At least seven rivers in this region
145 support populations of Arctic Char, five of which are commercially fished. Previous
146 studies in the region have demonstrated that many of the stocks mix at sea (Dempson &
147 Kristofferson 1987; Moore *et al.* 2016), and that genetic differentiation estimated from
148 microsatellite markers is low (Harris *et al*. 2016). Here, we aimed to better understand
149 patterns of homing and dispersal of Arctic Char in the region. Particularly, we are
150 interested in how variation in the harshness of freshwater migrations observed among
151 rivers in the region (Fig. 2; Table 1) influences freshwater habitat choice and local
152 adaptation. We addressed four empirical objectives related to this question. First, we used
153 acoustic telemetry data to show that many individuals used different freshwater habitats
154 in different years, and fish from all tagging locations preferentially used the shortest river
155 in the Wellington Bay region (Ekalluk River). Second, we used population genomic data
156 to document population structure and infer the natal origins of migrating individuals, thus
157 confirming that dispersal to non-natal habitats occur. Third, we used an Approximate
158 Bayesian Computation (ABC) framework (Csilléry *et al.* 2010) to determine whether the
159 asymmetric dispersal towards Ekalluk River is associated with asymmetric gene flow.
160 Finally, we used genome scans to test for evidence of local adaptation among
161 populations, and we characterized the chromosomes containing these markers and the

162 proximity of the markers to known genes in a related salmonid genome. This analysis
163 provided new insights on patterns and consequences of migrations in this ecologically,
164 economically, and culturally important species.
165
166 **Materials and Methods**
167 *Telemetry data and sampling*
168 Detailed methods for the acoustic telemetry data collection can be found in Moore
169 *et al*. (2016). In short, 42 moored Vemco VR2W passive acoustic receivers were
170 deployed between July 13 and September 8, 2013 (Fig. 2) along the southern shore of
171 Victoria Island, Nunavut, near the community of Cambridge Bay. Detections were
172 continually recorded until receivers were last retrieved between July 27 and August 31
173 2016. There are at least seven watersheds in the region that support runs of anadromous
174 Arctic Char, and these watersheds are drained by rivers that vary in length and elevation
175 gain (Fig. 2 & Table 1). We placed a receiver in six of these seven rivers to detect
176 migrating fish (the Jayko River was excluded from the telemetry study because of the
177 large distance between this site and the others). A total of 124 Arctic Char were
178 surgically implanted with Vemco V16 acoustic transmitters at three tagging locations
179 (Table 1). In all cases, the fish were collected at the river mouth near the ocean, and for
180 the Ekalluk and Surrey River, tagging occurred during the downstream migration (early
181 July), whereas tagging at the Halokvik River targeted upstream migrants (mid/late-
182 August). Because tagging occurred well before spawning, no information is available
183 about the breeding status of migrating individuals. In the absence of this information, we
184 refer to the use of non-natal freshwater systems as 'dispersal' throughout the manuscript,
185 while we use the term 'homing' to refer to the use of natal habitat regardless of the
186 purpose of the migration. We also refer to all migrations back to freshwater in the fall as
187 'return migrations' regardless of whether an individual dispersed or homed. A fin biopsy
188 was taken from each tagged fish for genomic analysis. Additional baseline samples for
189 genetic analysis were taken from sampling conducted during the upstream migrations to
190 monitor commercial stocks or were collected from commercial fishery catches (Table 1).
191 Samples from the Kitiga River (N69.29° W106.24°), which harbors an Arctic Char
192 population, were also obtained, but DNA extractions for these samples failed.

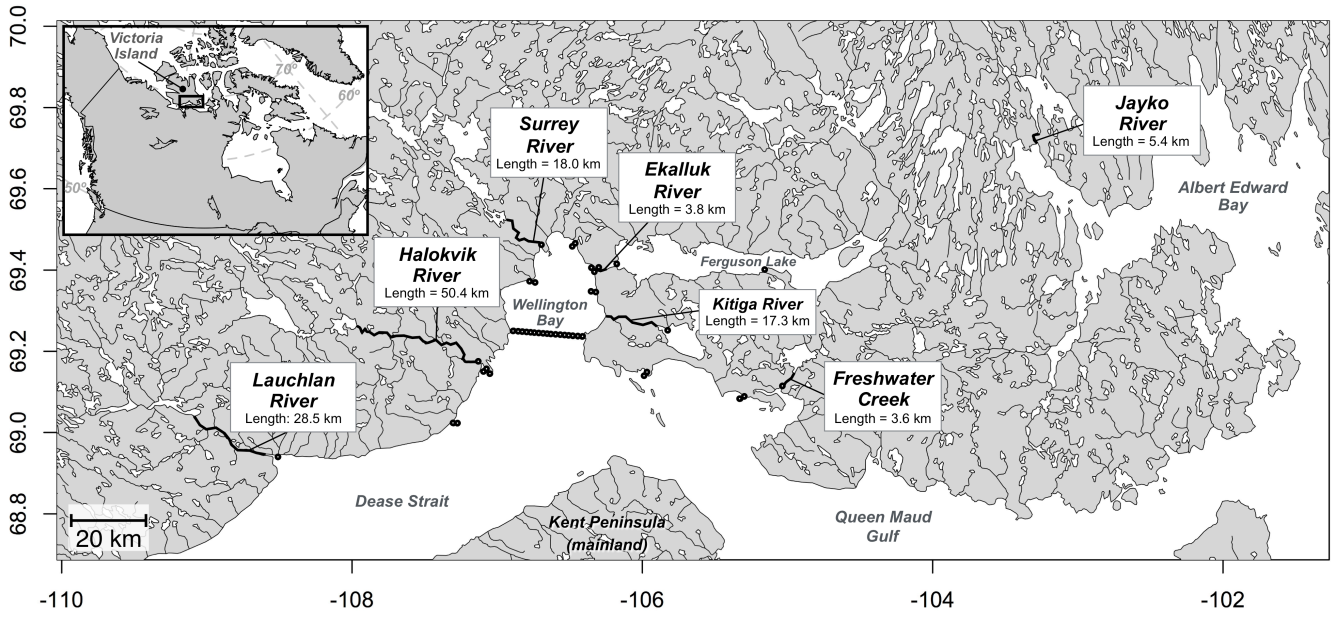


Figure 2. Map of the study area highlighting the names and lengths of the rivers sampled. The black circles with a white center represent the locations of the moored acoustic receivers used to track the movements of Arctic Char surgically implanted with transmitters at the Ekalluk, Surrey and Halokvik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

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 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

212 was also run on a different multiplex to evaluate sequencing errors (Mastreta-Yanes *et al.*
213 2015). Libraries were size-selected using a BluePippin prep (Sage Science), amplified by
214 PCR and sequenced on the Ion Torrent Proton P1v2 chip. The obtained reads were first
215 aligned to the closely related Rainbow Trout (*Oncorhynchus mykiss*) genome (Berthelot
216 *et al.* 2014) using GSNAP v2016-06-09 (Wu *et al.* 2016). The detailed methods for SNP
217 identification, SNP filtering and genotyping using STACKS v.1.40 (Catchen *et al.* 2013)
218 are presented in online supplementary materials.
219
220 *Genomic data: basic statistics, population structure, and identification of putative*
221 *dispersers*
222 Observed (H_O) and expected heterozygosity (H_E), and the inbreeding coefficient (G_{IS})
223 were estimated using GenoDive v2.0b27 (Meirmans and Van Tienderen 2004). Effective
224 population size (N_E) was estimated for the baseline samples using the linkage
225 disequilibrium method in NeEstimator V2.01 (Do *et al.* 2013) with a critical value for
226 rare alleles of 0.05. Pairwise population differentiation was quantified using F_{ST} (Weir &
227 Cockerman 1984) and their significance values estimated with 1000 permutations also in
228 GenoDive. A Discriminant Analysis of Principal Components (DAPC; Jombart *et al*.
229 2010) in the R package adegenet (Jombart 2008) was used to describe population
230 structure. First, the 'find.clusters' function was used with the number of clusters *K*
231 varying from 1 to 12. PCA scores of individuals were then plotted and the 'compoplot'
232 function was used to calculate their proportion of membership to the genetic clusters
233 identified. ADMIXTURE (Alexander *et al.* 2009) was also run varying the number of
234 clusters *K* between 1 and 12. Putative dispersers were identified on the basis of cluster
235 membership probabilities in DAPC and ADMIXTURE. Putative dispersers were defined
236 as having a >50% probability of membership to a different cluster than the majority
237 (>50%) of other sampled individuals at that location. These putative dispersers, as well as
238 individuals that had less than 75% probability of membership to any genetic clusters (i.e.
239 putatively admixed individuals) were removed from some analyses (when noted) to avoid
240 biases. Note that both of these thresholds are purposefully low to ensure all putative
241 dispersers are removed.
242

243 *Population assignments*

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

265

266 *Demographic simulations*

267 To better understand the relative contributions of recent divergence and current gene flow
268 on observed population structure, and to verify whether the observed patterns of
269 dispersion translated into realized gene flow, an ABC pipeline using coalescent
270 simulations was used (Csilléry et al. 2010). Following other approaches (e.g., Illera *et al*.
271 2014), three demographic models were compared and demographic parameters for the
272 best model were estimated. The analyses focused on comparing the Ekalluk and Lauchlan
273 samples and then the Ekalluk and Halokvik samples to limit computational demands.

274 Furthermore, these comparisons are the most relevant to test our hypotheses as they are
275 representative of the eastern (Ekalluk) and western (Lauchlan and Halokvik) group of
276 populations identified as the first level of hierarchical structure (see Results). We tested
277 three models: (1) a null hypothesis of panmixia; (2) a model of population divergence
278 with gene flow (i.e. the isolation with migration model, IM); and (3) a model of
279 population divergence without gene flow (strict isolation, SI). The simplest model of
280 panmixia was controlled by a single parameter, $\theta = 4\ N_1\mu$, where N_1 is the effective size
281 of the panmictic population and μ is the mutation rate per generation. The SI model is
282 characterized by six demographic parameters, namely $\theta_1=4N_1\mu$, $\theta_2=4N_2\mu$, and $\theta_A=4N_A\mu$,
283 with N_1 and N_2 the effective population size of the two daughter populations and N_A the
284 effective size of the ancestral population respectively. All these parameters are scaled by
285 $\theta_1=4N_{ref}\mu$ where N_{ref} is the effective size of a reference population. The two daughter
286 populations diverged from the ancestral T_{split} generations ago with $\tau = T_{split}/4N_{ref}$. The IM
287 model was further characterized by the migration rates $M_1 = 4Nm$ (migration into
288 population 1 from population 2) and $M_2 = 4Nm$ (migration into population 2 from
289 population 1) sampled independently and where Nm is the effective migration rate. Large
290 prior distributions following those commonly applied were used and drawn from uniform
291 distributions (e.g., Sousa & Hey 2013). Coalescent simulations ($n = 10^6$) were performed
292 under each model using msnam (Ross-Ibarra *et al*. 2008), which is a modified version of
293 the widely used ms simulator (Hudson 2002), under the infinite site model of DNA
294 mutation. The pipeline of Roux *et al*. (2013) was used with modifications to test for
295 panmixia, and priors were computed with a Python version of priorgen (Ross-Ibarra *et al*.
296 2008). Details of the summary statistics used, the model selection procedure, analysis of
297 robustness, parameter estimations and posterior predictive checks can be found in the
298 online supplementary materials. Values were not transformed into biological units since
299 the mutation rate (μ) was unknown. Instead coalescent units were kept, and ratios of θ
300 and M_1/M_2 are interpreted only.
301
302 *Genome scans and functional annotation of outliers*
303 Genomic studies of local adaptation using different methods to identify outlier loci often
304 only detect partially overlapping sets of loci (Gagnaire *et al*. 2015). A common practice

305 to partially circumvent these problems is to combine genome scans with genetic-
306 environment-association (GEA) methods as a way to more reliably identify the most
307 likely targets of selection (De Villemereuil *et al.* 2014). Therefore, we identified markers
308 putatively under selection using two genome scan methods and a GEA method. First,
309 Bayescan v1.2 (Foll & Gaggiotti 2008) was used to detect outlier loci with elevated F_{ST}
310 among the baseline populations (Table 1) with 5,000 iterations, and a burn-in length of
311 100,000. To increase our chances of finding outliers while also limiting type-two errors,
312 we tested values of 100, 1,000 and 5,000 for the prior odds (Lotherhos & Whitlock 2014).
313 This analysis was run on the dataset from which putative dispersers were removed to
314 avoid biases (n = 273 after dispersers removed). Second, we used PCAdapt (Luu *et al*.
315 2017), which uses PCA to control for population structure without *a priori* population
316 definition, a particularly useful feature in the current study where we expect significant
317 admixture among populations. Because of this feature, we ran the analysis on the full
318 dataset including the putative dispersers, which also increased the analytical power
319 associated with increased sample size (n = 318). Simulations have also shown that
320 PCAdapt is less prone to type-2 errors than Bayescan, and to be less affected by the
321 presence of admixed individuals in the samples (Luu *et al*. 2017). The optimal number of
322 principal components necessary to describe population structure was determined using
323 the graphical method described in Luu *et al.* (2017) and varying *K* from 1 to 10. Finally,
324 we also used latent factor mixed models (LFMM; Frichot *et al*. 2013) implemented in the
325 R package LEA to test directly for correlations between allele frequencies at specific loci
326 and migratory difficulty. We also varied K from 1 to 10 to identify the optimal number of
327 latent factors required to describe population structure. Migration harshness was
328 measured as 'work' (Crossin *et al*. 2004), that is the product of river length and altitude
329 gain, standardized to a mean of zero and a standard deviation of one. Because the
330 environmental data had to be assigned to a sampling location, we used the dataset with
331 the putative dispersers removed for this analysis (n = 273). As recommended by the
332 authors in both user manuals, for PCAdapt and LFMM, the missing data were imputed as
333 before using package *RandomForestSRC* (100 random trees and 10 iterations). All three
334 approaches controlled for false discovery rates (FDR) with an α of 0.05.

335 Given the absence of a reference genome for Arctic Char, we used the recently
336 developed high-density linkage map for the closely related Brook Char (*Salvelinus*
337 *fontinalis*; Sutherland *et al.* 2016) to determine approximate genomic positions of the
338 markers using the MapComp method (Sutherland *et al.* 2016). In brief, MapComp aligns
339 the flanking sequence of GBS markers of two genetic maps (here the Brook Char genetic
340 map and the anonymous Arctic Char markers) to a reference genome of a closely related
341 species. Here the Atlantic Salmon (*Salmo salar*) genome (Lien *et al.* 2016) was used
342 because it has longer scaffolds than the current Rainbow Trout assembly and thus
343 allowing more markers to be paired (see Sutherland *et al.* 2016). Then, markers from
344 each species that are closest to each other in nucleotide position on the reference genome
345 within a set distance are paired. The pairing was conducted in 10 iterations, removing
346 paired anonymous Arctic Char markers each iteration and rerunning the pairing with the
347 Brook Char genetic map to permit pairing of more than one anonymous marker with a
348 single genetic map marker, as previously described (Narum *et al.* 2017). Here, the
349 distance cutoff between the two markers on the reference genome was within 1 Mbp
350 maximum distance to find approximate locations for the greatest number of markers. All
351 significant outliers were also used in a BLAST query against the Atlantic Salmon
352 genome, which has a full annotation available in NCBI (Lien et al. 2016; NCBI Genome
353 ICSASG_v2 reference Annotation Release 100; GCA_000233375.4). Outlier markers
354 that were found by more than one method, but did not BLAST unambiguously to the
355 Atlantic Salmon genome, were also checked against the Rainbow Trout (*Oncorhynchus*
356 *mykiss*) genome.
357
358 **Results**
359 *Telemetry data*
360 The receivers located at the mouth of each river allowed the inference of likely
361 freshwater overwintering/spawning sites of 98 acoustically tagged Arctic Char for up to
362 three consecutive years (Table S1 in online supplementary materials). The telemetry data
363 suggested a pattern of asymmetric dispersal. Indeed, many fish from all three tagging
364 locations were detected one or more years in the Ekalluk River (Fig. 3), the shortest and
365 least harsh river around Wellington Bay (Fig. 1 and Table 1), and one of the shortest in

the entire area. First, of the 47 fish tagged in both 2013 and 2014 at the Ekalluk River, 44 returned to this river each year (94%). Second, 16 of the 21 fish (76%) tagged at the Surrey River returned to the Ekalluk River each year they were detected, while only one fish returned to the Surrey River. We therefore hypothesize 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 this river at least once (93%), but six (20%) returned to the Ekalluk River at least once. 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, we detected a majority of fish from all three tagging locations (72%) migrating to the Ekalluk River at least once.

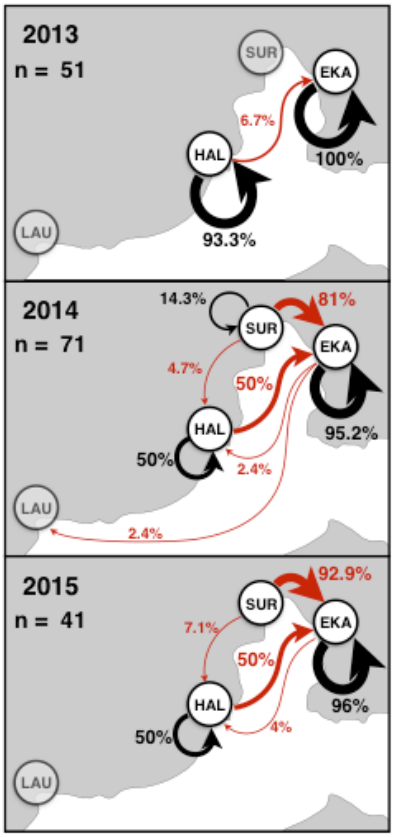


Figure 3. Choice of overwintering/spawning habitat by Arctic Char from three tagging locations inferred from acoustic telemetry data collected over three summers (2013, 2014, and 2015). The proportion of individuals returning to the same location where they were tagged (black) or dispersing to a different 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.

384 *Sequencing and SNP calling*

385 After cleaning and demultiplexing, a total of 1.9 billion reads were left with an average of
386 4.0 million reads per individual (coefficient of variation: 30.3%). The assembly resulted
387 in a catalog containing 2,963,980 loci and a total of 13,568,653 SNPs (over 1,272,427
388 polymorphic loci) after the population module of STACKS. Individuals with more than
389 25% missing genotypes were removed from all analyses (Table 1), and 6,230 high quality
390 SNPs were retained after filters. In addition, 94 putatively sex-linked markers were
391 removed (see Fig. S4 & S5 in supplementary materials and Benestan et al. 2017 for
392 details) leaving a total of 6,136 SNPs used in all subsequent analyses (Table 2). Replicate
393 individuals (n = 11) sequenced twice had identical genotypes at 92.0-97.1% of the
394 markers (mean 94.6%), which was comparable to error rates reported in Mastreta-Yanes
395 et al. (2015).

396

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

398 Levels of heterozygosity were comparable among populations (H_O: 0.099-0.105; H_E:
399 0.103-0.107), and although all G_{IS} values were significantly positive, indicating an excess
400 of homozygotes, the values were all small (0.001-0.037) (Table 1). Effective population
401 size estimates varied between 335 (Lauchlan River; 95% c.i.: 318-353) and 1081
402 (Ekalluk River; 95% c.i.: 935-1280). A weir enumeration study conducted in 1979-1983
403 found the Ekalluk River to be the most abundant population in the region, and the
404 Lauchlan River the least abundant, but the estimates of census size were much larger than
405 the N_E estimates reported here (183,203 for Ekalluk River and 10,850 for Lauchlan;
406 McGowan 1990). Population differentiation among rivers was weak albeit significant,
407 with an overall F_{ST} of 0.011 (95% c.i. 0.010-0.012; Fig 4a). The F_{ST} value between
408 Ekalluk and Surrey rivers was substantially lower than the other values (F_{ST} = 0.001;
409 95% c.i. 0.0006-0.0016), which was in agreement with our telemetry observations that
410 most fish captured at Surrey migrated to the Ekalluk River in the fall. It is thus likely that
411 sampling at the Surrey River resulted in the interception of Ekalluk River fish migrating
412 through the area, and those two sampling locations were combined for subsequent
413 analyses unless otherwise noted. When the putative dispersers were removed (*see below*),
414 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.

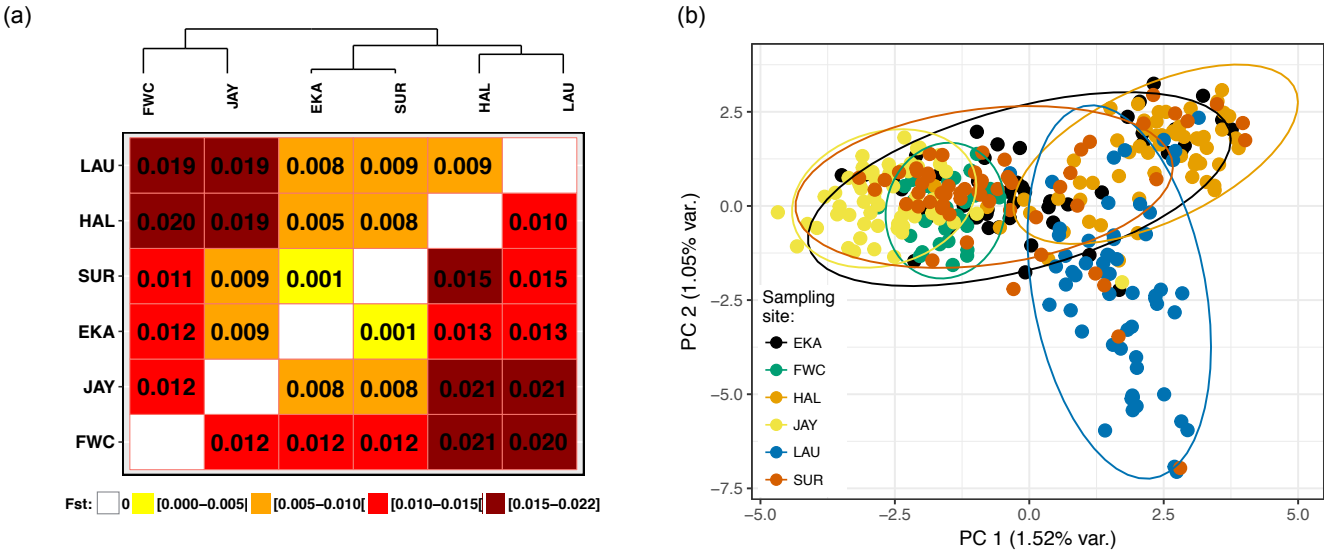
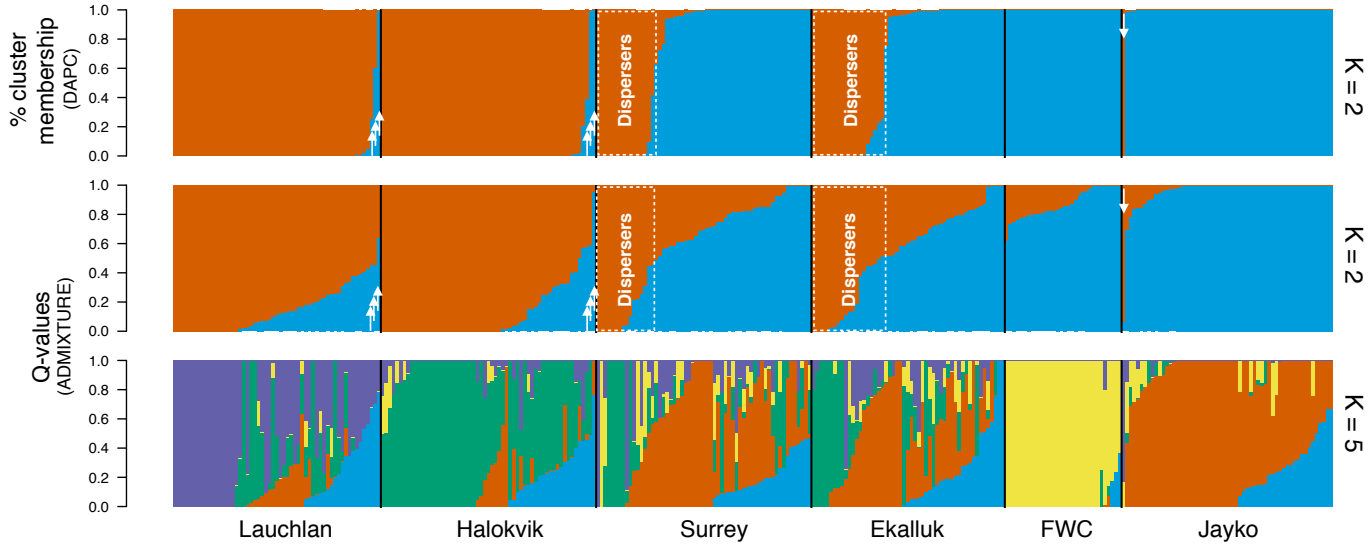


Figure 4. Description of population structure among the Arctic Char sampling sites: (a) heatmap of pairwise F_{ST} values before (above the diagonal) and after (below the diagonal) putative dispersers were removed from the samples (see text for details). All values are significantly different from zero ($p < 0.05$). (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 DPAC 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

437 Ekalluk River. Unlike the DAPC analysis, the cross-validation errors in the
438 ADMIXTURE analysis did not support the presence of two genetic groupings (K = 1 had
439 the lowest cross-validation error). Nonetheless, at *K* = 2, the individual Q-values were
440 consistent with the results of the DAPC, and all putative dispersers identified with the
441 DAPC had >50% probability of membership to the alternative genetic cluster in
442 ADMIXTURE as well. Furthermore, at K=5 the ADMIXTURE results suggested that
443 genetic differentiation among rivers was present, albeit weak.



444 **Figure 5.** Results of two clustering analyses used to identify putative dispersers. The top panel is
445 a compo-plot generated from a DAPC using two genetic clusters showing the percent cluster
446 membership of each individual in the analysis. The bottom two panels are the individual Q-values
447 from the ADMIXTURE analyses with two (middle) and five (bottom) clusters. The putative
448 dispersers are individuals with Q-values or percent cluster membership greater than 0.5 but
449 assigned to a cluster different from that of most other individuals sampled at the same sampling
450 location and are identified with a white arrow or a white dotted rectangle. Note that the order of
451 individuals varies among panels, but the identity of putative dispersers was the same in the two
452 analyses.

453

454 *Population assignments*

455 Overall, the results of the assignment tests using all 6,136 markers revealed that
456 there was sufficient power to infer the origin of individuals with high accuracy (84.3%-
457 94.5%). There was limited variation in assignment success when using different subsets
458 of individuals, indicating that the results are stable (*not shown*). Overall assignment

459 success was maximized when using 3,000 markers both when the putative dispersers
460 were retained or removed (Fig. 6a & 6c). We therefore present pairwise assignment
461 success per population based on 3,000 markers (Fig. 6b & 6d). As expected, when
462 dispersers were present in the dataset, assignment success was lower overall (84.3%
463 without imputations; 88.0% with imputations; range: 71-100%; Fig 6b), and many
464 individuals from Halokvik and Lauchlan River were assigned to Ekalluk and Surrey
465 River, and vice-versa. As expected, assignment success was improved when the putative
466 dispersers were removed from the dataset (92.3% without imputations; 94.5% with
467 imputations; range 81-100%; Fig 6d), and most of the mis-assignments involved fish
468 caught at Lauchlan River but assigned to Halokvik River, or fish caught at Jayko River
469 assigned to Ekalluk and Surrey River.

470 After confirming that there was sufficient assignment power, we used genomic
471 data to assign tagged individuals to their most likely river of origin (Table S1 online
472 supplementary materials). In this case, we used all the markers to simplify the treatment
473 of missing data (*see Methods*) and because assignment success based on all markers was
474 not substantially lower than that based on a subset of 3,000 markers (Fig. 6a & 6c).
475 Eleven of the 57 fish (19%) tagged at the Ekalluk River in 2013 and 2014 were assigned
476 to the Halokvik River and two to the Lauchlan River. Only two of the 11 putative
477 Halokvik-origin fish were detected overwintering/spawning in the Halokvik River at least
478 once, and all others were detected only in the Ekalluk system one or more years. In
479 contrast, all but one (29/30) of the fish tagged at the Halokvik River were assigned to the
480 Halokvik River itself. Six of those 29 individuals were detected overwintering/spawning
481 in both Ekalluk and Halokvik River in different years. Notably, all but one fish (32/33)
482 detected at least one year at Halokvik River were assigned to the Halokvik River. Finally,
483 fish tagged at Surrey River were assigned to all possible rivers (except Freshwater
484 Creek), but a majority (19/31) were assigned to the grouped Ekalluk-Surrey sample. In
485 short, population assignments also support the prevalence of asymmetric dispersal
486 towards the Ekalluk River.

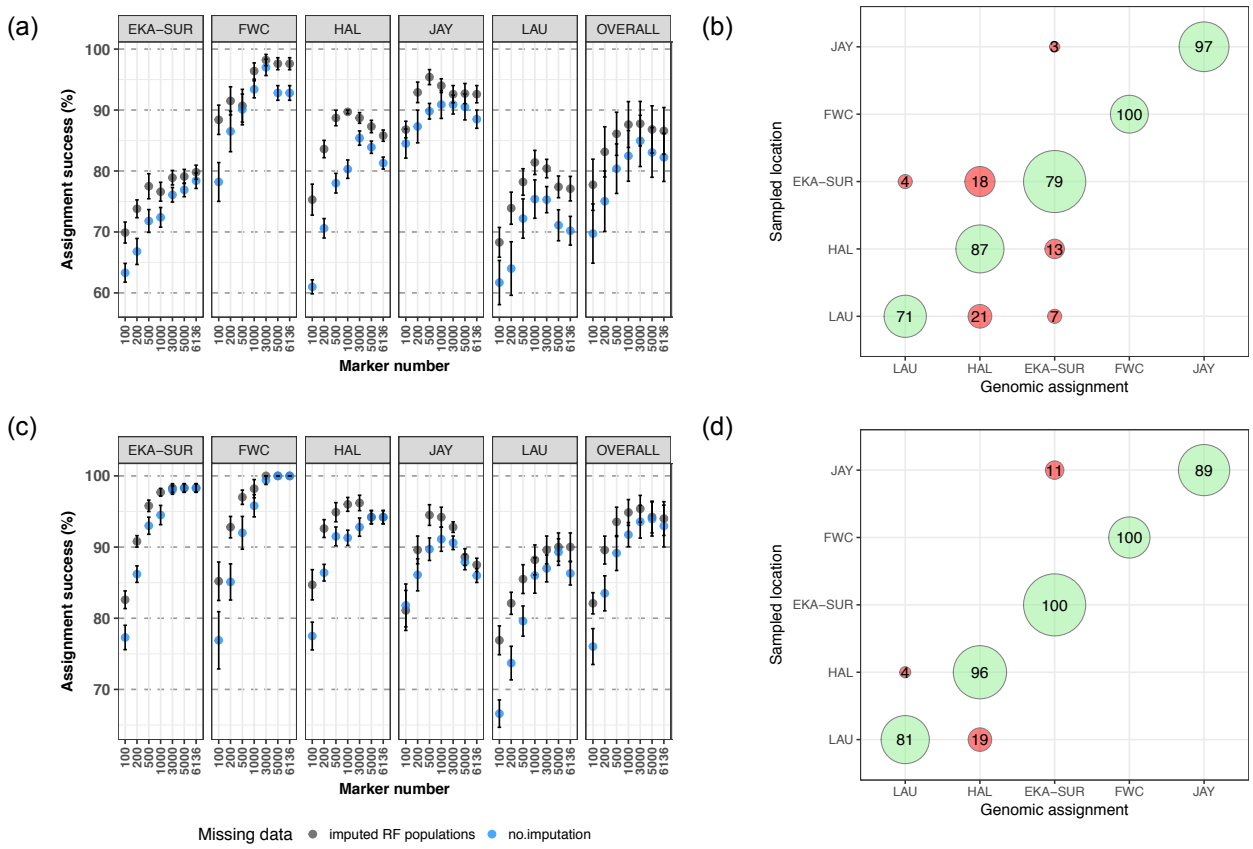


Figure 6. Results of assignment tests with all individuals (a & b) or with putative dispersers removed (c & d). The two left panels (a & c) show how assignment success per sampling location varied according to the number of markers used for the assignment, and on whether missing data were imputed using Random Forest (RF) or not. The two right panels (b & d) show results of the assignments with all markers with missing data imputed. The green circles on the diagonal display successful assignments, and the red 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.

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 S4 in online supplementary materials). This inference was highly significant with a robustness of 1 (Fig. S8 in online supplementary materials). Estimates of demographic parameters

503 produced confidence intervals of various widths, and differed most from the posterior in
504 the Ekalluk-Halokvik comparison than in the Ekalluk-Lauchlan comparison. These
505 estimates were mostly informative with regards to estimates of effective populations size,
506 divergence times (except in the Ekalluk-Lauchlan comparisons), and intensity of
507 migration rates, pointing especially to highly asymmetric gene flow (Fig. S9 & S10 in
508 online supplementary materials). This asymmetric gene flow, however, was in the
509 opposite direction of the asymmetric dispersal inferred from both the telemetry and the
510 genetic data. Indeed, gene flow was higher from Ekalluk to Lauchlan, with a ratio of
511 $M_2/M_1 = 3.08$, and from Ekalluk to Halokvik, with a ratio of $M_2/M_1 = 4.06$ (Table 3).
512 Ratios of effective population size (θ_1/θ_A) tended to indicate an expansion of the Ekalluk
513 population relative to the ancestral population in both comparisons (ratios of 9.68 and
514 3.45). Posterior distributions also indicated an expansion of the Halokvik population
515 ($\theta_2/\theta_A = 2.58$), but the Lauchlan population size was inferred to be smaller than the
516 ancestral population size ($\theta_2/\theta_A = 0.082$), although this should be interpreted cautiously
517 since the estimate of the ancestral population size was not highly accurate. Posterior
518 predictive checks in the Ekalluk-Lauchlan indicated that we were able to accurately
519 reproduce the summary statistics. In the Ekalluk-Halokvik comparison, however, three
520 statistics were significantly different from the observed data, namely the averaged
521 number of shared polymorphic sites and the averaged and standard deviations of the net
522 divergence (Da). In summary, the ABC approach provided robust estimates of
523 demographic parameters, and suggested that the asymmetric dispersal towards Ekalluk
524 River observed with both the telemetry and genomic data did not translate into
525 pronounced gene flow in that direction.
526
527 *Genome scans and functional annotation of outliers*
528 The three different methods of outlier detection identified a total of 90 markers putatively
529 under selection. The Bayescan analysis identified 30 outlier loci when the prior odds
530 were set at 100 (Fig. S11 online supplementary materials), 10 loci with prior odds of
531 1000, and 5 loci with prior odds of 5000, all with an FDR of 0.05. The PCAdapt analysis
532 was run assuming two genetic clusters after graphical evaluation of the eigenvalues
533 according to Luu et al. (2017). Consistent with the DAPC results, the first two PC axes

534 explained most variation and only those were retained to account for population structure
535 (Fig. S12 online supplementary materials). Six outliers were identified with FDR = 0.05
536 (Fig S13 online supplementary materials). Of those six, four were also identified by the
537 Bayescan run with prior odds of 100 and two with the run at prior odds of 5000. For the
538 LFMM analysis, cross-entropy was also minimal at K = 2 (Fig. S14 online supplementary
539 materials), and two latent factors were thus retained to control for population structure.
540 With an FDR = 0.05, 58 markers were identified as outliers by the LFMM analysis, none
541 of which overlapped with the markers identified by Bayescan or PCAdapt.
542 A total of 2070 of the 6136 markers were positioned onto the Brook Char linkage
543 map and appeared distributed throughout the genome on many chromosomes (Fig. 7). Of
544 the 90 unique markers identified combining the three outlier detection analyses, 34
545 provided unambiguous BLAST results against the Atlantic Salmon genome (see Table S5
546 for full list of successful BLAST results). Of those, 23 were found to be in a gene, while
547 nine were found within 100kb of a gene. Only one of the four outliers that were
548 significant in both the PCAdapt and the Bayescan had unambiguous BLAST results, but
549 three of them were located on the same 914Kb scaffold (*scaffold_363*) on the Rainbow
550 Trout genome. The marker with a successful BLAST result was within the gene
551 *FAM179B*, which codes for a protein required for the normal structure and function of
552 primary cilia, an organelle usually associated with the reception of extracellular
553 mechanical and chemical stimuli (Das *et al*. 2015). Several outlier-linked genes were
554 involved in muscle or cardiac muscle function and development (*Myocyte-specific*
555 *enhancer factor 2C*, *nebulette*, *glycogen phosphorylase muscle form*) or in
556 gluconeogenesis (*fructose-1,6-bisphosphatase 1-like*). Several other outlier-associated
557 genes have possible functions linked to axon guidance or neuronal development (*neural*
558 *cell adhesion molecule 2*, *neural cadherin-like isoform X1*; *Arf-GAP with GTPase*; *ANK*
559 *repeat and PH domain-containing protein 2*; *NEDD8-conjugating enzyme Ubc12*; *CD9*
560 *antigen-like isoform X2*). Finally, two outliers were found within immune-response genes
561 (*Interferon regulatory factor 4*; *Ig heavy chain V region-like*).

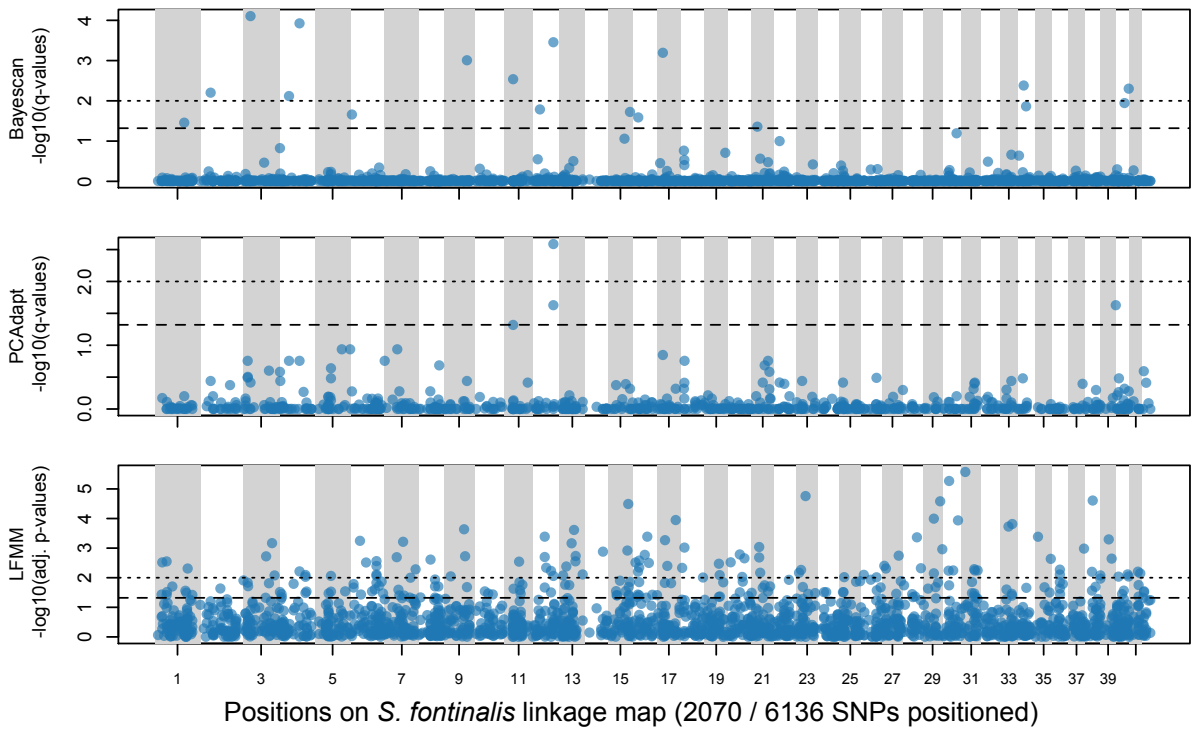


Figure 7. Manhattan plots showing the position of each marker that was successfully mapped to the closely related Brook Char genetic map (grey shading and number on the x-axis show linkage groups) and the test statistics from the two genome scan analyses (top: Bayescan and middle: PCAdapt) and the genetic environment association (bottom: LFMM). On the y-axes are the –log10 transformed values of the test statistics (q-values for Bayescan and PCAdapt and adjusted p-values for LFMM). The dashed lines represent the 0.05 significance threshold (used throughout) and the dotted line a 0.01 threshold (for reference only).

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 (Schaffer *et al*. 2016). 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 confirmed 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. Demographic simulations

582 using an ABC approach also revealed an asymmetry in gene flow, but in the opposite
583 direction, indicating that the observed asymmetric dispersal does not necessarily lead to
584 gene flow. Instead, the ABC simulations suggested that Ekalluk River, the most abundant
585 Arctic char population in the region (McGowan 1990), was a major source of gene flow
586 to surrounding areas. Together, our observations suggested that Arctic Char home to their
587 natal river to spawn, but may overwinter in rivers with the least harsh migratory route,
588 thus potentially minimizing the costs of migrations in non-breeding years. Another
589 finding facilitated by the integration of telemetry and genomic data was the possible
590 identification of local adaptation in migratory traits as indicated by the identification of
591 many outlier markers linked to genes demonstrated to have a role in muscle or cardiac
592 function. Related physiological traits have been demonstrated to evolve in response to
593 migration harshness in other anadromous salmonids (Eliason et al. 2011), thus suggesting
594 that this environmental factor could also drive local adaptation in Arctic Char.
595
596 *Asymmetric dispersal: overwintering habitat choice minimizes cost of migrations*
597 Unlike most other anadromous salmonids, which return to freshwater exclusively
598 to reproduce (Fleming 1998; Quinn 2005), Arctic Char and other *Salvelinus* species also
599 return yearly to freshwater to overwinter (Armstrong 1974; Johnson 1980). Given that
600 optimal habitats for reproduction and for overwintering likely differ, we might expect
601 individuals to migrate to separate locations depending on the purpose of the migration
602 (Dingle 2014). Precise homing to reproduction sites is the norm in salmonids, and it has
603 been suggested that local adaptation to nesting and rearing environments is an important
604 driver of philopatry (Hendry *et al*. 2004). Requirements to home, however, might be
605 relaxed when an individual is returning to freshwater to overwinter instead of to
606 reproduce (i.e., when there are no genetic or evolutionary consequences). Accordingly,
607 there is evidence that Arctic Char (and the closely related Dolly Varden; *Salvelinus*
608 *malma*) home to their natal habitats to spawn, but can use non-natal habitats to overwinter
609 (Armstrong 1974; Gyselman 1994; Moore *et al*. 2013; Gilbert *et al*. 2016). Here, we
610 documented the use of alternative freshwater habitats by the same individuals in
611 consecutive years using acoustic telemetry data. In addition, we used population genomic
612 data to infer the natal origins of individuals, and confirmed that the asymmetric dispersal

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

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

635 Taken together, the observations of asymmetric dispersal with both genomic and
636 telemetry data, but the lack of realized gene flow in the same direction, are consistent
637 with the hypothesis that individuals in breeding condition home to spawn but may select
638 overwintering habitats that minimize the costs of migrations in the years when they are
639 not in breeding condition. Indeed the Ekalluk River is one of the most easily accessible
640 rivers in the region (measured as work; Table 1). While other rivers are also easily
641 accessible (e.g., Freshwater Creek, Jayko River), they are further away from Wellington
642 Bay, and comparatively fewer fish from the Wellington Bay tagging locations travel there
643 (Moore *et al*. 2016). Other features of the freshwater habitats may also explain the

644 observed habitat choice. For example, the Ekalluk River drains the largest lake on
645 Victoria Island, Ferguson Lake, which could offer more abundant or better overwintering
646 habitats, perhaps resulting in lower overwintering mortalityNonetheless, other studies of
647 Arctic Char have also found spawning habitat accessibility to be a major constraint to
648 migrations (Gyselman 1994; Gilbert *et al.* 2016), and the easier migrations afforded by
649 the short Ekalluk River constitutes a very plausible explanation for the observed patterns
650 of dispersal.

651 This hypothesis, however, assumes the ability of Arctic Char to assess which river
652 provides the most readily accessible overwintering habitat. One possibility is that
653 individuals explore a variety of freshwater habitats during their summer migrations and
654 assess their accessibility. Observations of back and forth movement during homing in
655 other salmonids (Quinn 2005), and the regular use of different estuaries by individuals
656 during the summer documented by our telemetry work (Moore *et al.* 2016) suggest this is
657 plausible. Another possibility is that the decision of where to overwinter relies in part on
658 collective navigation (i.e., increased ability to navigate through social interactions), a
659 possibility that has been discussed in relation to salmon homing (Berdahl *et al*. 2014).
660 Given that the Ekalluk River has the most abundant population of Arctic Char in the
661 region (McGowan 1990; Day & Harris 2013), it is also possible that collective navigation
662 biases dispersal towards the river with the greater number of individuals. Continued
663 collection of telemetry data will allow future tests of some of these hypotheses.

664 In summary, the integration of telemetry and genomic data allowed nuanced
665 insights into the complex migratory behaviour of anadromous Arctic Char. Notably, it
666 allowed circumventing logistical challenges posed by the sampling of baseline samples.
667 In most salmonids, sampling in the freshwater is used to ensure the origin of sampled fish
668 for baseline collections, but here the use of different freshwater habitats by adults within
669 their lifetime precludes any certainty in inferring origins of sampled adults even if
670 collected in freshwater. Sampling juveniles prior to their first marine migration (e.g.,
671 Moore *et al*. 2013) would also be an option, but attempts at collecting juveniles in 2010
672 and 2015 failed. Given the complex migratory behaviour of Arctic Char and the logistical
673 challenges associated with sampling, the integration of telemetry and genomic data
674 provided a powerful tool for understanding the ecology of this species.

675

676 *Evidence for local adaptation in migratory traits*

677 Anadromous salmonids display tremendous diversity in migratory traits both
678 among and within species, and their study has contributed to our understanding of
679 migration ecology (Hendry *et al*. 2004; Dingle 2014). The application of molecular tools
680 has provided important insights to the study of the genetic basis of migratory traits in
681 salmonids. For example, O'Malley & Banks (2008) used a candidate gene approach to
682 link variation in two circadian rhythm genes, *OtsClock1a* and *OtsClock1b*, to variation in
683 run timing in anadromous Chinook salmon (*O. tshawytscha*). Other studies have
684 identified SNP markers linked to migration timing in both Chinook salmon (Brieuc *et al.*
685 2015) and Steelhead Trout (*O. mykiss*; Hess *et al*. 2016). Finally, Hecht *et al*. (2015)
686 linked variation in RAD markers with 24 environmental correlates using redundancy
687 analysis in Chinook salmon, finding that freshwater migration distance was the
688 environmental variable explaining most of the genomic variation. This last result parallels
689 our own conclusions regarding the importance of migratory harshness, and is also
690 corroborated by many classic studies demonstrating the importance of migratory
691 difficulty as a selective factor driving adaptation in morphological, physiological, and
692 life-history traits (Shaffer & Elson 1975; Bernatchez & Dodson 1987; Crossin *et al*.
693 2004; Eliason *et al*. 2011).

694 In the present study, we used genome scans and genetic-environment correlations
695 with migratory difficulty to identify markers putatively under divergent selection. These
696 approaches have been used extensively to identify putative targets of selection, but also
697 have several limitations, which have been amply discussed in the literature (Lotterhos &
698 Whitlock 2014; Haasl & Payseur 2015; Bernatchez 2016; Hoban *et al*. 2016). For
699 example, recent demographic history and patterns of isolation by distance have been
700 shown to lead to high levels of false positives (Lotterhos & Whitlock 2014; Hoban *et al*.
701 2016). The geographical proximity of the populations under study, however, precludes
702 important differences in terms of demographic history since the re-colonization of the
703 area from a single source population in the last 6,500 years (Moore *et al*. 2015). In
704 addition, Lotterhos & Whitlock (2015) concluded that maximum power in genome scan
705 studies could be achieved by examining geographically proximate and genetically similar

706 populations that experience contrasting environments, criteria that our system fulfilled.
707 Nevertheless, few of the outlier markers identified by the three methods overlapped, a
708 common finding (de Villemereuil *et al*. 2014; Gagnaire *et al*. 2015). Interestingly, three
709 of the four markers that overlapped between the PCAdapt and the Bayescan approaches
710 were co-located on the same 914 Kb scaffold (*scaffold_363*) on the Rainbow Trout
711 genome (two of the markers failed to provide unambiguous BLAST results on the better
712 annotated but more distantly related Atlantic Salmon genome). The co-localization of
713 three outliers with evidence from two separate methods therefore suggests the importance
714 of this genomic region, and provides some confidence in the success of the method.
715 In accordance with our *a priori* expectation that migration harshness is a major
716 selective factor differing among rivers, several of the outlier SNPs were linked to muscle
717 and cardiac functions and development. One of the outlier SNPs was located within 75Kb
718 (5' side) of *myocyte enhancer factor 2* (MEF2), a transcription factor acting as an
719 important regulator of vertebrate skeletal muscle differentiation and heart development
720 (Potthof & Olson 2007). Supporting the relevance of this protein for migrating salmonids,
721 cardiac mRNA levels of MEF2 were elevated in Atlantic salmon following 10 weeks of
722 experimental exercise training (Castro *et al*. 2013). An outlier was also located within the
723 gene coding for *nebulette*, a cardiac-specific actin-binding protein essential in the
724 structure of the sarcomere Z-disc (Bonzo *et al*. 2008). Another marker was within the
725 gene coding for *muscle glycogen phosphorylase* (PYGM), the enzyme responsible for
726 breaking up glycogen into glucose subunits to power muscle cells (Kitaoka 2014).
727 Mutations in this gene in humans are linked to McArdle disease, the major symptom of
728 which is exercise intolerance in the form of premature fatigue, myalgia and muscle
729 cramps (Kitaoka 2014). Finally, another marker was within the gene coding for *fructose-*
730 *1,6-biphosphatase*, which is involved in the conversion of glycerol into glucose (i.e.,
731 gluconeogenesis; Lamont *et al*. 2006). These results are consistent with the hypothesis
732 that migration harshness is a major selective force in anadromous Arctic Char, and
733 provide a basis to formulate hypotheses regarding the functional traits underlying long-
734 distance migrations.
735 The annotations of several other genes are also noteworthy in the context of
736 homing, which in salmonids involves memory, learning, and chemosensing (Quinn

737 2005). First, the protein coded by the previously mentioned MEF2 gene is also enriched
738 in neurons, in the cerebellum, the cerebral cortex and the hippocampus, and its
739 involvement in synaptic plasticity suggests a role in learning and memory (Potthof &
740 Olson 2007). Second, the only marker from Rainbow Trout *scaffold_363* that could be
741 localized is within the gene coding for *FAM179B*, a protein required for the structure and
742 function of the primary cilia, an organelle usually associated with the reception of
743 extracellular mechanical and chemical stimuli, and which has been linked to chemotaxis
744 in *C. elegans* (Das et al. 2015). Note, however, that this scaffold contains 36 predicted
745 ORFs, and other linked genes could easily be the targets of selection instead of
746 *FAM179B*. Finally, several other markers were localized in or near genes coding for
747 proteins that have functions in the nervous system, particularly axon guidance (*neural*
748 *cell adhesion molecule 2*, *neural cadherin-like isoform X1*; *Arf-GAP with GTPase*; *ANK*
749 *repeat and PH domain-containing protein 2*; *NEDD8-conjugating enzyme Ubc12*; *CD9*
750 *antigen-like isoform X2*; see table S4 for details). There is some evidence from Pacific
751 salmon that homing has both a learned and an innate, genetically determined component
752 (McIsaac & Quinn 1988). The abundance of putative targets of selection identified here
753 linked to the development and function of the nervous system is therefore interesting in
754 that regard, but much additional work would obviously be required to establish causal
755 links (Barrett & Hoekstra 2011). Further, the low density of markers used in the present
756 study did not allow an exhaustive assessment of the targets of natural selection
757 throughout the Arctic Char genome. Nonetheless, the evidence that local adaptation in a
758 variety of biological functions evolved despite low genetic differentiation and in the short
759 time since deglaciation appears fairly robust.
760
761 *Conclusions*
762 Migration harshness has long been identified as a major driver of adaptation in
763 anadromous salmonids (Shaffer & Elson 1975; Bernatchez & Dodson 1987; Eliason *et al*.
764 2011; Hecht *et al*. 2015). We here used an integrative approach, linking genome-wide
765 data with telemetry, to provide evidence that migration harshness drives both
766 overwintering habitat choice and local adaptation in anadromous Arctic Char. The
767 genomic data suggested the importance of migratory harshness in driving local adaptation

through the identification of several putative targets of selection linked to muscle and cardiac functions. Future work in this system will build on these results to describe physiological differences among populations in relations to migration harshness. These will be particularly relevant in the context of adaptation to a changing Arctic, where migratory environments could exert an important selective pressure on populations (e.g., Eliason *et al*. 2011). Many of our conclusions would have been difficult, if not 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.

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.

Barrett RDH, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. *Nature reviews: Genetics*, **12**, 767–780.

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.

887 de Villemereuil P, Frichot E, Bazin E, François O, Gaggiotti OE (2014) Genome scan
888 methods against more complex models: when and how much should we trust
889 them? *Molecular Ecology*, **23**, 2006–2019.
890 Dingle, H. (2014) Migration: the biology of life on the move. 2nd edition. Oxford
891 University Press, Oxford UK. 306 pp.
892 Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J., and Ovenden, J.R. (2014)
893 NeEstimator v2: re-implementation of software for the estimation of
894 contemporary effective population size (Ne) from genetic data. *Molecular*
895 *Ecology Resources,* **14**, 209–214.
896 Dutil, J.-D. (1986) Energetic constraints and spawning interval in the anadromous Arctic
897 Charr (*Salvelinus alpinus*). *Copeia*, **4**, 945–955.
898 Eliason, E.J., Clark, T.D., Hague, M.J., Hanson, L.M., Gallagher, Z.S., Jeffries, K.M.,
899 Gale, M.K., Patterson, D.A., Hinch, S.G., and Farrell, A.P. (2011) Differences in
900 thermal tolerance among sockeye salmon populations. *Science,* **332**, 109–112.
901 Fleming IA (1998) Pattern and variability in the breeding system of Atlantic salmon
902 (Salmo salar), with comparisons to other salmonids. *Canadian Journal of*
903 *Fisheries and Aquatic Sciences*, **55(Suppl. 1)**, 59–76.
904 Foll, M., and Gaggiotti, O. (2008) A Genome-Scan Method to Identify Selected Loci
905 Appropriate for Both Dominant and Codominant Markers: A Bayesian
906 Perspective. *Genetics,* **180**, 977–993.
907 Franchini P, Irisarri I, Fudickar A *et al.* (2017) Animal tracking meets migration
908 genomics: transcriptomic analysis of a partially migratory bird species. *Molecular*
909 *Ecology*, DOI: 10.1111/mec.14108
910 Frichot, E., Schoville, S.D., Bouchard, G., and François, O. (2013) Testing for
911 associations between loci and environmental gradients using latent factor mixed
912 models. *Molecular Biology and Evolution,* **30**, 1687–1699.
913 Gagnaire P-A, Broquet T, Aurelle D *et al.* (2015) Using neutral, selected, and hitchhiker
914 loci to assess connectivity of marine populations in the genomic era. *Evolutionary*
915 *Applications*, **8**, 769–786.
916 Gilbert, M., Donadt, C.R., Swanson, H.K., and Tierney, K.B. (2016) Low Annual
917 Fidelity and Early Upstream Migration of Anadromous Arctic Char in a Variable
918 Environment. *Transactions of the American Fisheries Society,* **145**, 931–942.
919 Gosselin, T, Anderson, E. C., Bradbury, I. (2016). assigner: Assignment Analysis with
920 GBS/RAD Data using R. Rpackage version 0.4.1.
921 https://github.com/thierrygosselin/assigner. doi : 10.5281/zenodo.51453
922 Gross MR, Coleman RM, McDowall RM (1988) Aquatic productivity and the evolution
923 of diadromous fish migration. *Science*, **239**, 1291–1293.
924 Haasl RJ, Payseur BA (2016) Fifteen years of genomewide scans for selection: trends,
925 lessons and unaddressed genetic sources of complication. *Molecular Ecology*, **25**,
926 5–23.
927 Harris, L.N., Moore, J.S., Bajno, R., and Tallman, R.F. (2016) Genetic Stock Structure of
928 Anadromous Arctic Char in Canada's Central Arctic: Potential Implications for
929 the Management of Canada's Largest Arctic Char Commercial Fishery. *North*
930 *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, 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.

976 Lamont BJ, Visinoni S, Fam BC *et al.* (2006) Expression of human fructose-1,6-
977 bisphosphatase in the liver of transgenic mice results in increased glycerol
978 gluconeogenesis. *Endocrinology*, **147**, 2764–2772.
979 Liedvogel, M., Akesson, S., and Bensch, S. (2011) The genetics of migration on the
980 move. *Trends in Ecology & Evolution* **26**, 561–569.
981 Lien, S., Koop, B.F., Sandve, S.R., Miller, J.R., Kent, M.P., Nome, T., Hvidsten, T.R.,
982 Leong, J.S., Minkley, D.R., Zimin, A., Grammes, F., Grove, H., Gjuvsland, A.,
983 Walenz, B., Hermansen, R.A., Schalburg, von, K., Rondeau, E.B., Di Genova, A.,
984 Samy, J.K.A., Olav Vik, J., Vigeland, M.D., Caler, L., Grimholt, U., Jentoft, S.,
985 Inge Våge, D., de Jong, P., Moen, T., Baranski, M., Palti, Y., Smith, D.R., Yorke,
986 J.A., Nederbragt, A.J., Tooming-Klunderud, A., Jakobsen, K.S., Jiang, X., Fan,
987 D., Hu, Y., Liberles, D.A., Vidal, R., Iturra, P., Jones, S.J.M., Jonassen, I., Maass,
988 A., Omholt, S.W., and Davidson, W.S. (2016) The Atlantic salmon genome
989 provides insights into rediploidization. *Nature,* **533**, 200–207.
990 Lotterhos, K.E., and Whitlock, M.C. (2014) Evaluation of demographic history and
991 neutral parameterization on the performance of FST outlier tests. *Molecular*
992 *Ecology,* **23**, 2178–2192.
993 Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local
994 adaptation depends on sampling design and statistical method. *Molecular*
995 *Ecology*, **24**, 1031–1046.
996 Luu, K., Bazin, E., and Blum, M.G.B. (2017) PCAdapt: an R package to perform genome
997 scans for selection based on principal component analysis. *Molecular Ecology*
998 *Resources,* **17**, 67–77.
999 Mastretta-Yanes A, Arrigo N, Alvarez N *et al.* (2015) Restriction site-associated DNA
1000 sequencing, genotyping error estimation and de novo assembly optimization for
1001 population genetic inference. *Molecular Ecology Resources*, **15**, 28–41.
1002 McGowan DK (1990) Enumeration and Biological Data from the Upstream Migration of
1003 Arctic Charr, *Salvelinus alpinus* (L.), in the Cambridge Bay Area, Northwest
1004 Territories, 1979-1983. Canadian Data Report of Fisheries and Aquatic Sciences
1005 811: iv + 27 p.
1006 McIsaac, D. O., and Quinn, T. P. (1988) Evidence for a hereditary component in homing
1007 behavior of chinook salmon (*Oncorhynchus tshawytscha*). Can. J. Fish. Aquat.
1008 Sci. **45**: 2201-2205.
1009 Meirmans, P.G., and van Tienderen, P.H. (2004) GENOTYPE and GENODIVE: two
1010 programs for the analysis of genetic diversity of asexual organisms. *Molecular*
1011 *Ecology Notes,* **4**, 792–794.
1012 Mascher, M., Wu, S., Amand, P.S., Stein, N., and Poland, J. (2013) Application of
1013 genotyping-by-sequencing on semiconductor sequencing platforms: a comparison
1014 of genetic and reference-based marker ordering in barley. *PLoS ONE,* **8**, e76925.
1015 Moore, J.-S., Harris, L.N., Tallman, R.F., and Taylor, E.B. (2013) The interplay between
1016 dispersal and gene flow in anadromous Arctic Char (*Salvelinus alpinus*):
1017 implications for potential for local adaptation. *Canadian Journal of Fisheries and*
1018 *Aquatic Sciences*, **70**, 1327–1338.
1019 Moore JS, Bajno R, Reist JD, Taylor EB (2015) Post‐glacial recolonization of the North
1020 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.

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.

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., 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*

1067 *leucas*) despite nuclear gene flow: implication for the endangered belugas in
1068 eastern Hudson Bay (Canada). *Conservation genetics,* **13**, 419–433.
1069 Weir, B.S., and Cockerham, C.C. (1984) Estimating F-statistics for the analysis of
1070 population structure. *Evolution*, **38**, 1358–1370.
1071 Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ (2016) GMAP and GSNAP for
1072 Genomic Sequence Alignment: Enhancements to Speed, Accuracy, and
1073 Functionality. In: *Statistical Genomics* Methods in Molecular Biology. pp. 283–
1074 334. Springer New York, New York, NY.
1075

1076

Table 1. Sample information, river characteristics, and basic genetic diversity estimates for the six sampling locations used in the present study. Work is the product of river length and elevation gain, standardized to a mean of zero and a standard deviation of one. 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'). 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)	Work (std.)	Sampling year	Sample type	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	0.175	2012	Baseline	57/57	0.102	0.105	0.03	335 (318 – 353)
Halokvik River	N69.16° W107.09°	50.4	118	1.948	2013	Baseline	60/59	0.099	0.103	0.037	823 (736 - 934)
					2013	Tagged	34/30				-
Surrey River	N69.45° W106.68°	18	35	-0.338	2013	Baseline	60/59	0.102	0.105	0.023	892 (798 – 1009)
					2014	Tagged	31/31				-
Ekalluk River	N69.41° W106.31°	3.8	15	-0.584	2013	Baseline	60/53	0.104	0.106	0.017	1081 (935 – 1280)
					2013	Tagged	30/26				-
					2014	Tagged	31/31				-
Freshwater Creek	N69.12° W105.00°	3.6	7	-0.598	1992	Baseline	32/32	0.105	0.105	0.001	389 (351 – 436)
Jayko River	N69.72° W103.28°	5.4	3	-0.602	2014	Baseline	60/58	0.104	0.107	0.022	368 (351 – 387)

Table 2. Summary of the filter steps and number of SNPs or genotypes removed at each step. Allelic imbalance corresponds to the ratio of the number of reads for the major and minor allele in heterozygous genotypes. Note that each filter was run independently and that only SNPs/genotypes that passed all filters were retained, therefore the total number of SNPs/genotypes filtered out does not equal the sum of the number of SNPs filtered out at each step.

Filter [value]	**Number of SNPs/genotypes filtered out**
Depth of coverage [≤ 10000]	3 SNPs removed
Depth of coverage [≥ 5]	6,930,852 genotypes removed
Maximum allelic imbalance [≤ 4]	5,021 genotypes removed
Genotype likelihood [≥ 6]	5,937,959 genotypes removed
Proportion of individuals with a genotype for the SNP [≥ 70%]	153,788 SNPs removed
MAF global [≥0.01]	236,678 SNPs removed
MAF sampling site [≥0.05]	249,951 SNPs removed
Heterozygosity [≤0.6]	3,558 SNPs removed
Minimum Fis [≥ -0.4]	4,237 SNPs removed
Maximum Fis [≤ 0.4]	31,294 SNPs removed
Number of SNPs per locus [≤ 10]	57,694 SNPs removed
Summary	
Total number of SNPs/loci in unfiltered catalogue	286,584 SNPs/81,919 loci
Total number of SNPs filtered out	276,875 SNPs (96.6%)
Total number of SNPs retained	9,709 SNPs
Only one SNP per locus (1st) retained	6,230 SNPs
Sex-linked SNPs removed	6,136 SNPs

Table 3. 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]