Leveraging whole genomes, mitochondrial DNA, and haploblocks to decipher complex demographic histories: an example from a broadly admixed arctic fish

Xavier Dallaire*[1,2], Eric Normandeau[3], Thomas Brazier[4], Les Harris[5]; Michael M. Hansen, Michael[6], Claire Mérot[4], Jean-Sébastien Moore[1,2]

[1] Institut de biologie intégrative et des systèmes, Université Laval, Québec, Canada

[2] Centre d'Études Nordiques, Université Laval, Québec, Canada

[3] Plateforme de bio-informatique de l'IBIS, Université Laval, Québec, Canada

[4] CNRS, UMR 6553 ECOBIO, Université de Rennes, Rennes, France

[5] Freshwater Institute, Fisheries and Oceans Canada, Canada

[6] Department of Biology, Aarhus University, Aarhus, Denmark

*Corresponding author's email: xavier.dallaire.2@ulaval.ca

Abstract

The study of phylogeography has transitioned from mitochondrial haplotypes to genome-wide analyses, blurring the line between this field and population genomics. Whole-genome sequencing offers the opportunity to join use both and provides the density of markers necessary to investigate genetic linkage and recombination along the genome. This facilitates the unraveling of complex demographic histories of admixture between divergent lineages, as is often the case in species evolving in recently deglaciated habitats. In this study, we sequenced 1120 Arctic Char genomes from 33 populations across Canada and Western Greenland to characterize patterns of genetic variation and diversity, and how they are shaped by hybridization between the Arctic and Atlantic glacial lineages. Several lines of evidence supported mito-nuclear discordance in lineage

distribution, with all Canadian populations under the 66th parallel being characterized by introgression from the Atlantic lineage, leading to higher nuclear genetic diversity. By scanning the genome using local PCAs, we identified putative low-recombining haploblocks as local ancestry tracts from either lineage and described the impacts of recombination on the introgression landscape in admixed populations. Finally, we inferred conflicting origins of recolonization using whole genomes vs. ancestry tracts for the Arctic lineage, suggesting that haplotypes sheltered from introgression by low recombination could enlighten complex post-glacial histories. Overall, we argue that Whole-Genome Sequencing, even at low depths of coverage, provides a versatile approach to the study of phylogeographic dynamics.

Introduction

Understanding a species' evolutionary and demographic history is central to interpreting the processes that produced its contemporary genetic diversity and population structure. Since its inception, the field of phylogeography has aimed to interrogate the micro-evolutionary processes responsible for the geographical distribution of genetic lineages (Avise et al. 1987). For two decades, phylogeography was dominated by the study of gene trees based on mitochondrial DNA (mtDNA) sequences because of their easy sequencing, short length, and simple uniparental inheritance without recombination (Avise et al. 1987; Edwards et al. 2015). With the advent of high-throughput sequencing, phylogeographic studies now commonly use thousands of nuclear DNA (nuDNA) markers such as Single Nucleotide Polymorphisms (SNPs). Frequent cases of discordance between mitochondrial and nuclear genetic variation have highlighted the limitations of

47 mtDNA (Ballard and Whitlock 2004; Zink and Barrowclough 2008; Toews and Brelsford
48 2012), e.g. the effect of sex-biased migration on genealogies for these matrilineally
49 inherited markers. Despite these limitations, mtDNA remains a valuable tool in
50 phylogeographic studies: being haploid, non-recombining, and having a lower effective
51 population size, they provide a rich source of information independent from the nuclear
52 genome. Used in conjunction with nuclear markers, they offer a more comprehensive
53 understanding of species' evolutionary histories and genetic diversity.

54 The integration of genome-wide nuclear markers into phylogeography has also
55 blurred the distinctions between this field and population genetics (Edwards et al. 2015).
56 Population genomics continues to rely heavily on averaging the genome-wide genetic
57 variation of SNPs, for example, as a means to investigate population structure, while
58 ignoring fluctuations in genetic variation at a finer genomic scale (Bhatia et al. 2013; Peter
59 2022). However, these "vertical" approaches stacking information from individual genetic
60 markers across individuals usually ignore or purposely try to reduce the impacts of linkage
61 disequilibrium, neglecting the "horizontal" signal contained in haplotypes (Leitwein et al.
62 2020). Haplotypes are usually defined as DNA sequences where alleles for multiple
63 markers are inherited as a block. Unlike SNP-based approaches focusing on mutations
64 and assuming independence between markers, haplotypes carry information about
65 recombination events and have the potential to increase accuracy in the inference of
66 selection, gene flow, and population structure (Martin et al. 2017; Shipilina et al. 2023). In
67 particular, long haplotypes, sometimes called haploblocks, may be maintained in low
68 recombination regions or within structural rearrangements, providing information about

69 deeper coalescent times than short haplotypes within high recombination regions (Lawson
70 et al. 2012).

71 Phylogeographic studies are particularly relevant for arctic species, as their
72 habitats and ranges were among the most impacted by the glacial cycles of the
73 Pleistocene. As demographic bottlenecks and extended periods in allopatry lead to the
74 accumulation of divergence between populations that survived in distinct glacial refugia
75 (Hewitt 1996; Hewitt 2000; Shafer et al. 2010), post-glacial recolonizations are commonly
76 the stage for secondary contact between previously isolated lineages (Bernatchez and
77 Wilson 1998; Hewitt 1999). These cases of hybridization between lineages lead to an
78 admixture of genetic backgrounds, producing a mosaic of local ancestry tracts that tend
79 to shorten through the recombination of chromosomes over generations of backcrossing
80 (Liang and Nielsen 2014; Leitwein et al. 2020). Whole-genome averaging approaches can
81 struggle to properly identify admixture at a relatively recent glacial timescale, as isolation-
82 by-distance can produce similar genomic signatures (Wiens and Colella 2024). Therefore,
83 integrating multiple aspects of the genome (e.g., mtDNA and haplotype information) may
84 help distinguish between true admixture events and patterns that arise due to gradual
85 changes in genetic variation over geographical space.

86 Arctic char (*Salvelinus alpinus*) is a widespread circumpolar species with the
87 northernmost distribution observed in freshwater or diadromous fishes, most of which
88 coincides with regions covered by ice sheets at the last glacial maximum, around 18,000
89 years ago (Hocutt and Wiley 1986; Bernatchez and Wilson 1998). Arctic Char has been
90 called “the most diverse vertebrate on earth” (Klemetsen 2010) with many localities

91 harboring morphs diverging in size, diet, habitat choice, phenology, migratory tactic, or life
92 trait history traits (Reist et al. 2013). Some authors recognize dozens of subspecies, and
93 five main glacial lineages have been defined using mtDNA (Brunner et al. 2001; Moore et
94 al. 2015). However, these fail to offer a perfectly discrete classification of wild populations
95 around the globe, as post-glacial secondary contact zones have been described between
96 most pairs of neighboring mtDNA lineages (Salisbury et al. 2019; Dallaire et al. 2021;
97 Gordeeva et al. 2021; Jacobsen et al. 2022). With their unique blend of moderate
98 philopatry, higher dispersal rate than other anadromous salmonids (Moore et al. 2013),
99 and range spanning the most recently deglaciated region around the globe, Arctic Char
100 offers unique opportunities for the study of the distribution of genetic diversity in post-
101 recolonization populations. Notably, the origin of the post-glacial recolonization of the
102 'Arctic' mtDNA lineage, presently found throughout most of the Canadian Arctic, has
103 remained cryptic, while both Beringia and the Arctic Archipelago have been suggested as
104 potential refugia (Moore et al. 2015).

105 In this study, we used low-coverage whole-genome sequencing to investigate the
106 recolonization of the recently deglaciated Canadian Arctic by several lineages of Arctic
107 Char and their putative hybridization. We sequenced 1,120 Arctic Char genomes (Fig. 1a;
108 average coverage = 2X, with 100 samples in 5 populations further sequenced at a depth
109 of 8X) from 33 river systems across the 4,000 km wide range of the previously defined
110 Arctic mitochondrial lineage in northern Canada and western Greenland to characterize
111 patterns of genetic variation, diversity, and admixture with the neighboring Atlantic
112 lineage. We combined information from both mitochondrial and nuclear genomes

113 including mtDNA haplotype networks, genome-wide admixture and ancestry, low-
114 recombining haploblocks, and demographic inference to resolve the history of the
115 secondary contact between those glacial lineages. Finally, to highlight the complexity of
116 this post-glacial evolutionary history, we aimed to infer the origin of recolonization of the
117 Arctic lineage by comparing derived allele distributions across the genome and inside
118 Arctic haploblocks potentially carrying different non-recombining ancestry.

119 Results

120 *Distribution of nuclear genetic variation across the North American Arctic*

121 Nuclear genome-wide variation at 5.98 million SNPs (average coverage = 2X, Fig. S1) in
122 1,120 anadromous Arctic Char from 33 Canadian and Greenlandic rivers (Fig. 1a, Table
123 S1) displayed a strong substructure associated with rivers, which grouped into two major
124 genetic clusters corresponding to Northern and Southern areas. Genome-wide pairwise
125 F_{ST} ranged from 0.012 to 0.381 among Canadian rivers, while comparisons including
126 Greenlandic sites reached up to F_{ST} = 0.632 (pop. 1 and 33) (Fig 2a). The general
127 structure associated with longitude was visible on the first axis of a principal components
128 analysis (Fig 1b), explaining 63.9% and 5.5% of the genome-wide variance, respectively
129 before and after stringent LD-pruning (375,236 SNPs). As expected for homing salmonids,
130 each river system appeared genetically differentiated in an NGSadmix analysis (Fig. S2).

131 Pairwise differentiation among populations, both based on the F_{ST} index and allele
132 frequency differences (AFD), highlighted a genetic divide between the northern (pop. 1-
133 16) and southern (pop. 17-30) Canadian Arctic around the 67th parallel (Fig. 2a). While

134 populations from both regions exhibited patterns of isolation-by-distance (IBD), southern
135 populations were more differentiated at equal marine distances than northern populations
136 (Fig. 2b-c). Greenlandic populations and pop. 30 (south-eastern Nunavik) were
137 comparatively more differentiated than other populations and did not fit the Canadian IBD
138 patterns. Intra-population genetic diversity, as measured by nucleotide diversity (Θ_π), was
139 also higher in southern (median widowed per-site Θ_π = 1.13-2.77 x 10^{-3}) than in northern
140 (Θ_π = 0.72-2.09 x 10^{-3}) Canadian populations (Fig. 3d).

141 *Admixture between lineages is a major determinant of variation in southern populations*

142 Several lines of evidence support the hypothesis that the large intra- and inter-population
143 diversity described earlier in the Southern Canadian Arctic and Greenland is due to
144 admixture between the Arctic and Atlantic glacial lineages.

145 First, we realigned and genotyped data from 20 Arctic Char collected in a Swedish
146 population (pop. 34, Saha et al. 2024; depth of coverage = 5X) to serve as an outgroup
147 from the Atlantic lineage and conducted an admixture analysis (K=2, NGSadmix).
148 Southern Canadian and north-western Greenlandic populations had mixed genome-wide
149 ancestry for the two groups, distributed along a cline-like northwest-to-southeast gradient,
150 suggesting admixture between the Arctic (i.e., northern Canadian populations) and
151 Atlantic (i.e., Swedish outgroup) lineages (Fig. 3a). Within population variance in
152 proportions of Atlantic ancestry was low (Fig. 3a), and putatively admixed populations had
153 higher observed heterozygosity (Fig. 3b, Fig. S3) and nucleotide diversity (Fig. 3d),
154 supporting the hypothesis of secondary contact between the Arctic and Atlantic lineages.

155 Second, to further test the hypothesis of admixture between lineages in the
156 southern part of the study area, we computed f_4 statistics (Reich et al. 2009; Patterson et
157 al. 2012) using a Dolly Varden (*Salvelinus malma*) population from Yukon, Canada, as an
158 outgroup. This test provided support for introgression from the Atlantic lineage (Sweden)
159 toward the southern Arctic populations (Fig. 4c). Finally, using demographic modeling
160 performed with fastsimcoal, with the Swedish population and 8X data for COP (pop. 4)
161 and TJL (pop. 25), a model including ancient gene flow between the Atlantic lineage and
162 the southern Canadian Arctic was better supported than models with no or more recent
163 gene flow (Fig. 4a-b; see demographic parameter estimation in Fig. S4).

164 *Mito-nuclear discordance in admixture and diversity patterns*

165 To leverage an independent source of historical demographic information and to compare
166 our results with previous studies, we extracted mitochondrial D-loop sequences (993 bp)
167 from whole-genome data (Fig 1c). Consistent with previously published data based on
168 mtDNA, but contrary to our results using the nuclear genome, the Arctic lineage dominates
169 in populations of the entire Arctic archipelago outside Labrador (Moore et al. 2015;
170 Salisbury et al. 2019; Dallaire et al. 2021), with only one Canadian sample from the
171 present study carrying an Atlantic D-loop haplotype (in pop. 28). In contrast, the Atlantic
172 and Arctic mitochondrial lineages co-occurred in western Greenland (Jacobsen et al.
173 2021). As previously observed, a single mitochondrial haplotype dominated the Arctic
174 lineage (77.5% of samples shared the most common D-loop haplotype). In Greenlandic
175 populations, the individuals carrying haplotypes from either mtDNA lineages did not differ

176 in their nuclear genome ancestry level inferred with NGSadmix (t-test per population: $p >$
177 0.05).

178 Whole-mitogenome genetic diversity inside the Arctic mtDNA lineage also
179 displayed different trends compared to the nuclear genome. While nuclear diversity was
180 the highest in the admixed populations of the South Arctic and Greenland, both
181 mitochondrial Θ_W and haplotype diversity were higher near the westernmost population
182 than in the southeastern part of the study area (Fig. 3c, Fig. S5a,d). Mitochondrial
183 nucleotide diversity (Θ_π) and Tajima's D were lower at intermediate distances from the
184 westernmost population, i.e. the center of the Arctic mtDNA lineage range. Finally, the
185 frequency of the most frequent whole-mitogenome haplotype, present in 34% of all Arctic
186 lineage samples, was lower in populations at both extremities of the range of the lineage,
187 i.e. in Inuvialuit and Greenland (Fig. S5e).

188 *Heterogeneity in the introgression along the genome*

189 To investigate the fine-scale genomic landscape of admixture between the Arctic
190 and Atlantic lineages, we scanned the genome for distinct patterns of genetic variation
191 and detected putative local ancestry tracts for both lineages.

192 We used local Principal Component Analyses (PCA) and multidimensional scaling
193 (MDS) in 100 SNPs-wide windows to identify genomic regions displaying patterns of
194 genetic variation that were significantly different from the chromosome-wide population
195 structure. We retained 117 independent regions that contained at least 10 MDS outlier
196 windows (e.g. Fig. 5a, see all outlier regions in Fig S6-7). These regions ranged from 98
197 kb to 5.26 Mb in length (median = 825 kb, total = 133 Mb). Of these regions, we visually

198 identified 46 that showed three clear clusters on a first PC explaining most of the variance
199 (e.g., Fig. 5b). We interpret these regions as putative haplotype blocks, or haploblocks,
200 i.e. regions of low recombination where polymorphic sites are inherited in high linkage,
201 creating groups of haplotypes, or haplogroups. As such, the three observed clusters
202 correspond to the two homozygotes and the heterozygote for these haplogroups. Using
203 populational-scale recombination maps based on 8X data (Fig. S8), we confirmed that the
204 putative haploblocks tended to have lower recombination rates than the chromosome-
205 wide average (Fig. S9), further supporting this interpretation.

206 Most of these putative haploblocks corresponded to islands of differentiation
207 between the Arctic and Atlantic lineages. Twenty-three haploblocks (median length = 783
208 kb, total = 20 Mb) were differentially fixed between the north-western Canadian Arctic
209 (Inuvialuit, Kitikmeot; henceforth identified as haplogroup 'A') and Sweden (haplogroup
210 'B') while showing heterozygotes in southern populations of the Canadian Arctic (e.g., Fig.
211 5c), suggesting that they may be tracts of different ancestry that are yet unbroken by
212 recombination. The individual distribution of heterozygotes differed between haploblocks,
213 supporting their independent genealogy, but haplogroup frequencies by population were
214 correlated with the Arctic-Atlantic whole-genome ancestry inferred earlier (NGSadmix, K
215 = 2). Such correlations further support a scenario of secondary contact and introgression
216 at the nuclear level by revealing blocks of linked polymorphic markers that appear to be
217 inherited from ancestral lineages, after their divergence in allopatry during past glacial
218 periods.

219 Interestingly, two other MDS outliers (e.g., win21 and win58 in Fig. S7) also
220 featured three distinct genotypes on PC1, but their spatial distribution did not match the
221 Arctic-Atlantic admixture gradient. Those could be associated with recent and polymorphic
222 structural variants suppressing recombination such as chromosomal inversions, allowing
223 the accumulation of genetic divergence in sympatry.

224 *Inference of origin of recolonization*

225 Directionality analysis conducted on either the whole genome or restricted to variation
226 inside Arctic haplotypes supported conflicting recolonization histories, respectively from
227 the east and west of the lineage range. Based on genome-wide data, we inferred an origin
228 of recolonization around the southeast of the Canadian Arctic by comparing the spatial
229 distribution of genome-wide derived alleles in Canadian populations through the
230 directionality index (ψ) and Time Difference of Arrival (TDoA) (Fig. 6a). This is
231 accomplished by testing every potential origin (cells on a spatial raster) for a positive
232 correlation between population-pairwise ψ (i.e., difference in derived allele frequencies)
233 and geographical distance from the origin (maximum $R^2 = 0.638$ before rescaling). This
234 inference is nevertheless to be taken with caution because Canadian populations deviate
235 from the single-origin assumption of TDoA analysis and are precisely characterized by
236 extensive introgression in the southeast.

237 To confirm the directionality analysis on a putatively non-introgressed fraction of
238 the genome, we leveraged the previously identified local ancestry tracts to compute ψ on
239 derived allele frequencies in homozygotes for the Arctic lineage-specific alleles (e.g.
240 haplogroup A in Fig. 5) of these local ancestry tracts (n = 23, total length = 20 Mb). Despite

241 a reduced genomic size (1.3 % of the total genome), the resulting 100 replicate runs of
242 TDoA weakly supported an origin of recolonization in the western part of the sampling
243 area (maximum R^2 ranging from 0.007 to 0.107, Fig 6b, Fig S10a), contrary to the eastern
244 origin of recolonization inferred using genome-wide data.

245 To control for the small number of non-recombining haploblocks and the difficulty
246 in precisely delimiting these putative local ancestry tracts, we ran the same analyses on
247 random genomic regions of equivalent lengths. These analyses supported a south-
248 eastern origin of recolonization, which was highly congruent with the whole-genome TDoA
249 (maximum R^2: 0.199 – 0.483, Fig. 6b, Fig. S10b). This suggests that, although our current
250 prediction power is low, genetic diversity inside local ancestry tracts could carry distinct
251 lineage-specific information that can be leveraged to elucidate more complex evolutionary
252 histories.

253

254 Discussion

255 It has long been recognized that making phylogeographic inferences based on
256 mitochondria alone can lead to biased interpretations (Galtier et al. 2009). However, the
257 non-recombining nature of mtDNA can offer advantages in specific cases since it can
258 'record' past events that are not blurred by generations of hybridization and admixture. At
259 the same time, vast amounts of nuclear data are sometimes insufficient in revealing
260 historical patterns if averaged over the genome, and evolutionary inferences could benefit
261 from identifying haploblocks that retain historical signals (Leitwein et al. 2020; Shipilina et

262 al. 2023). In this study, we sought to maximize the use of whole-genome sequencing data
263 by integrating mitochondrial and nuclear analyses, both at the genome-wide and
264 haplotypic levels, to assess the extent of admixture between two glacial lineages of an
265 arctic fish. Intraspecific hybridization between glacial lineages and the resulting increased
266 nuclear diversity observed in southern populations was a dominant feature of the study
267 system. Such introgression across lineages obscured much of the genomic signal for
268 post-glacial recolonization from either single lineage. We scanned the genome for regions
269 of distinct genetic variation and identified putative local tracts of conserved Arctic or
270 Atlantic ancestry. By comparing genetic variation within Arctic tracts to the whole genome,
271 we uncovered a putative cryptic glacial refugium for the lineage, despite a heavily admixed
272 genetic background.

273 *Extensive admixture between lineages and mito-nuclear discordance*

274 In Arctic Char, we show evidence for the genetic admixture of the Arctic and Atlantic
275 lineages in most of Arctic Canada under the 66th parallel. This is in stark contrast with the
276 previously described distribution of mitochondrial haplotypes, with observed co-
277 occurrence zones of both lineages restricted to Labrador (Salisbury et al. 2019) and
278 western Greenland (Jacobsen et al. 2022). While previous phylogeographic work on the
279 species hinted at the introgression from the Atlantic lineage reaching further west at the
280 nuclear level (Moore et al. 2015; Dallaire et al. 2021), here we inferred the presence of
281 Atlantic nuclear alleles thousands of kilometers away from the range limit of the Atlantic
282 mitochondrial lineage, both using f_4 tests and indirect detection of lineage-divergent
283 haploblocks.

284 This kind of mito-nuclear discordance has been observed in numerous natural
285 systems (Toews and Brelsford 2012) and could be explained by multiple mechanisms.
286 First, as non-recombinant haploid sequences, mitogenomes have a population effective
287 size (Ne) that is ¼ of their diploid nuclear counterparts (Hudson and Turelli 2003) and are
288 thus more impacted by genetic drift. Alternatively, sex-biased dispersion has been
289 suggested in Arctic Char with males appearing more mobile than females (Dempson and
290 Kristofferson 1987; Moore et al. 2016). As mitochondrial genomes are matrilineally
291 inherited, this could contribute to a discordance in the distribution of nuclear DNA and
292 mitochondrial DNA variation.

293 *Local PCA reveals more than just inversions*

294 Short-read whole-genome sequencing has become increasingly accessible in non-
295 model organisms. New analytical frameworks allow for the study of population genetics
296 questions even with low individual depth of coverage, which has opened the door for
297 whole-genome studies of unprecedented sampling sizes (Lou et al. 2021). However,
298 phasing (i.e. haplotype estimation) might be impossible in these designs without a
299 reference panel (Rubinacci et al. 2021), as haploblocks are most reliably defined and
300 delimited using long- or linked-read data (Meier et al. 2021). It has been suggested that,
301 when they are large and divergent enough (e.g., in regions of low recombination),
302 haploblocks might also be indirectly detectable through genome scans for distinct patterns
303 of genetic variation (Ishigohoka et al. 2024). One such method is the use of local PCAs,
304 an approach that has risen as a popular approach to detect large recombination-
305 suppressing structural variants (e.g., inversions) (Huang et al. 2020). The typical signal

306 expected around non-recombining haploblocks is a PCA forming three distinct clusters on
307 the first axes (generally PC1), with the middle cluster carrying more heterozygous sites
308 than the two others, and higher linkage disequilibrium among clusters than inside each
309 one (Huang et al. 2020).

310 In this study, we identified local PCA outliers by performing multidimensional
311 scaling (MDS) at the chromosomal level that allowed for the detection of relatively short
312 (250 kb to 4 Mb) candidate regions on every chromosome. We analyzed the five first MDS
313 axes, but Fig. S7 suggests that many additional outlier regions could be identified on
314 subsequent axes. PC1 shows the expected three-cluster pattern indicative of haploblocks,
315 i.e. the two homozygotes and the heterozygote for a region of low recombination, and
316 inversions may underlie some of those haploblocks. For example, one of our MDS outlier
317 regions indicative of a haploblock (win16) was located on the distal end of LG12, which
318 was recently identified as a putative inversion in Nunavut Arctic Char (Hale et al. 2021).
319 However, we observed that the clustering in three groups was far from perfect in most of
320 the haploblocks, with many intermediate individuals (e.g. Fig. 5B). This suggests a
321 suppression of recombination less severe than what has been observed in recent
322 publications investigating inversions (Huang et al. 2020; Mérot et al. 2021). Intriguingly,
323 some candidate haploblocks (including win16) showed additional structure inside an
324 inferred homozygote group on the second axis of the PCA, resulting in a triangular pattern
325 when considering PC1 and 2 (e.g., win50, win89, win91, and win109 in Fig. S7), which
326 may result from nested inversions or regions of low recombination maintaining more than
327 two distinct groups of haplotypes. Therefore, overall, we proposed (and confirmed for

328 some) that those regions are characterized by low recombination but do not further
329 speculate with the present data about the precise mechanism protecting the observed
330 haploblocks from recombination.

331 More interesting is the coherent spatial pattern observed in multiple independent
332 genomic loci, with the two alternative haplogroups being differentially fixed in the Western
333 Canadian Arctic and Sweden. As we expect genetic variation in these two regions to be
334 respectively associated with the Arctic and Atlantic lineages, this local PCA approach thus
335 reveals putative local ancestry tracts for either lineage in admixed individuals, which were
336 found mostly in the Southern Canadian Arctic and Western Greenland. This constitutes a
337 rather indirect method compared to phasing algorithms based on the Li and Stephens
338 model (Li and Stephens 2003) which requires a reference panel (e.g. HAPMIX (Price et
339 al. 2009) and FLARE (Browning et al. 2023)). However, no solution currently exists for the
340 accurate phasing of low-coverage unlinked short-read sequencing data. By exploring how
341 local PCAs highlight local genetic patterns coherent with haplotype structure, we hope to
342 spur interest in improving how lcWGS can be used to their fullest.

343 Here, we only scratched the surface of the haplotypic structure and likely had the
344 power to detect only the longest and most geographically structured haploblocks. A quick
345 look at the structure outside MDS outlier regions (Fig. 5a, Fig S8) shows how most of the
346 genome segregates similarly at a finer scale, i.e. in smaller blocks. The delimitation of
347 haploblocks could likely be refined by tuning the window size and filtering parameters of
348 the MDS outlier detection. However, some of the imprecision might be linked to reasons
349 both technical (e.g., the uncertainty inherent to the analytical framework of low-coverage

350 data) and biological (the heterogeneity of recombination along the genome and among
351 populations, breaking ancestry tracks into regions of varying lengths). In the next section,
352 we discuss one possible application for such haplotypic information in datasets of large
353 sample sizes.

354 *Leveraging local ancestry to disentangle multiple origins of recolonization*

355 In cases of post-glacial secondary contact, pinpointing the origin of recolonization – i.e.
356 the position of the glacial refugium – of each lineage can help provide additional context,
357 as the process of recolonization itself and its direction can have predictable impacts on
358 within-lineage genetic variation (Hewitt 2000; Peischl et al. 2015). One such impact is
359 increased genetic drift through sequential founder effects along the recolonization front
360 (Hallatschek et al. 2007; Slatkin and Excoffier 2012; Braga et al. 2019), leading to a
361 decrease in genetic diversity and an accumulation of derived and sometimes deleterious
362 alleles via increased genetic drift (Koski et al. 2019; Rougemont et al. 2020; Rougemont
363 et al. 2023). The directionality index (ψ) and TDoA approach (Peter and Slatkin 2013)
364 uses this prediction to triangulate the origin of recent range expansions by comparing the
365 spatial distribution of derived alleles (e.g. (Prior et al. 2020; Ioannidis et al. 2021; Luqman
366 et al. 2023)).

367 In our data, applying this method to the genome-wide variation pointed toward an
368 origin in the south-eastern Canadian Arctic. However, the study system shows clear signs
369 of admixture between lineages of at least two distinct origins, as is often the case in similar
370 studies (Sommer and Zachos 2009; Nugent et al. 2024). This violates a major assumption
371 of the model (Peter and Slatkin 2013), as the high divergence between lineages most

372 likely overshadows any within-lineage signal of directionality. The Arctic lineage is
373 expected to have recolonized from a smaller source population than other lineages
374 (Moore et al. 2015), suggesting a demographic bottleneck that could explain the higher
375 frequency of derived alleles in the northern, non-admixed populations. With the goal of
376 revealing patterns specific to the Arctic lineage, we focused on genetic variation inside
377 putative local ancestry tracts identified earlier. By considering only homozygotes for the
378 haplogroup that we attributed to the Arctic lineage, we hypothesized that we could limit
379 the influence of the introgression from the Atlantic lineage and investigate the origin of
380 recolonization of the Arctic lineage. While these inferences had lower explanatory power
381 and replicability, they globally pointed in a direction opposite to the genome-wide analysis,
382 highlighting the interest in dissecting horizontal information present in genomes and
383 further refinement of local ancestry tracts.

384 Inferring range expansions from the distribution of derived alleles has several
385 limitations, as other evolutionary mechanisms can create asymmetry in two-dimensional
386 site-frequency spectra, such as demographic changes unrelated to the expansion (Mestre
387 et al. 2022). Boundary effects can also act as confounding factors, as populations on the
388 outer part of the distribution will experience increased drift compared to central ones,
389 which can create clines in ψ independently from range expansions (Kemppainen et al.
390 2024). In light of this and in the absence of rigorous hypothesis testing, our key takeaway
391 focuses less on our confidence in the inference for this specific system and more on the
392 potential of genetic information preserved in large blocks of known local ancestry to
393 unravel complex, multi-layered evolutionary histories.

394 Despite previously discussed limitations, other sources of data were consistent with
395 a post-glacial recolonization from the northwestern Canadian Arctic, such as
396 mitochondrial diversity being higher in this region than elsewhere in the range of the Arctic
397 lineage. While species are intuitively expected to have survived glacial periods in refugia
398 south of the ice extent, fossil and genetic evidence suggests that higher-latitude refugia
399 might have existed (Sommer and Zachos 2009; Premoli et al. 2010; Shafer et al. 2010;
400 Wójcik et al. 2010; Mee and Moore 2014). Although recent reconstructions suggest a
401 complete glaciation of the Arctic archipelago at the last maximum (Fig. 1) (Dalton et al.
402 2020), data from rodents (Fedorov and Stenseth 2002), birds (Holder et al. 1999), and
403 plants (Tremblay and Schoen 1999; Abbott et al. 2000) point to cryptic refugia in the High
404 Arctic, and Banks Island (Inuvialuit, North West Territories) has been proposed as a
405 refugium for caribou (Klütsch et al. 2017). Ultimately, the use of local ancestry as a lens
406 through which to study recolonization could offer a promising avenue for disentangling
407 complex post-glacial evolutionary histories of the High Arctic and elsewhere.

408 *Conclusion*

409 Until recently, applications of whole-genome sequencing (WGS) to population genomics
410 have been limited by high costs, resulting in low sample sizes and restricted study areas.
411 Here, our use of low-coverage WGS data showcases how versatile this approach can be
412 in investigating intraspecific evolutionary processes for widely sampled organisms. The
413 inclusion of cytoplasmic DNA (from mitochondria or chloroplasts in plants) in WGS also
414 promises a more systematic comparison of markers from different independently evolving
415 segments of the genome, providing new phylogeographic insight while integrating earlier

416 research in long-studied systems. Finally, WGS data offers far more than an increased
417 SNP count, and we are confident that the scientific community will benefit from exploring
418 innovative approaches for leveraging horizontal genetic signals, like linkage and
419 recombination, to deepen our understanding of evolutionary dynamics.

420 Material & methods

421 *Library preparation, sequencing, and data preprocessing*

422 Anadromous Arctic Char adipose fins were collected at the mouth of rivers across Arctic
423 Canada and Western Greenland between 2007 and 2023 and stored in ethanol 95% (4°C)
424 or RNAlater (-20°C). Whole-genome libraries were prepared following Therkildsen &
425 Palumbi (2017). Briefly, DNA was extracted using Nucleomag kits, cleaned with Axygen
426 magnetic beads, quantified with AccuClear Ultra High Sensitivity dsDNA Quantification
427 kits, and diluted to 1.5 ng/μl. Sample extracts were then used for Nextera library
428 preparation: tagmentation (i.e. fragmentation and addition of adapters), 2 PCR steps to
429 attach barcodes, and size selection for fragments 400-700bp in length using Axygen
430 magnetic beads. To achieve a final depth of coverage of 2X, we performed sequencing
431 on lanes of Illumina HiSeq 4000, with around 96 individual libraries per lane (range of 75
432 to 106), yielding on average 31.6 million reads per sample for 1178 samples in 33
433 populations. For the medium-coverage dataset, we selected 100 random samples in 5
434 populations (20 per population) and repeated sequencing on 5 lanes to obtain 135.9
435 million reads per sample (adding ~6X).

436 All data were prepared following the pipeline described at
437 https://github.com/enormandeau/wgs_sample_preparation. In brief, raw sequences were

438 trimmed using fastp (Chen et al. 2018) and aligned on the ASM291031v2 reference
439 genome with bwa mem (minimum alignment quality of 10) (Li 2013). Note that the
440 reference genome used was assembled using sequencing data from a Dolly Varden
441 (*Salvelinus malma*) or a *S. alpinus* x *S. malma* hybrid (Christensen et al. 2021) but remains
442 the closest high-quality reference for Arctic Char at the time of our analysis. Duplicate
443 reads were then removed using picard, indels were realigned, and overlapping ends of
444 paired reads were clipped. The average per-base depth of coverage was then estimated
445 using mosdepth (Pedersen and Quinlan 2018), and 61 samples with genome-wide
446 coverage under 1X were excluded from analyses on nuclear DNA. For the 5 populations
447 targeted by additional sequencing, the 2X and 6X aligned reads were merged so that
448 these populations could be analyzed at a coverage of either 2 or 8X.

449 *mtDNA and nuDNA genotyping and filtering*

450 The depth of coverage was much higher on the mitochondrial genome, allowing us to call
451 haplotypes for the whole mitogenome (16,659 bp) for every individual in the 2X dataset
452 using angsd -doFasta 1 (Korneliussen et al. 2014). Sequences were cleaned in Geneious
453 2022.0.1 (https://www.geneious.com) by removing positions, then individuals, with more
454 than 5% missing data. We used popart (Leigh and Bryant 2015) to construct a median-
455 joining network of haplotypes extracted for the D-loop control region (998 bp), using an ε
456 parameter of zero. An individual sampled in population 30 matched the D-loop sequence
457 for *Salvelinus fontinalis* and was kept as an outgroup for the haplotype network but was
458 removed from every subsequent analysis.

459 Nuclear SNP positions were first detected using ANGSD v0.937 (Korneliussen et
460 al. 2014) (`-GL 2 -remove_bads 1 -minMapQ 30 -minQ 20 -skipTriallelic`
461 `1 -uniqueOnly 1 -only_proper_pairs 1 -min_Maf = 0.01`) on both the 2X
462 and 8X datasets independently with the following coverage filters (for 2X data: at least 1X
463 in 75% of samples, maximum average depth of 4X; for 8X data: at least 3X in 75% of
464 samples, maximum average depth of 20X). The sex of sampled fish was unknown. As
465 such, SNPs on sex-linked regions identified by Beemelmanns et al. (2024) were not
466 considered, to avoid potentially unbalanced sex ratios being confounded with population
467 differentiation.

468 The 15.6 (2X) and 21.5 (8X) million retained SNP positions were tested for their
469 deviation of Mendelian inheritance using the calcLR in ngsParalog (Linderoth 2018).
470 ngsParalog likelihood ratios were compared to a χ^2 distribution and SNPs with a P-value
471 (adjusted by applying the Bonferroni procedure) under the threshold of 0.001 were
472 considered as deviant (2X: 9.63 million, 62% of all SNPs; 8X: 11.20 million, 52%). Deviant
473 SNPs were removed from the site list for each dataset, as they are expected to have their
474 inferred genotype biased by mismapped reads related to paralogy or repeated elements
475 (Dallaire et al. 2023). For analyses based on site frequency spectra (SFS), we masked
476 the reference genome in 150bp regions centered on every deviant SNP, as well as all
477 coding genes and a 1-kb region around them.

478 We estimated genotype likelihoods for 5.98 million non-deviant SNPs in the 2X
479 dataset using `-doGlf -GL2` in ANGSD. We used ngsLD to estimate linkage between
480 SNPs within 500 kb using a random subset of half the samples and pruned the dataset

481 with the graph method until no SNP pairs within 200kb had an r^2 above 0.1, leaving
482 375,236 SNPs in the LD-pruned dataset.

483 *Population structure, differentiation, and diversity*

484 The general distribution of genetic variation and population structure was first assessed
485 by performing a principal components analysis on the LD-pruned 2X dataset in PCAngsd
486 1.10 (Meisner and Albrechtsen 2018). We then explored the hierarchical population
487 structure using NGSadmix (Skotte et al. 2013) on the LD-pruned 2X data to estimate
488 individual admixture while varying the number of ancestral populations (K) from 2 to 33.

489 Pairwise population differentiation was estimated using allele frequency difference
490 (AFD) and F_{ST} index. For AFD, we estimated allele frequency in every population for all
491 5.98 million SNPs (2X) using angsd -domaf 1, then calculated the average absolute MAF
492 difference between each pair of populations. For F_{ST}, we used ANGSD and the reference
493 genome masked for deviant SNPs and genes to estimate sample allele frequency (SAF)
494 for each population, then we used the winSFS streaming option (Rasmussen et al. 2022)
495 to compute two-population SFS while accounting for linkage disequilibrium. The SAF and
496 2D-SFS served as input for the *realsfs fst index*, *print*, *stats,* and *stats2* functions in
497 ANGSD to calculate genome-wide weighted F_{ST} (Reynolds et al. 1983). We tested for the
498 presence of isolation-by-distance through linear regression between the marine distance
499 between population pairs, estimated using marmap (Pante and Simon-Bouhet 2013), and
500 either their AFD or linearized F_{ST} (Slatkin 1993).

501 We estimated nuclear genetic diversity in populations with Watterson's estimator
502 (Θ_W) and nucleotide diversity (Θ_π), as well as individual heterozygosity. Θ statistics were

503 computed with ANGSD's *realsfs saf2theta* and *thetaStat do_stat* using the SAF by
504 population described above and a 1D-SFS produced by winsfs. To account for the variable
505 coverage along the genome and across populations, we compared per-site Θ in windows
506 of 100 kb (step of 10 kb) and excluded windows with a number of sites (denominator)
507 under the 5th percentile (2,600 sites) in any population. Individual heterozygosity was
508 estimated by computing a per-sample SAF and dividing the number of heterozygous sites
509 (SAF = 1) by the number of sites.

510 We also estimated diversity metrics on whole mitogenome sequences, after removing
511 haplotypes from the Atlantic lineage, those with more than 5% of missing data, and
512 masking positions with remaining missing values. The 1,074 Arctic haplotypes were
513 16,483 bp long and were analyzed by population with the *pegas* (Paradis 2010) package
514 in R for 1) Θ_W with *theta.s* function, 2) Θ_π with the *nuc.div* function, 3) Tajima's D with the
515 Tajima.test function, 4) haplotype diversity using the *haploFreq* function and the formula
516 corrected for sample size from (Nei and Roychoudhury 1974), and 5) the frequency of the
517 Arctic haplotype most frequent in the whole dataset. We tested for geographical patterns
518 in populational mtDNA diversity using regression models for those 5 variables with the
519 marine distance from the westernmost population (pop.1) as the explicative variable. We
520 used the Akaike Information Criterion (AIC) (Akaike 1974) to compare the null, simple,
521 and 2nd-degree polynomial models.

522 *Secondary contact, introgression, and demographic inference*

523 To test hypotheses regarding the admixture between the Arctic and Atlantic lineage of
524 Arctic Char in Canada, we downloaded *S. alpinus* raw sequences (PRJEB62707) from an

525 inland lake in Sweden (Lilla Stensjön) (Saha et al. 2024), which we assumed represents
526 a population descending purely from the Atlantic lineage. We also acquired unpublished
527 data from a Dolly Varden (*Salvelinus malma*) population in Babage River, Yukon, Canada
528 (Labrecque et al., in prep). Those sequences were trimmed, aligned, and cleaned as
529 described in the "data preprocessing" section, then we estimated genotype likelihoods
530 and population MAF for the LD-pruned SNP list. We obtained an average depth of
531 coverage of 5X for the Swedish population and 2X for *S. malma*.

532 We repeated the NGSadmix analysis for K=2 with the Swedish population. We
533 converted the estimated MAF of all 35 populations of *S. alpinus* and *S. malma* to
534 approximate allele count by multiplying the MAF by the number of sampled alleles (twice
535 the sample size) to fit the Treemix file format. We computed f_4 statistics in Treemix 1.13
536 (Pickrell and Pritchard 2012) with *S. malma* as the outgroup (H_4), the Swedish population
537 as the potential introgressor (H_3), and all pairs of Canadian and Greenlandic populations
538 as H_1 and H_2.

539 Based on the previous results, we build alternative demographic models to test the
540 presence of ancient or recent gene flow between the Atlantic lineage, represented by the
541 Swedish population and the Southern Arctic Canada (see Fig. 4B, Table S). We estimated
542 the unfolded 3D-SFS between COP (North, pop. 4), TJL (South, pop. 25), and LLS
543 (Sweden, pop. 34) using winsfs. Alleles were polarized based on an ancestral reference
544 genome constructed by aligning WGS data from four closely related species (Atlantic
545 Salmon; Lake Trout, *Salvelinus namaycush*; Rainbow Trout, *Oncorhynchus mykiss*; and
546 Chinook Salmon, *Oncorhynchus tshawytscha*; SRA accession number in supplementary

547 table S6, Supplementary Material online) on the *Salvelinus sp.* reference genome
548 (ASM291031v2). We used the most common allele to create a consensus for each
549 species using *angsd -doFasta*, then merged the consensus in Geneious 2022.0.1 to keep
550 only positions covered in at least 2 consensuses. The fit between the observed SFS and
551 models with 0) no, 1) ancient, and 2) recent gene flow was compared by running 100 runs
552 of each in fastsimcoal 2.7 with the following parameters: `-d -C 10 -n 500000 -L 50`
553 `-s 0 -M`. [Parameter estimation]

554 *Introgression and recombination landscapes*

555 We identified genomic regions displaying distinct genetic variation from the genome-wide
556 signal using local PCAs as described in (Huang et al. 2020; Mérot et al. 2021). Briefly, we
557 scanned each chromosome independently by computing the covariance matrix between
558 individuals in non-overlapping windows of 100 SNPs in PCAngsd and detecting clusters
559 of consecutive windows presenting similar PCA patterns using the *pcdist* function and
560 multidimensional scaling (MDS) in the R package *lostruct*. Windows with MDS values
561 beyond 3 standard deviations from the chromosomal mean on one of the first five MDS
562 axes were kept, and clusters within 20 windows of each other were merged. We only
563 considered clusters longer than 10 windows.

564 In each outlier cluster, we tested for the presence of patterns on PC1 indicative of the
565 three genotypes of a chromosomal inversion or other types of nonrecombinant
566 haploblock. The R package *dbscan* allowed for the grouping of individuals based on their
567 density distribution on PC1 while allowing for uncategorized individuals between clusters.
568 We iteratively ran *dbscan* while increasing the ε parameter by increments of 0.005,

569 starting from 0.01, until three groups were detected or ε = 0.15 was reached. We further
570 cleaned the *dbscan* results by setting individuals with PC1 values outside 3 standard
571 deviations from the mean of any group as intermediate genotype. The three remaining
572 groups were labeled haplogroup AA, AB, and BB, starting from the most frequent one,
573 and the intermediate genotypes were labelled A0 and 0B.

574 To test if the putative haploblocks were associated to lower recombination rates, we
575 estimated recombination using LDhat for each chromosome and population in the 8X
576 dataset. To produce the bcf file in input, we genotyped previously identified non-deviant
577 SNPs in ANGSD using `-doBcf 1 -doGeno -4 -doPost 1 -postCutoff 0.95`
578 with more stringent filters (population MAF > 0.05, minimum of 4X in 90% of individuals).
579 Population-scaled recombination rates (ρ/kb) was estimated in LDhat 2.2 (Auton and
580 McVean 2007) while controlling for demography with LDpop (Kamm et al. 2016) using a
581 θ estimate of 0.001, a block penalty of 5 and other parameters set following the pipeline
582 in (Brazier and Glémin 2024). The LDhat output was cleaned by removing plateaus of
583 invariable recombining rates longer than 22 kb (99th percentile of the length of all
584 plateaus). The weighted mean of recombination rates inside each putative haploblocks
585 and other MDS outliers was compared to the chromosome-wide average through mixed-
586 effect models with the chromosome and population set as random effects. Chromosomes
587 and haploblocks were only considered in recombination maps that covered at least 60%
588 of their length.

589 *Inference of the origin of recolonization*

590 To triangulate the origin of the recolonization after the last glacial maximum, and thus a
591 putative glacial refugium, we computed the directionality index (ψ) for each population pair
592 from their genome-wide unfolded 2D-SFS, following Peter & Slatkin (2013). Ancestral
593 alleles private to either population were excluded. We then used the Time Difference of
594 Arrival algorithm (Gustafsson and Gunnarsson 2003) on the ψ matrix following Prior et al.
595 (2020). Briefly, a pairwise comparison of ψ and geographical distance between focal
596 populations and the assumed origin was performed for every possible origin, using a 0.5
597 decimal degrees step across the study area. For origins with a positive slope between ψ
598 and geographic distance, we rescaled the coefficient of regression (R^2) from 0 to 1, and
599 origins with a higher R^2 were considered more likely. For this analysis, we excluded
600 populations out of Canada, as well as population 30, which had a high level of Atlantic
601 lineage ancestry.

602 To test how introgression from the Atlantic lineage impacted the inference of the
603 origin of recolonization, we repeated the *ψ* and TDoA analyses while focusing on the
604 diversity inside the Arctic alleles for the putative haploblocks identified using local PCAs.
605 For each population pair and haploblock, we randomly selected 10 homozygotes for Arctic
606 allele (AA) by population and computed their SAFs and 2D-SFS. We then summed the
607 per-haploblock SFS where both populations had at least 10 AA individuals and computed
608 a directionality index from the summed spectrum. TDoA was performed on the resulting
609 ψ matrix as described above.

610 To assess the robustness of this analysis, we repeated this process over 100
611 replicates on resampled AA individuals and a random subset of 90% of putative
612 haploblocks, as well as over 100 replicates where each haploblock was replaced by a
613 random genomic region of equal length on the same chromosome. TDoA R^2 maps from
614 the putative haploblocks and random regions were each rescaled as above, then
615 averaged and compared to the genome-wide TDoA signal.

616

Data Accessibility Statement

Part of the *Salvelinus alpinus* sequencing data analyzed in this manuscript is available on Short Read Archive as part of projects PRJNA1031558 (Canada and Greenalnd) and PRJEB62707 (Sweden). Following submission, all new genetic data will be deposited on SRA. Sequencing data on *Salvelinus malma* is also planned to be deposited as part of a separate publication, presently in preparation.

Acknowledgment

This work was supported by a Large-Scale Applied Research Project grant from Genome Canada named “FISHES: Fostering Small-scale Fisheries for Health, Economy and Food Security”. Sampling was made possible by the collaboration of Fisheries and Ocean Canada; Ministère de l’Environnement, de la Lutte contre les changements climatiques, de la Faune et des Parcs (Québec; Julien Mainguy); Government of Nunavut; Makivik Corporation (Nunavik) and numerous Inuit communities, organizations, and local fishers across Canada and Greenland. Thanks to Anne Beemelmanns, Charles Babin, Isabeau Caza-Allard, Louis-Philippe Collin, Alysse Perreault-Payette, and Gabriel Piette-Lauzière for their help with laboratory work and coordination. We also thank Rennes Metropole, ECOBIO (Université de Rennes), and ISEM (Université de Montpellier) for providing support, funding, or access to their facility which were invaluable in the completion of this work.

References

Abbott RJ, Smith LC, Milne RI, Crawford RMM, Wolff K, Balfour J. 2000. Molecular Analysis of Plant Migration and Refugia in the Arctic. *Science* [Internet] 289:1343–1346. Available from: https://www.jstor.org/stable/3077629

Akaike H. 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control* [Internet] 19:716–723. Available from: https://ieeexplore.ieee.org/document/1100705

Auton A, McVean G. 2007. Recombination rate estimation in the presence of hotspots. *Genome Res.* [Internet] 17:1219–1227. Available from: https://genome.cshlp.org/content/17/8/1219

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC. 1987. Intraspecific Phylogeography: The Mitochondrial DNA Bridge Between Population Genetics and Systematics. *Annu. Rev. Ecol. Syst.* 18:489–522.

Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. *Mol. Ecol.* [Internet] 13:729–744. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-294X.2003.02063.x

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

Bhatia G, Patterson N, Sankararaman S, Price AL. 2013. Estimating and interpreting F_{ST}: The impact of rare variants. *Genome Res.* [Internet] 23:1514–1521. Available from: http://genome.cshlp.org/lookup/doi/10.1101/gr.154831.113

Braga RT, Rodrigues JFM, Diniz-Filho J a. F, Rangel TF. 2019. Genetic Population Structure and Allele Surfing During Range Expansion in Dynamic Habitats. *An. Acad. Bras. Ciênc.* [Internet] 91:e20180179. Available from: https://www.scielo.br/j/aabc/a/bjgBZQhBLJh5MFQNfC8v9Qh/?lang=en

Brazier T, Glémin S. 2024. Diversity in Recombination Hotspot Characteristics and Gene Structure Shape Fine-Scale Recombination Patterns in Plant Genomes. *Mol. Biol. Evol.* 41.

Browning SR, Waples RK, Browning BL. 2023. Fast, accurate local ancestry inference with FLARE. *Am. J. Hum. Genet.* [Internet] 110:326–335. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9943733/

Brunner PC, Douglas MR, Osinov A, Wilson CC, Bernatchez L. 2001. Holarctic Phylogeography of Arctic Charr (Salvelinus Alpinus L.) Inferred From Mitochondrial Dna Sequences. *Evolution* [Internet] 55:573–586. Available from: http://doi.wiley.com/10.1111/j.0014-3820.2001.tb00790.x

Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* [Internet] 34:i884–i890. Available from: https://academic.oup.com/bioinformatics/article/34/17/i884/5093234

Christensen KA, Rondeau EB, Minkley DR, Leong JS, Nugent CM, Danzmann RG, Ferguson MM, Stadnik A, Devlin RH, Muzzerall R, et al. 2021. Retraction: The Arctic charr (Salvelinus alpinus) genome and transcriptome assembly. *PLOS ONE* [Internet] 16:e0247083. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0247083

Dallaire X, Bouchard R, Hénault P, Ulmo-Diaz G, Normandeau E, Mérot C, Bernatchez L, Moore J-S. 2023. Widespread Deviant Patterns of Heterozygosity in Whole-Genome Sequencing Due to Autopolyploidy, Repeated Elements, and Duplication. *Genome Biol. Evol.* [Internet] 15:evad229. Available from: https://doi.org/10.1093/gbe/evad229

Dallaire X, Normandeau É, Mainguy J, Tremblay J-É, Bernatchez L, Moore J-S. 2021. Genomic data support management of anadromous Arctic Char fisheries in Nunavik by highlighting neutral and putatively adaptive genetic variation. *Evol. Appl.* [Internet] n/a. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/eva.13248

Dalton AS, Margold M, Stokes CR, Tarasov L, Dyke AS, Adams RS, Allard S, Arends HE, Atkinson N, Attig JW, et al. 2020. An updated radiocarbon-based ice margin chronology for the last deglaciation of the North American Ice Sheet Complex. *Quat. Sci. Rev.* 234.

Dempson JB, Kristofferson AH. 1987. Spatial and Temporal Aspects of the Ocean Migration of Anadromous Arctic Char. *Am. Fish. Soc. Symp.* 1:340–357.

Edwards SV, Shultz AJ, Campbell-Staton SC. 2015. Next-generation sequencing and the expanding domain of phylogeography. *Folia Zool.* [Internet] 64:187–206. Available from: http://www.bioone.org/doi/10.25225/fozo.v64.i3.a2.2015

Fedorov, Stenseth. 2002. Multiple glacial refugia in the North American Arctic: inference from phylogeography of the collared lemming (Dicrostonyx groenlandicus). Available from: https://royalsocietypublishing-org.acces.bibl.ulaval.ca/doi/epdf/10.1098/rspb.2002.2126

Galtier N, Nabholz B, Glémin S, Hurst GDD. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Mol. Ecol.* [Internet] 18:4541–4550. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-294X.2009.04380.x

Gordeeva NV, Alekseyev SS, Kirillov AF, Romanov VI, Pichugin MYu. 2021. New Data about the Distribution of Three Phylogenetic Lineages of Arctic Charr Salvelinus alpinus (Salmonidae) in their Contact Zones in the North of East Siberia. *J. Ichthyol.*

[Internet] 61:701–708. Available from: https://link.springer.com/10.1134/S0032945221050064

Gustafsson F, Gunnarsson F. 2003. Positioning using time-difference of arrival measurements. In: 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). Vol. 6. p. VI–553.

Hale MC, Campbell MA, McKinney GJ. 2021. A candidate chromosome inversion in Arctic charr (*Salvelinus alpinus*) identified by population genetic analysis techniques.Macqueen D, editor. *G3 GenesGenomesGenetics* [Internet]:jkab267. Available from: https://academic.oup.com/g3journal/advance-article/doi/10.1093/g3journal/jkab267/6329827

Hallatschek O, Hersen P, Ramanathan S, Nelson DR. 2007. Genetic drift at expanding frontiers promotes gene segregation. *Proc. Natl. Acad. Sci. U. S. A.* 104:19926–19930.

Hewitt GM. 1996. Some genetic consequences of ice ages, and thier role in divergence. *Biol. J. Linn. Soc.* 58:247–276.

Hewitt GM. 1999. Post-glacial re-colonization of European biota. *Biol. J. Linn. Soc.* [Internet] 68:87–112. Available from: https://doi.org/10.1111/j.1095-8312.1999.tb01160.x

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

Hocutt CH, Wiley EO. 1986. The Zoogeography of North American freshwater fishes. New York: Wiley Available from: https://bac-lac.on.worldcat.org/oclc/299840901

Holder K, Montgomerie R, Friesen VL. 1999. A TEST OF THE GLACIAL REFUGIUM HYPOTHESIS USING PATTERNS OF MITOCHONDRIAL AND NUCLEAR DNA SEQUENCE VARIATION IN ROCK PTARMIGAN (LAGOPUS MUTUS). *Evol. Int. J. Org. Evol.* 53:1936–1950.

Huang K, Andrew RL, Owens GL, Ostevik KL, Rieseberg LH. 2020. Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Mol. Ecol.* [Internet] 29:2535–2549. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/mec.15428

Hudson RR, Turelli M. 2003. STOCHASTICITY OVERRULES THE "THREE-TIMES RULE": GENETIC DRIFT, GENETIC DRAFT, AND COALESCENCE TIMES FOR NUCLEAR LOCI VERSUS MITOCHONDRIAL DNA. *Evolution* [Internet] 57:182–190. Available from: https://academic.oup.com/evolut/article/57/1/182/6755810

Ioannidis AG, Blanco-Portillo J, Sandoval K, Hagelberg E, Barberena-Jonas C, Hill AVS, Rodríguez-Rodríguez JE, Fox K, Robson K, Haoa-Cardinali S, et al. 2021. Paths

and timings of the peopling of Polynesia inferred from genomic networks. *Nature* [Internet] 597:522–526. Available from: https://www.nature.com/articles/s41586-021-03902-8

Ishigohoka J, Bascón-Cardozo K, Bours A, Fuß J, Rhie A, Mountcastle J, Haase B, Chow W, Collins J, Howe K, et al. 2024. Distinct patterns of genetic variation at low-recombining genomic regions represent haplotype structure. Available from: http://biorxiv.org/lookup/doi/10.1101/2021.12.22.473882

Jacobsen MW, Jensen NW, Nygaard R, Præbel K, Jónsson B, Nielsen NH, Pujolar JM, Fraser DJ, Bernatchez L, Hansen MM. 2022. A melting pot in the Arctic: Analysis of mitogenome variation in Arctic char (*Salvelinus alpinus*) reveals a 1000-km contact zone between highly divergent lineages. *Ecol. Freshw. Fish* [Internet] 31:330–346. Available from: https://onlinelibrary.wiley.com/doi/10.1111/eff.12633

Kamm JA, Spence JP, Chan J, Song YS. 2016. Two-Locus Likelihoods Under Variable Population Size and Fine-Scale Recombination Rate Estimation. *Genetics* [Internet] 203:1381–1399. Available from: https://academic.oup.com/genetics/article/203/3/1381/6065801

Kemppainen P, Schembri R, Momigliano P. 2024. Boundary Effects Cause False Signals of Range Expansions in Population Genomic Data. *Mol. Biol. Evol.* [Internet] 41:msae091. Available from: https://doi.org/10.1093/molbev/msae091

Klemetsen A. 2010. The Charr Problem Revisited: Exceptional Phenotypic Plasticity Promotes Ecological Speciation in Postglacial Lakes. *Freshw. Rev.* 3:49–74.

Klütsch CFC, Manseau M, Anderson M, Sinkins P, Wilson PJ. 2017. Evolutionary reconstruction supports the presence of a Pleistocene Arctic refugium for a large mammal species. *J. Biogeogr.* 44:2729–2739.

Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* [Internet] 15:356. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-014-0356-4

Koski MH, Layman NC, Prior CJ, Busch JW, Galloway LF. 2019. Selfing ability and drift load evolve with range expansion. *Evol. Lett.* [Internet] 3:500–512. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/evl3.136

Lawson DJ, Hellenthal G, Myers S, Falush D. 2012. Inference of Population Structure using Dense Haplotype Data. *PLOS Genet.* [Internet] 8:e1002453. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002453

Leigh JW, Bryant D. 2015. popart: full-feature software for haplotype network construction. *Methods Ecol. Evol.* [Internet] 6:1110–1116. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12410

Leitwein M, Duranton M, Rougemont Q, Gagnaire P-A, Bernatchez L. 2020. Using Haplotype Information for Conservation Genomics. *Trends Ecol. Evol.* [Internet] 35:245–258. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0169534719303040

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv.org* [Internet]. Available from: https://arxiv.org/abs/1303.3997v2

Li N, Stephens M. 2003. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics* [Internet] 165:2213–2233. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1462870/

Liang M, Nielsen R. 2014. The Lengths of Admixture Tracts. *Genetics* [Internet] 197:953–967. Available from: https://doi.org/10.1534/genetics.114.162362

Linderoth T. 2018. Identifying Population Histories, Adaptive Genes, and Genetic Duplication from Population-Scale Next Generation Sequencing.

Lou RN, Jacobs A, Wilder A, Therkildsen NO. 2021. A beginner's guide to low-coverage whole genome sequencing for population genomics. Preprints Available from: https://www.authorea.com/users/380682/articles/496574-a-beginner-s-guide-to-low-coverage-whole-genome-sequencing-for-population-genomics?commit=a25912ea40ed4aa7f84aeab0b40f1e4d2bea282c

Luqman H, Wegmann D, Fior S, Widmer A. 2023. Climate-induced range shifts drive adaptive response via spatio-temporal sieving of alleles. *Nat. Commun.* [Internet] 14:1080. Available from: https://www.nature.com/articles/s41467-023-36631-9

Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, Daly MJ, Bustamante CD, Kenny EE. 2017. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* [Internet] 100:635–649. Available from: https://www.cell.com/ajhg/abstract/S0002-9297(17)30107-6

Mee JA, Moore J-S. 2014. The ecological and evolutionary implications of microrefugia. *J. Biogeogr.* [Internet] 41:837–841. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/jbi.12254

Meier JI, Salazar PA, Kučka M, Davies RW, Dréau A, Aldás I, Box Power O, Nadeau NJ, Bridle JR, Rolian C, et al. 2021. Haplotype tagging reveals parallel formation of hybrid races in two butterfly species. *Proc. Natl. Acad. Sci.* [Internet] 118:e2015005118. Available from: https://www.pnas.org/doi/full/10.1073/pnas.2015005118

Meisner J, Albrechtsen A. 2018. Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics* [Internet] 210:719–731. Available from: https://doi.org/10.1534/genetics.118.301336

Mérot C, Berdan EL, Cayuela H, Djambazian H, Ferchaud A-L, Laporte M, Normandeau E, Ragoussis J, Wellenreuther M, Bernatchez L. 2021. Locally Adaptive Inversions Modulate Genetic Variation at Different Geographic Scales in a Seaweed Fly.Charlesworth D, editor. *Mol. Biol. Evol.* [Internet] 38:3953–3971. Available from: https://academic.oup.com/mbe/article/38/9/3953/6272233

Mestre F, Barbosa S, Garrido-García JA, Pita R, Mira A, Alves PC, Paupério J, Searle JB, Beja P. 2022. Inferring past refugia and range dynamics through the integration of fossil, niche modelling and genomic data. *J. Biogeogr.* [Internet] 49:2064–2076. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/jbi.14492

Moore J-S, 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. *J. Biogeogr.* 42:2089–2100.

Moore J-S, Harris LN, Kessel ST, Bernatchez L, Tallman RF, Fisk AT. 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. *Can. J. Fish. Aquat. Sci.* 53:1689–1699.

Moore J-S, Harris LN, Tallman RF, B TE. 2013. The interplay between dispersal and gene flow in anadromous Arctic char (Salvelinus alpinus): implications for potential for local adaptation. *Can. J. Fish. Aquat. Sci.* [Internet] 1338:1327–1338. Available from: http://www.nrcresearchpress.com/doi/abs/10.1139/cjfas-2013-0138

Nei M, Roychoudhury AK. 1974. SAMPLING VARIANCES OF HETEROZYGOSITY AND GENETIC DISTANCE. *Genetics* [Internet] 76:379–390. Available from: https://academic.oup.com/genetics/article/76/2/379/5990983

Nugent CM, Kess T, Langille BL, Beck SV, Duffy S, Messmer A, Smith N, Lehnert SJ, Wringe BF, Kent M, et al. 2024. Post-glacial recolonization and multiple scales of secondary contact contribute to contemporary Atlantic salmon (Salmo salar) genomic variation in North America. *J. Biogeogr.* [Internet] 51:1767–1782. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/jbi.14852

Pante E, Simon-Bouhet B. 2013. marmap: A Package for Importing, Plotting and Analyzing Bathymetric and Topographic Data in R. *PLoS ONE* 8:6–9.

Paradis E. 2010. pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* [Internet] 26:419–420. Available from: https://doi.org/10.1093/bioinformatics/btp696

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D. 2012. Ancient Admixture in Human History. *Genetics* [Internet] 192:1065–1093. Available from: https://doi.org/10.1534/genetics.112.145037

Pedersen BS, Quinlan AR. 2018. Mosdepth: quick coverage calculation for genomes and exomes.Hancock J, editor. *Bioinformatics* [Internet] 34:867–868. Available from: https://academic.oup.com/bioinformatics/article/34/5/867/4583630

Peischl S, Kirkpatrick M, Excoffier L. 2015. Expansion Load and the Evolutionary Dynamics of a Species Range. *Am. Nat.* [Internet] 185:E81–E93. Available from: https://www.journals.uchicago.edu/doi/10.1086/680220

Peter BM. 2022. A geometric relationship of F_2, F_3 and F_4-statistics with principal component analysis. *Philos. Trans. R. Soc. B Biol. Sci.* [Internet] 377:20200413. Available from: https://royalsocietypublishing.org/doi/10.1098/rstb.2020.0413

Peter BM, Slatkin M. 2013. Detecting Range Expansions from Genetic Data. *Evolution* [Internet] 67:3274–3289. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/evo.12202

Pickrell JK, Pritchard JK. 2012. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLOS Genet.* [Internet] 8:e1002967. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002967

Premoli A, Mathiasen P, Kitzberger T. 2010. Southern-most Nothofagus trees enduring ice ages: Genetic evidence and ecological niche retrodiction reveal high latitude (54°S) glacial refugia. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 298:247–256.

Price AL, Tandon A, Patterson N, Barnes KC, Rafaels N, Ruczinski I, Beaty TH, Mathias R, Reich D, Myers S. 2009. Sensitive Detection of Chromosomal Segments of Distinct Ancestry in Admixed Populations. *PLOS Genet.* [Internet] 5:e1000519. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000519

Prior CJ, Layman NC, Koski MH, Galloway LF, Busch JW. 2020. Westward range expansion from middle latitudes explains the Mississippi River discontinuity in a forest herb of eastern North America. *Mol. Ecol.* [Internet] 29:4473–4486. Available from: https://onlinelibrary.wiley.com/doi/10.1111/mec.15650

Rasmussen MS, Garcia-Erill G, Korneliussen TS, Wiuf C, Albrechtsen A. 2022. Estimation of site frequency spectra from low-coverage sequencing data using stochastic EM reduces overfitting, runtime, and memory usage.Gravel S, editor. *Genetics* [Internet] 222:iyac148. Available from: https://academic.oup.com/genetics/article/doi/10.1093/genetics/iyac148/6730749

Reich D, Thangaraj K, Patterson N, Price AL, Singh L. 2009. Reconstructing Indian population history. *Nature* [Internet] 461:489–494. Available from: https://www.nature.com/articles/nature08365

Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of the Coancestry Coefficient: Basis for a Short-Term Genetic Distance. *Genetics* [Internet] 105:767–779. Available from: https://academic.oup.com/genetics/article/105/3/767/5996242

Rougemont Q, Leroy T, Rondeau EB, Koop B, Bernatchez L. 2023. Allele surfing causes maladaptation in a Pacific salmon of conservation concern. *PLOS Genet.* [Internet] 19:e1010918. Available from: https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010918

Rougemont Q, Moore J-S, Leroy T, Normandeau E, Rondeau EB, Withler RE, Doornik DMV, Crane PA, Naish KA, Garza JC, et al. 2020. The role of historical contingency in shaping geographic pattern of deleterious mutation load in a broadly distributed Pacific Salmon. Available from: http://www.xkcd.com/446/#

Rubinacci S, Ribeiro DM, Hofmeister RJ, Delaneau O. 2021. Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat. Genet.* [Internet] 53:120–126. Available from: https://www.nature.com/articles/s41588-020-00756-0

Saha A, Kurland S, Kutschera VE, Díez-del-Molino D, Ekman D, Ryman N, Laikre L. 2024. Monitoring genome-wide diversity over contemporary time with new indicators applied to Arctic charr populations. *Conserv. Genet.* [Internet]. Available from: https://doi.org/10.1007/s10592-023-01586-3

Salisbury SJ, McCracken GR, Keefe D, Perry R, Ruzzante DE. 2019. Extensive secondary contact among three glacial lineages of Arctic Char (Salvelinus alpinus) in Labrador and Newfoundland. *Ecol. Evol.* [Internet] 9:2031–2045. Available from: http://doi.wiley.com/10.1002/ece3.4893

Shafer ABA, Cullingham CI, Côté SD, Coltman DW. 2010. Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Mol. Ecol.* 19:4589–4621.

Shipilina D, Pal A, Stankowski S, Chan YF, Barton NH. 2023. On the origin and structure of haplotype blocks. *Mol. Ecol.* [Internet] 32:1441–1457. Available from: https://onlinelibrary.wiley.com/doi/10.1111/mec.16793

Skotte L, Korneliussen TS, Albrechtsen A. 2013. Estimating Individual Admixture Proportions from Next Generation Sequencing Data. *Genetics* [Internet] 195:693–702. Available from: https://academic.oup.com/genetics/article/195/3/693/5935455

Slatkin M. 1993. Isolation by Distance in Equilibrium and Non-Equilibrium Populations. *Evolution* [Internet] 47:264–279. Available from: https://academic.oup.com/evolut/article/47/1/264/6870150

Slatkin M, Excoffier L. 2012. Serial Founder Effects During Range Expansion: A Spatial Analog of Genetic Drift. *Genetics* [Internet] 191:171–181. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3338258/

Sommer RS, Zachos FE. 2009. Fossil evidence and phylogeography of temperate species: 'glacial refugia' and post-glacial recolonization. *J. Biogeogr.* [Internet] 36:2013–2020. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2699.2009.02187.x

Toews DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Mol. Ecol.* [Internet] 21:3907–3930. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1365-294X.2012.05664.x

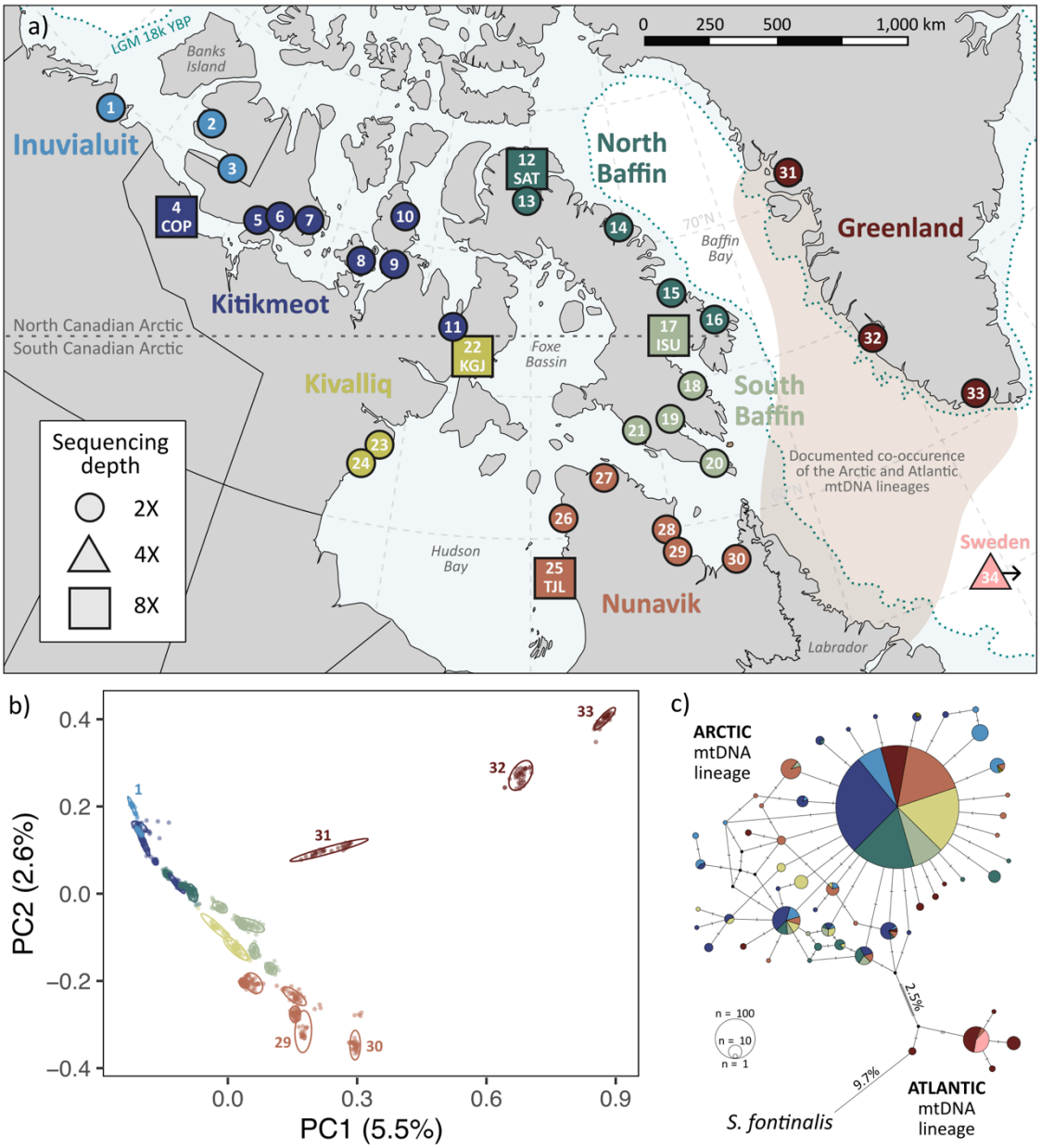
Tremblay NO, Schoen DJ. 1999. Molecular phylogeography of Dryas integrifolia: glacial refugia and postglacial recolonization. *Mol. Ecol.* [Internet] 8:1187–1198. Available from: http://doi.wiley.com/10.1046/j.1365-294x.1999.00680.x

Wiens BJ, Colella JP. 2024. That's not a Hybrid: How to Distinguish Patterns of Admixture and Isolation-by-Distance. Available from: http://biorxiv.org/lookup/doi/10.1101/2024.04.15.589658

Wójcik JM, Kawałko A, Marková S, Searle JB, Kotlík P. 2010. Phylogeographic signatures of northward post-glacial colonization from high-latitude refugia: a case study of bank voles using museum specimens. *J. Zool.* [Internet] 281:249–262. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-7998.2010.00699.x

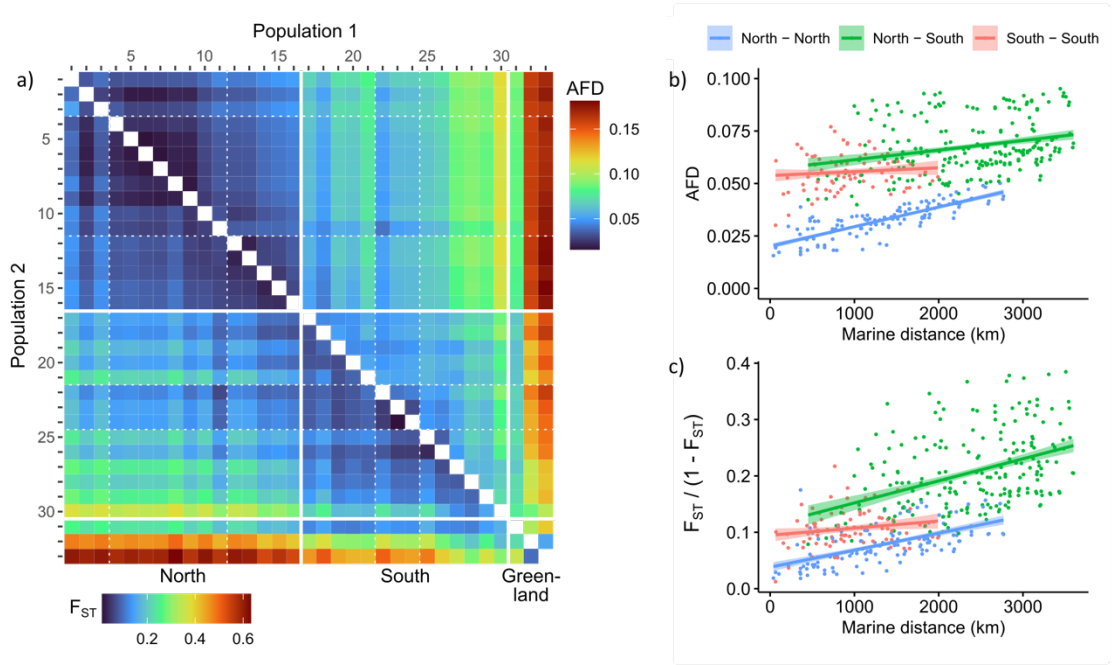
Zink RM, Barrowclough GF. 2008. Mitochondrial DNA under siege in avian phylogeography. *Mol. Ecol.* [Internet] 17:2107–2121. Available from: https://onlinelibrary.wiley.com/doi/10.1111/j.1365-294X.2008.03737.x

Figure 1: A highly structured distribution of genetic variation across the Arctic



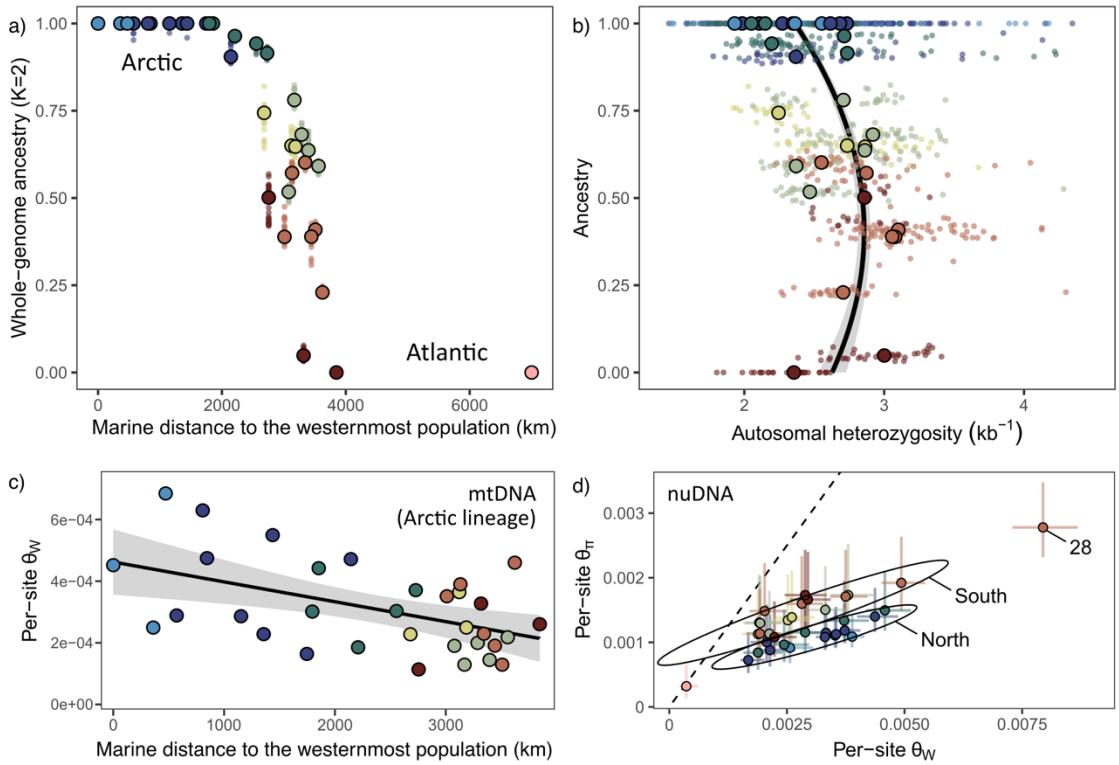
a) Sampling sites for 1140 Arctic Chars (Salvelinus alpinus) sequenced at a coverage of 2X (circle), 5X (triangle), or 8X (square). Colors indicate sampling regions, and the same color coding is used in all figure panels. Dotted blue lines delimit the ice margin at the last glacial maximum (18 ka, Dalton et al. 2020). b) Principal components analysis based on 396,113 independent SNPs. 95% confidence interval ellipses are drawn around each sampling site. c) Median-joining haplotype network for the mitochondrial D-loop control region (993 bp), showing divergence between the Arctic (left) and Atlantic (right) mitochondrial lineages. For each haplotype shared by n>1 individual, colored sections of the pie chart indicate the proportional representation of individuals from each region.

Figure 2: A North-South divide in genetic variation



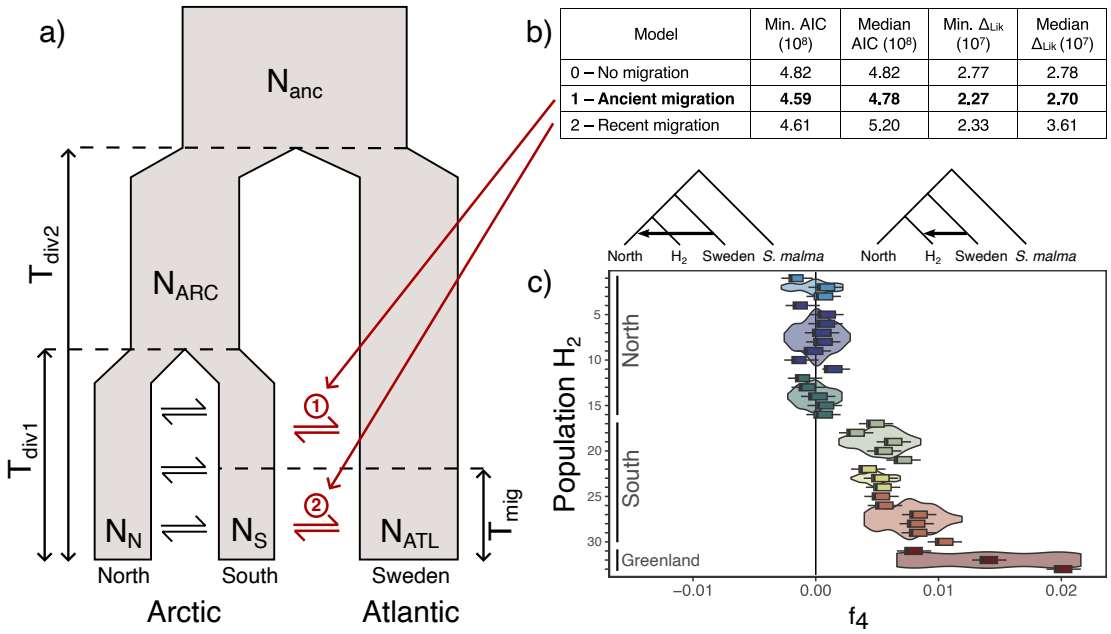
a) Pairwise genetic differentiation among populations estimated using allele frequency difference (AFD, above diagonal) and genome-wide weighted FST (below diagonal). Greenlandic and Canadian populations above and below the 66th parallel are divided by solid white lines, and dotted lines divide regions defined in Figure 1a. b-c) Isolation-by-distance as estimated by the linear regression between genetic (either b) AFD or c) linearized FST) and marine distance between populations 1-29. Population pairs were categorized based on their inclusion in the North (blue, pop. 1-16) or South (red, pop. 17-29) group, or including a Northern and a Southern population (green).

Figure 3: Higher nuclear diversity in admixed populations



a) Genome-wide Arctic ancestry, estimated by admixture proportion in NGSadmix (K = 2), of individuals (smaller points) and sampling sites (mean ancestry, circled points) along a spatial gradient from Inuvialuit (left) to Sweden (right). b) Individual autosomal heterozygosity (heterozygote sites per kb) is higher in the secondary contact zone (intermediate ancestry), as visualized by a fitted quadratic equation. c) Per-site Watterson's estimator (Θ_W) in mitochondrial haplotypes of the Arctic lineage by population. d) Per-site nucleotide diversity (Θ_π) and Watterson's estimator by population, estimated in 100 kb windows along the nuclear genome. The median (point), 25th, and 75th percentiles (error bars) are shown. The Northern (pop. 1-16) and Southern (pop. 17-30) Canadian populations are highlighted by 95% confidence interval ellipses and a dotted line follows $\Theta_\pi = \Theta_W$ (Tajima's D = 0).

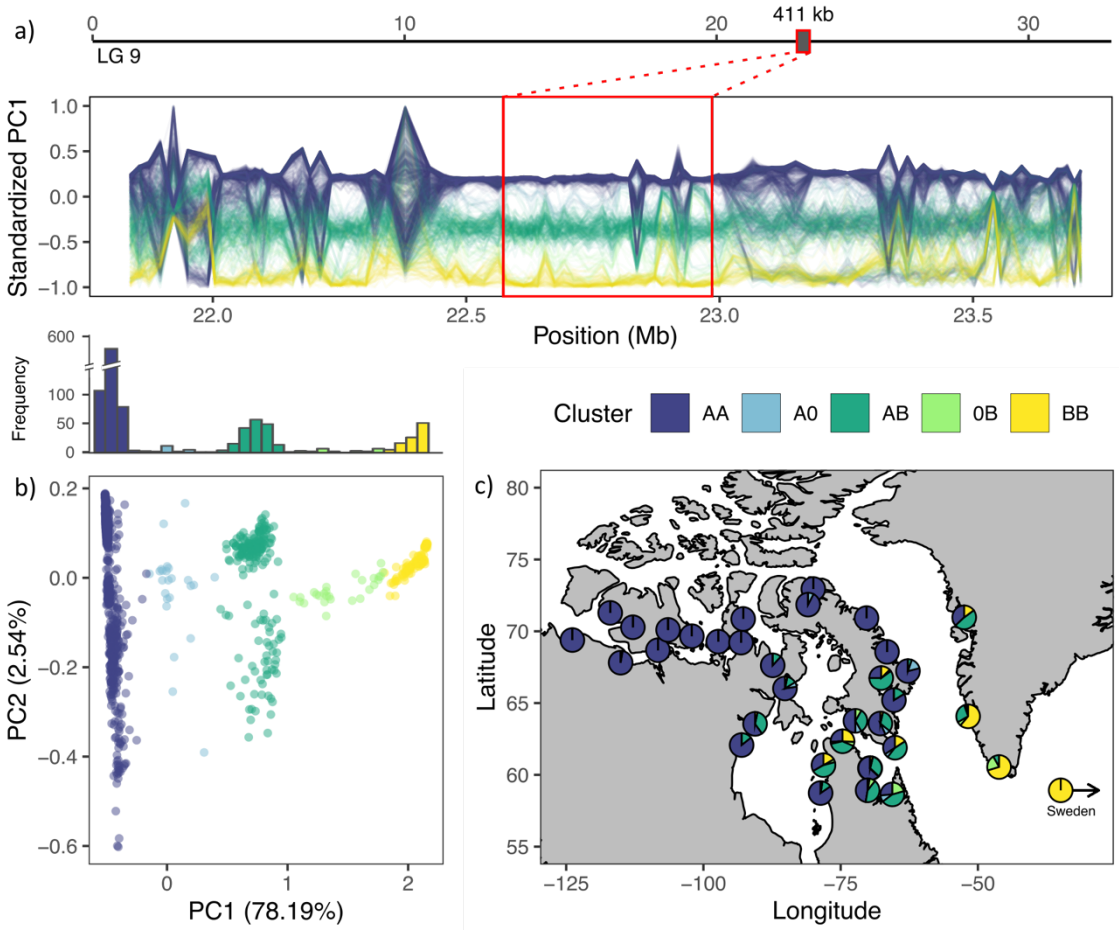
Figure 4: Evidence for gene flow between the Atlantic and South Arctic



Model	Min. AIC (10^8)	Median AIC (10^8)	Min. Δ_{Lik} (10^7)	Median Δ_{Lik} (10^7)
0 – No migration	4.82	4.82	2.77	2.78
1 – Ancient migration	**4.59**	**4.78**	**2.27**	**2.70**
2 – Recent migration	4.61	5.20	2.33	3.61

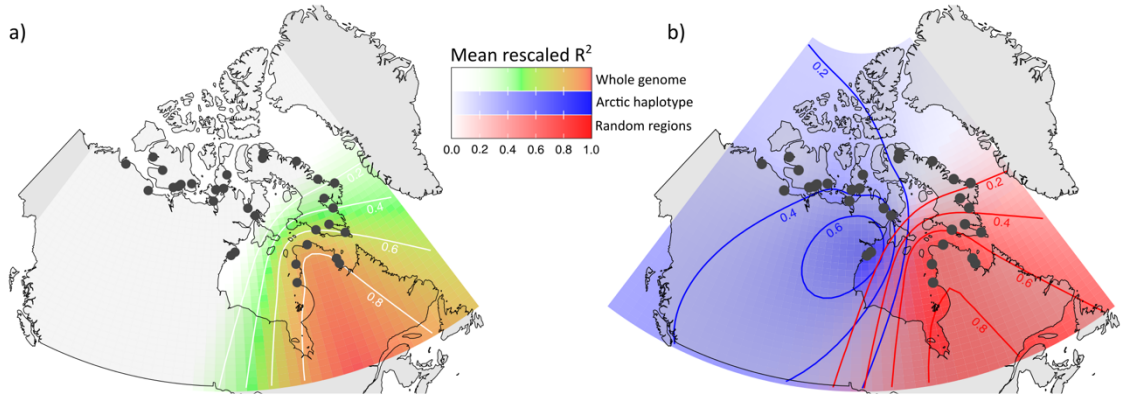
a) Simplified demographic model showing ancient divergence between the Arctic and Atlantic, followed by isolation with migration of the North (COP, pop. 4) and South (TJL, pop. 24) Arctic. b) Support (AIC and Δ_{Lik}, i.e. difference between observed and predicted likelihood of the SFS) for alternative fastsimcoal models (100 runs per model) for 0) absence of gene flow, 1) ancient or 2) recent gene flow between the South Arctic and Atlantic lineage (represented by the Swedish population), as highlighted in red in panel a). c) f_4 test for introgression from the Atlantic lineage (represented by the Swedish population), where a negative f_4 statistic supports introgression in population H_1 (only tests with populations from northern regions are shown) and a positive f_4 supports introgression in population H_2 (detailed on the y-axis). Similar lcWGS data from Dolly Varden (Salvelinus malma) was used as an outgroup.

Figure 5: Local ancestry inferred through local PCA



Representative case of a low-recombining diverged putative haploblock corresponding to a local PCA outlier region (win8) on LG9 (NC_036849.1). a) Position of the MDS outlier region (red) on the linkage group (above) and standardized PC1 coordinates in PCAs for successive non-overlapping 100-SNP windows inside and around the outlier region (below). Each sample is represented by a line colored according to their haplogroup inferred from panel b). b). PCA for SNPs in the outlier region where the three major clusters found on PC1 are attributed to the 2 homozygotes (AA, blue, left; BB, yellow, right) and their heterozygote (AB, green, middle). Intermediate genotypes (A0, light blue; 0B, light green) were assigned to individuals outside of the cluster distributions. The histogram (above) represents the density of individuals along the first axis of the PCA. c) Proportion of inferred haplogroups for the win8 haploblock across sampling sites.

Figure 6: Genome-wide and Arctic lineage-specific inference of the origin of recolonization



a) Time Difference of Arrival (TDoA) analysis for inference of the origin of recolonization, where red color indicates a higher correlation (R^2) between distance from the putative origin and whole-genome directionality index (ψ). b) TDoA for ψ calculated in Arctic haplogroup homozygotes (10 individuals per population) in 23 putative local ancestry tracks (total length = 20 Mb). The mean rescaled R^2 for 100 replicate runs is shown in blue, and the mean rescaled R^2 for 100 runs on 23 random genomic regions (total length = 20 Mb) is shown in red.