Genetic population structure and variation at phenology-related loci in anadromous Arctic char (Salvelinus alpinus)

Rikke P. A. Madsen1 | Magnus W. Jacobsen1 | Kathleen G. O’Malley2 | Rasmus Nygaard3 | Kim Præbel4 | Bjarni Jónsson5 | Jose M. Pujolar1 | Dylan J. Fraser6 | Louis Bernatchez7 | Michael M. Hansen1

1Department of Bioscience, Aarhus University, Aarhus, Denmark
2Department of Fisheries and Wildlife, Hatfield Marine Science Center, Coastal Oregon Marine Experiment Station, Oregon State University, Newport, OR, USA
3Greenland Institute of Natural Resources, Nuuk, Greenland
4Norwegian College of Fishery Science, UiT the Arctic University of Norway, Tromsø, Norway
5North West Iceland Nature Center, Sudárkökur, Iceland
6Department of Biology, Concordia University, Montreal, QC, Canada
7IBIS (Institut de Biologie Intégrative et des Systèmes), Université Laval, Québec, QC, Canada

Correspondence
Michael M. Hansen, Department of Bioscience, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus C, Denmark.
Email: mmh@bios.au.dk

Present address
Magnus W. Jacobsen, National Institute of Aquatic Resources, Technical University of Denmark, Silkeborg, Denmark
Jose M. Pujolar, National History Museum of Denmark, University of Copenhagen, Copenhagen, Denmark

Funding information
The Danish Council for Independent Research, Natural Science, Grant/Award Number: 1323-00158A

Abstract
The Arctic will be especially affected by climate change, resulting in altered seasonal timing. Anadromous Arctic char (Salvelinus alpinus) is strongly influenced by sea surface temperature (SST) delimiting time periods available for foraging in the sea. Recent studies of salmonid species have shown variation at phenology-related loci associated with timing of migration and spawning. We contrasted genetic population structure at 53 SNPs versus four phenology-related loci among 15 anadromous Arctic char populations from Western Greenland and three outgroup populations. Among anadromous populations, the time period available for foraging at sea (>2°C) ranges from a few weeks to several months, motivating two research questions: (a) Is population structure compatible with possibilities for evolutionary rescue of anadromous populations during climate change? (b) Does selection associated with latitude or SST regimes act on phenology-related loci? In Western Greenland, strong isolation by distance at SNPs was observed and spatial autocorrelation analysis showed genetic patch size up to 450 km, documenting contingency and gene flow among populations. Outlier tests provided no evidence for selection at phenology-related loci. However, in Western Greenland, mean allele length at OtsClock1b was positively associated with the time of year when SST first exceeded 2°C and negatively associated with duration of the period where SST exceeded 2°C. This is consistent with local adaptation for making full use of the time period available for foraging in the sea. Current adaptation may become maladaptive under climate change, but long-distance connectivity of anadromous populations could redistribute adaptive variation across populations and lead to evolutionary rescue.

Keywords
Arctic char, climate change, clock gene, phenology, sea surface temperature, spatial autocorrelation
1 | INTRODUCTION

Ongoing anthropogenic climate change has the potential to profoundly affect the living conditions of biota, involving, for example physiological stress during warm periods, altered ecological interactions and colonisation of new species (Hoffmann & Sgro, 2011; Parmesan, 2006; Pörtner & Peck, 2010; Thackeray et al., 2016). A much debated issue concerns whether or not organisms are able to respond to rapid climate change by genetically based microevolution or have to rely on phenotypic plasticity (Hansen et al., 2012; Hoffmann & Sgro, 2011; Merila & Hendry, 2014). Crozier and Hutchings (2014) found that very few studies of fishes had documented adaptive change that could be ascribed to changing climate, with a few notable exceptions such as a study of altered migration timing in pink salmon (Oncorhynchus gorbuscha) (Kovach, Gharrett, & Tallmon, 2012). Nevertheless, several studies have presented results consistent with adaptation to extant climate and temperature regimes in fishes at phenotypic traits and/or candidate genes that supposedly reflect evolution over longer time spans than those over which anthropogenic climate change occurs (Bernatchez, 2016; Bradbury et al., 2010; Harrisson et al., 2017; Jensen et al., 2008; Koskinen, Haugen, & Primmer, 2002; Narum, Campbell, Kozfky, & Meyer, 2010; Perrier, Ferchaud, Sirois, Thibault, & Bernatchez, 2017). Adaptations to current climate conditions could become increasingly maladaptive as the climate changes, but could also act as a source of genetic variation for future evolutionary rescue, through the influx of genetic variation into populations via gene flow to allow adaptation to altered environmental conditions (Gonzalez, Ronce, Ferriere, & Hochberg, 2013).

It has been argued that in temperate and Arctic regions, the most pronounced changes to living conditions concern altered seasonal timing, including later arrival of winter and earlier arrival of spring, rather than increased temperature per se (Bradshaw & Holzapfel, 2006, 2008). This means that phenological traits, such as timing of migration and reproduction, may be particularly important for the future persistence of organisms. Many phenological traits are regulated by an internal clock that is synchronised particularly by photoperiods and temperature. A core set of genes form and regulate the circadian clock system across vertebrate taxa: Clock, Bmal, Period and Cryptochrome (Idda et al., 2012; Lincoln, Andersson, & Loudon, 2003; Lowrey & Takahashi, 2004). Clock, in particular, has received considerable attention. A critical domain in this gene is the carboxyl-terminal polyglutamine repeat motif (polyQ), in which increases and decreases in the number of polyQ repeats affect gene expression (Darlington et al., 1998; Hayasaka, LaRue, & Green, 2002). Several studies of birds have revealed positive associations between clock (polyQ) allele lengths and breeding latitude (Bazzi et al., 2016; Jøhnsen et al., 2007), but also examples of no association in some species (Dor et al., 2012).

Arctic regions are particularly affected by climate change (Leduc, Matthews, & Elia, 2016). For instance, the decade from 2001 to 2010 was the warmest period on record in Greenland from 1784 to the present, and by 2050, temperature is projected to have increased by 3°C in winter, 4°C in spring and 2°C in summer and autumn (Cappelen & Vinther, 2014). Arctic char (Salvelinus alpinus) is a cold water-adapted salmonid widely distributed in the northern circumpolar Arctic region (Klemetsen et al., 2003), and in Greenland, anadromous populations are found throughout coastal regions. They exhibit a complex life history involving repeat spawning interrupted by years of no spawning. It is generally assumed that anadromous populations spawn around October (Klemetsen et al., 2003). Due to logistic constraints, no systematic records of spawning time are available for Arctic char in Greenland. However, ripe and spent spawners were observed in late September–early October in Southern Greenland during the course of the present study, and it is assumed that spawning takes place earlier in more northern regions.

Both spawning and nonspawning anadromous char overwinter in freshwater, the latter presumably in order to avoid osmotic stress in the marine environment during cold Arctic winters (Klemetsen et al., 2003; Moore et al., 2017). Experimental work by Finstad, Nilssen, and Arnesen (1989) demonstrated osmotic stress and high mortality when Arctic char were exposed to high salinity and a temperature of 1°C during winter, but not when they were exposed to the same conditions during summer. This suggests that complex interactions exist between osmoregulatory capacity and seasonal change, possibly regulated by photoperiod. In general, the total length of the season that anadromous Arctic char are able to spend foraging at sea, as determined by the sea temperature, is assumed to be a critical parameter determining growth and life history (Dutil, 1986). Greenlandic anadromous char populations are distributed at a range of more than 20 latitudinal degrees, implying that considerable geographical variation in the length of the growth season must be expected, leading to the possibility of local adaptation of associated phenological traits.

The goal of this study was to address two key research questions: (a) Is the genetic structure and differentiation among anadromous populations compatible with possibilities for evolutionary rescue during climate change? (b) Does selection associated with latitude or marine temperature regimes act on the phenology-related markers? Towards this end, the genetic structure of anadromous char populations in Western Greenland was analysed along with ‘outgroup’ populations from Eastern Greenland, Iceland and Norway, and the latter two represented by landlocked lake populations. Two data sets of fifty-three presumably neutral SNPs (single nucleotide polymorphisms) and four phenology-related loci (OtsClock1b, OtsClock2b, Cryptochrome2b.2 and Cryptochrome3), respectively, were analysed in 18 populations. Moreover, remotely sensed data were extracted on sea surface temperature close to the mouths of the sampled rivers and lakes to estimate the onset, end and duration of the periods of time that local populations could potentially spend at sea.

2 | MATERIALS AND METHODS

2.1 | Samples

Adipose fin clips were collected from 2005 to 2016 by angling, net fishing and electrofishing. We aimed for sample sizes of twenty, as
higher sample sizes generally do not improve estimates of standard population genetic statistics as compared to increasing number of loci (Takezaki & Nei, 1996). Among the 18 populations included in the study, 15 were anadromous populations located along the West coast of Greenland. Three additional populations represented anadromous char from Eastern Greenland and two landlocked lake populations from Iceland and Norway (see Figure 1 and Table 1). Collection and handling of samples in Greenland took place according to survey licences G14-034 and G15-013 from the Government of Greenland.

2.2 | Molecular analyses

DNA was extracted using the E.Z.N.A DNA Tissue Extraction Kit (Omega Bio-Tek) according to the manufacturer’s recommendations. Two sets of loci were analysed: (a) 53 single nucleotide polymorphisms (SNPs) developed for Arctic char (Jacobsen et al., 2017) and assumed to represent neutral markers as based on outlier tests conducted in Christensen, Jacobsen, Nygaard, and Hansen (2018) and (b) four candidate loci assumed to be involved in phenology. SNPs were genotyped on a 96.96 Dynamic Array on the Fluidigm Biomark platform (Fluidigm Corporation). As explained in Jacobsen et al. (2017), the initial set consisted of 96 SNPs, of which 43 could not be scored reliably due particularly to the presence of paralogs presumably resulting from ancient tetraploidy in salmonid fishes (Allendorf et al., 2015). Genotypes were scored using the associated Fluidigm® SNP Genotyping Analysis software.

The candidate loci consisted of the polyQ region of the Clock gene OtsClock1b, microsatellites closely linked to the two duplicated copies Cryptochrome2b.2 and Cryptochrome3 of the circadian rhythm gene Cryptochrome, and a microsatellite Ots515NWFSC, which is a QTL for spawning time and body weight in rainbow trout (O’Malley, Sakamoto, Danzmann, & Ferguson, 2003). Primer sequences for the loci are described in Naish and Park (2002), O’Malley, Camara, and Banks (2007) and O’Malley, McClelland, and Naish (2010). The forward primers of OtsClock1b, Ots515NWFSC, Cryptochrome2b.2 and Cryptochrome3 were labelled with the fluorescent dyes PET, NED, FAM and VIC respectively. The loci were PCR amplified at an annealing temperature of 55°C in 30 µl reactions containing 15 µl QIAGEN Multiplex PCR Mastermix (QIAGEN), 3 µl 100 µM primer mix, 10 µl fluorescently labelled primer and 10 µl reverse primer, 11 µl H2O and 1 µl sample DNA (concentrations between ca. 80 and 400 ng/µl). Genotyping was outsourced to Macrogen Inc., where fragments were resolved on an ABI 3730XL capillary sequencer using a 600 LIZ internal size standard (Applied Biosystems). Scoring of genotypes was conducted using the software Geneious 10.0.7 (Kearse et al., 2012).

Salmonid fishes are ancient tetraploids, and simple Mendelian inheritance cannot always be assumed (Allendorf et al., 2015; Allendorf & Thorgaard, 1984). Also, scoring of multiallelic loci may in itself be complicated. In order to validate Mendelian inheritance and scoring of the phenology-related loci, two full-sib family crosses were therefore established, based on two males and two females sampled in October 2013 in the NUUK-2 population (see Table 1 and Figure 1). Fertilised eggs were incubated in Petri dishes at 5°C following Wedekind and Muller (2004). This took place at the Greenland Institute of Natural Resources, Nuuk, where Petri dishes were inspected daily, and upon hatching, the larvae were euthanised and stored in 96% ethanol at −18°C. The parents and 10 offspring from each family were genotyped.

2.3 | Genetic population structure

For all analyses of population structure, SNPs and candidate loci were analysed separately. Mean heterozygosity was estimated using GENEPOP version 4.2 (Rousset, 2008), and the same software was used to test for Hardy–Weinberg equilibrium at all loci in all populations. Genetic differentiation for the two data sets was analysed by (a) an AMOVA (analysis of molecular variance) involving all populations and (b) a hierarchical AMOVA involving populations from Western Greenland, as implemented in ARLEQUIN version 3.5.2.2 (Excoffier, Guillaume, & Schneider, 2005). For this study, five regional groups of Western Greenland populations were defined by the geographical location of populations: region 1 (UUMM-1, UUMM-2 and DISK-1), region 2 (KANG-1 and SISI-1), region 3 (MANI-1 and MANI-2), region
TABLE 1  Overview of samples and localities showing sample codes, localities, geographical coordinates, major geographic regions, year of sampling, life history of populations, sample size (N) and mean expected heterozygosity (H_e) for SNPs and phenology-related markers respectively

<table>
<thead>
<tr>
<th>Sample code</th>
<th>Locality</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Major geographic region</th>
<th>Year of sampling</th>
<th>Life history form</th>
<th>N</th>
<th>H_e (SNPs)</th>
<th>H_e (phenology-related)</th>
</tr>
</thead>
<tbody>
<tr>
<td>QAAN-1</td>
<td>Qaanaaq</td>
<td>77.46°N</td>
<td>−69.23°W</td>
<td>Western Greenland</td>
<td>2012</td>
<td>Anadromous</td>
<td>18</td>
<td>0.11</td>
<td>0.18</td>
</tr>
<tr>
<td>UUMM-1</td>
<td>Umivik</td>
<td>71.66°N</td>
<td>−54.10°W</td>
<td>Western Greenland</td>
<td>2015</td>
<td>Anadromous</td>
<td>20</td>
<td>0.29</td>
<td>0.35</td>
</tr>
<tr>
<td>UUMM-2</td>
<td>Sermeerlat</td>
<td>70.54°N</td>
<td>−50.77°W</td>
<td>Western Greenland</td>
<td>2015</td>
<td>Anadromous</td>
<td>20</td>
<td>0.26</td>
<td>0.27</td>
</tr>
<tr>
<td>DISK-1</td>
<td>Disko Island</td>
<td>69.25°N</td>
<td>−53.51°W</td>
<td>Western Greenland</td>
<td>2014</td>
<td>Anadromous</td>
<td>20</td>
<td>0.28</td>
<td>0.40</td>
</tr>
<tr>
<td>KANG-1</td>
<td>Robinson River</td>
<td>66.71°N</td>
<td>−51.43°W</td>
<td>Western Greenland</td>
<td>2014</td>
<td>Anadromous</td>
<td>20</td>
<td>0.22</td>
<td>0.59</td>
</tr>
<tr>
<td>SISI*1</td>
<td>Sisimiut</td>
<td>66.43°N</td>
<td>−53.61°W</td>
<td>Western Greenland</td>
<td>2014</td>
<td>Anadromous</td>
<td>20</td>
<td>0.32</td>
<td>0.51</td>
</tr>
<tr>
<td>MANI-1</td>
<td>Kangerdluaassuk</td>
<td>65.57°N</td>
<td>−52.38°W</td>
<td>Western Greenland</td>
<td>2014</td>
<td>Anadromous</td>
<td>20</td>
<td>0.30</td>
<td>0.58</td>
</tr>
<tr>
<td>MANI-2</td>
<td>Kangia</td>
<td>65.31°N</td>
<td>−51.97°W</td>
<td>Western Greenland</td>
<td>2015</td>
<td>Anadromous</td>
<td>20</td>
<td>0.26</td>
<td>0.65</td>
</tr>
<tr>
<td>NUUK-1</td>
<td>Kapisilil</td>
<td>64.24°N</td>
<td>−50.20°W</td>
<td>Western Greenland</td>
<td>2012</td>
<td>Anadromous</td>
<td>18</td>
<td>0.22</td>
<td>0.47</td>
</tr>
<tr>
<td>NUUK-2</td>
<td>Kobbefjord</td>
<td>64.14°N</td>
<td>−51.38°W</td>
<td>Western Greenland</td>
<td>2013</td>
<td>Anadromous</td>
<td>19</td>
<td>0.27</td>
<td>0.55</td>
</tr>
<tr>
<td>NUUK-3</td>
<td>Praestefjord</td>
<td>64.00°N</td>
<td>−51.24°W</td>
<td>Western Greenland</td>
<td>2013</td>
<td>Anadromous</td>
<td>20</td>
<td>0.28</td>
<td>0.50</td>
</tr>
<tr>
<td>NUUK-4</td>
<td>Qarajat</td>
<td>63.99°N</td>
<td>−51.48°W</td>
<td>Western Greenland</td>
<td>2012</td>
<td>Anadromous</td>
<td>20</td>
<td>0.25</td>
<td>0.51</td>
</tr>
<tr>
<td>NUUK-5</td>
<td>Eqaluit</td>
<td>64.13°N</td>
<td>−50.47°W</td>
<td>Western Greenland</td>
<td>2012</td>
<td>Anadromous</td>
<td>20</td>
<td>0.30</td>
<td>0.63</td>
</tr>
<tr>
<td>QAOQ-1</td>
<td>Lakseelv</td>
<td>60.89°N</td>
<td>−45.84°W</td>
<td>Western Greenland</td>
<td>2014</td>
<td>Anadromous</td>
<td>20</td>
<td>0.16</td>
<td>0.34</td>
</tr>
<tr>
<td>QAOQ-2</td>
<td>Eqaluit</td>
<td>60.76°N</td>
<td>−45.54°W</td>
<td>Western Greenland</td>
<td>2014</td>
<td>Anadromous</td>
<td>20</td>
<td>0.15</td>
<td>0.41</td>
</tr>
<tr>
<td>SCOR-1</td>
<td>Scoresbysund</td>
<td>70.35°N</td>
<td>−28.14°W</td>
<td>Eastern Greenland</td>
<td>2012</td>
<td>Anadromous</td>
<td>20</td>
<td>0.08</td>
<td>0.17</td>
</tr>
<tr>
<td>ICEL-1</td>
<td>Vatnshlidarvatn</td>
<td>65.52°N</td>
<td>−19.64°W</td>
<td>Iceland</td>
<td>2016</td>
<td>Landlocked</td>
<td>20</td>
<td>0.07</td>
<td>0.59</td>
</tr>
<tr>
<td>NORW-1</td>
<td>Biggijavri</td>
<td>69.33°N</td>
<td>23.45°W</td>
<td>Norway</td>
<td>2005</td>
<td>Landlocked</td>
<td>16</td>
<td>0.06</td>
<td>0.34</td>
</tr>
</tbody>
</table>
4 (NUUK-1, NUUK-2, NUUK-3, NUUK-4 and NUUK-5) and region 5 (QAQO-1 and QAQO-2). The geographically remote QAAN-1 population could not be meaningfully included in a regional group with other populations and was omitted from this analysis. Finally, $F_{ST}$ between all pairs of populations was estimated, also using ARLEQUIN.

The genetic relationships among populations at the SNPs were further analysed by DAPC (discriminant analysis of principal components) (Jombart, Devillard, & Balloux, 2010), implemented in the R package adegenet (Jombart, 2008). Briefly, the method defines clusters of individuals without prior knowledge of their sample of origin and identifies discriminant functions that distinguish clusters while at the same time minimizing variation within clusters. We first identified the most likely number of clusters and the individuals belonging to them based on k-means clustering and Bayesian information criterion, followed by choosing the optimal number of principal components (using cross-validation) and discriminant axes, as detailed in the documentation for DAPC.

Isolation by distance (IBD) for the two classes of markers was tested using Mantel tests implemented in the software Isolation-By-Distance, web service version 3.23 (Jensen, Bohonak, & Kelley, 2005). Pairwise $F_{ST}$ estimates were used as genetic distance, and geographical distance (shortest waterway distance) was estimated using Google Earth. Moreover, IBD was visualized by genetic-geographical distance scatter plots along with their regression lines and 95% confidence intervals. The analyses focused exclusively on the 15 populations from Western Greenland (i.e., excluding the geographically distant SCOR-1, ICEL-1 and NORW-1 populations).

Finally, we used spatial autocorrelation analysis (Sokal & Oden, 1991) implemented in GenAlEx 6.5 (Peakall & Smouse, 2006, 2012; Smouse & Peakall, 1999) in order to assess the geographical scale in Western Greenland over which individual genotypes show nonrandom association. This was based on all pairwise individual genetic distances (Smouse & Peakall, 1999) and a corresponding geographical distance matrix based on waterway distances between sites, as described for the isolation by distance analyses. We assumed a geographical distance of 0 for individuals from the same rivers. In order to balance the number of individuals within geographical distance classes, we assumed classes with increments of 50 km from 0 to 500 and subsequently with increments of 500 km. Both the 95% confidence interval of distance-class specific $r$ values and the 95% confidence interval in case of no spatial structure of individuals were estimated by bootstrapping over pairs of individuals 9,999 times.

### 2.4 | Sea surface temperature data

Remotely sensed sea surface temperature data (in the following denoted SST), encompassing a resolution of 0.25 degree latitude × 0.25 degree longitude on a global grid and measured for each day, were provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Website at http://www.esrl.noaa.gov/psd/. Data from 1984, 1994, 2004 and 2014 were used, hence covering temperatures for a time span of 40 years. Data for each day of the year from the position closest to the sampled river/lake mouths inhabited by anadromous char (hence excluding the resident populations ICEL-1 and NORW-1) were retrieved using the function extractOISSTdaily from the R script NOAA_OISST_ncdf4.R (http://lukemiller.org/index.php/2014/11/extracting-noaa-sea-surface-temperatures-with-ncdf4/). Subsequently, the mean temperature per day over the total time period was calculated. As anadromous char experience osmotic stress at 1°C (Finstad et al., 1989), SST < 2°C was tentatively defined as unfavourable to char in the sea. For each locality, the time period (in the following denoted SST window) was estimated during which SST was ≥2°C. The start and end points of the SST window, measured in numbers of days starting from 1 January, and the duration of the SST window were subsequently used for some of the selection tests (see below).

### 2.5 | Selection tests

Outlier tests implemented in ARLEQUIN (Excoffier, Hofer, & Foll, 2009) were used for assessing possible selection at the phenotype-related loci, with the SNP data set included to provide a putatively neutral baseline of differentiation (Christensen et al., 2018). The first, involving all populations was the $F_{ST}$-based test by Beaumont and Nichols (1996). The second was an extension of this test by Excoffier et al. (2009), which takes underlying hierarchical structure of populations into account. The latter test was based on the same populations and regional groups in Western Greenland as described for the hierarchical AMOVA (see above). The analyses were based on 10,000 simulations.

A third outlier test was conducted, that is BAYESCENV (de Villemereuil, Gaggiotti, & O’Hara, 2015) which tests for association between loci and environmental parameters. It is an extension of the outlier test BAYESCAN (Foll & Gaggiotti, 2008) and distinguishes between (a) neutrality, (b) a locus-specific effect, possibly representing selection but not associated with the environmental parameter tested, and (c) an effect of the environmental parameter on a specific locus which could represent selection. The total set of SNPs and phenotype-related loci were included, and the environmental parameters tested were the start dates, end dates and duration of SST windows, along with latitude of the sample localities. The recommended default settings of the programme were used (20 pilot runs each consisting of 2,000 steps, burn-in of 50,000 steps followed by 50,000 steps and a thinning interval size of 10).

Finally, we tested for an association between mean allele lengths (assumed to represent polyQ copy number variation) in populations at OtsClock1b and (a) latitude, (b) start, (c) end dates and (d) duration of SST windows, using linear models (as in e.g., O’Malley and Banks (2008)) implemented in R (R Core Team 2018).

### 3 | RESULTS

#### 3.1 | Mendelian inheritance of phenotype-related genes

The experimental crosses were informative for resolving inheritance except for Cryptochrome2b.2 (Table S1). At Ots515NWESC and
Weinberg equilibrium yielded significant outcomes (p < .001) after false discovery rate (FDR) correction by the B-Y method (Narum, 2006) (Table S2). Hence, the populations can be assumed to be in Hardy–Weinberg equilibrium.

3.2 Summary statistics and genetic population structure

Among 18,603 genotypes in the SNP data set (351 individuals × 53 loci), only 57 could not be resolved, leading to 0.3% missing data. Estimated mean heterozygosity across SNPs per population varied from 0.06 (NORW-1) to 0.32 (SISI-1). There was a distinct pattern of lower heterozygosity across phylogeographic-related loci from Western Greenland (p < .001 as determined by a permutation test in FSTAT 2.9.3 (Goudet, 1995; see also Table 1 and Table S2). The phenology-related loci encompassed 1,404 genotypes (351 individuals × 4 loci), of which only 13 (0.9%) could not be resolved. Estimated mean heterozygosity across phylogeographic-related loci ranged from 0.18 (QAAN-1) to 0.65 (MANI-2) (Table 1, Table S2). In contrast to SNPs, these loci were all multiallelic with numbers of alleles ranging from 4 to 24 per locus (Table S2). Three out of a total of 741 tests for Hardy–Weinberg equilibrium yielded significant outcomes (p < .001) after false discovery rate (FDR) correction by the B-Y method (Narum, 2006) (Table S2). Hence, the populations can be assumed to be in Hardy–Weinberg equilibrium.

Overall genetic differentiation (FST) across all populations and over all SNPs was 0.27 (p < .001). The hierarchical AMOVA involving only Western Greenland populations showed that the largest part of differentiation was distributed among geographic groups of populations (FCT = 0.11, p < .001), whereas a relatively smaller part was distributed among populations within geographic groups (FSC = 0.09, p < .001). Genetic differentiation at phylogeographic-related loci was similar, with overall FST = 0.23 (p < .001) across all populations. For the hierarchical AMOVA, FCT was 0.10 (p < .001) and FSC was 0.06 (p < .001). FST between pairs of populations for the SNP data set ranged from 0.02 (NUUK-2 vs. NUUK-3 and NUUK-2 vs. NUUK-4) to 0.67 (QAAN-1 vs. NORW-1), whereas for the phylogeographic-related loci FST ranged from 0.02 (several pairs of populations) to 0.47 (QAAN-1 vs. SCOR-1; Table S3).

For the DAPC analysis of the SNP data, the most likely number of groups represented by the individual multi-locus genotypes was 9, as determined by the Bayesian information criterion (see Figure S1). Grouping of individuals (Figure 2a) showed that the northernmost populations (QAAN-1, UUMM-1, UUMM-2 and DISK-1) were composed of three clusters (Clusters 1, 7 and 9), and individuals from KANG-1 belonged exclusively to Cluster 2. Individuals from the populations SISI-1, MANI-1, MANI-2, NUUK-1, NUUK-2, NUUK-3, NUUK-4 and NUUK-5 were distributed across Clusters 1, 2, 3, 4, 5, 6, 7 and 8. QAQO-2 individuals were assigned to Clusters 3 and 8. Finally, all individuals from SCOR-1, ICEL-1 and NORW-1 were assigned to Cluster 3. The first 25 principal components and seven discriminant axes were retained for the DAPC scatterplot. Axes 1 and 2 (Figure 2b) demonstrated a strong geographic structure among the nine inferred clusters, with Clusters 9, 1 and 7 (northernmost populations in Western Greenland) representing one end of a continuum and Cluster 3 (south-western and Eastern Greenland, Iceland and Norway) representing the other end. Hence, the results of DAPC showed good correspondence with the geographical location of populations, justifying the groupings of populations used for the hierarchical AMOVA.

The close relationships between geographical and genetic relationships were further illustrated for both SNPs and candidate loci by analysis of isolation by distance involving only the anadromous Western Greenland populations (Figure 3a,b). Hence, there was significant correlation between genetic differentiation and geographical distance for SNPs (R2 = .92, p = .0000) and for phenology-related loci (R2 = .55, p = .0000).

The spatial autocorrelation analysis (Figure 4) showed a mean correlation among individuals from the same freshwater localities of 0.330 and subsequently declined and reached its first intercept with the x-axis at 450 km. This value is usually referred to as the genetic patch size (Smouse & Peakall, 1999; Sokal & Wartenberg, 1983). Using distance classes of 100 km instead of 50 km yielded a similar genetic patch size (data not shown).

3.3 Sea surface temperature data

Sea surface temperature (SST) data were retrieved from all coastal regions close to the river mouths of the sampled anadromous populations. In the case of NUUK-2, NUUK-3, NUUK-4 and NUUK-5, the geographical distances between river mouths were short. Therefore, these populations shared the same pixel of the SST grid and thereby similar temperature regimes. The SST windows, defined by the time periods during the year when SST exceeded 2°C, varied considerably across populations (Figure 5, Table S4). Hence, SST exceeded 2°C for only a few weeks in the northernmost populations QAAN-1, UUMM-1, UUMM-2 and in SCOR-1 from Eastern Greenland (Figure 5a,b,c,m). In contrast, SST exceeded 2°C for several months in most of the other populations, potentially leaving longer time periods for Arctic char to forage in the sea. The lower temperatures in the south-western localities QAQO-1 and QAQO-2 (Figure 5k,l) as opposed to the more northern localities DISK-1, SISI-1, KANG-1, MANI-1, MANI-2 and NUUK-1 to 5 (Figure 5d-j) reflects the influence of the West Greenland Current (Lloyd, Kuijpers, Long, Moros, & Park, 2007). Hence, variation in SST windows did not merely reflect latitudinal variation.

3.4 Selection tests

The FST-based outlier test (Beaumont & Nichols, 1996) involving all populations identified three SNPs (Contig7991, Contig11261 and
Contig10740_78) to be high-divergence outliers, whereas seven SNPs and one phenology-related locus Ots515NWFSC showed lower F_{ST} than expected under neutrality (Figure S2a). The hierarchical outlier test (Excoffier et al., 2009) involving only populations from Western Greenland identified only Contig10740_78 as a high-divergence outlier and also again identified Ots515NWFSC as a low-divergence outlier along with two SNPs (Figure S2b). The results for Ots515NWFSC are likely to reflect the higher allelic diversity (microsatellite; 24 alleles) relative to bi-allelic SNPs. Hence, its outlier status is assumed to represent differences in mutation rate between microsatellites and SNPs rather than evidence for balancing selection. The absence of clearly identifiable selection was also evident from the landscape outlier test analyses using the method by de Villemereuil et al. (2015). Hence, there were no significant associations between any of the loci and (a) latitude, (b) start of SST window, (c) end of SST window and (d) duration of SST-window. Also, none of the loci were outliers without association with environmental parameters (data not shown). In order to rule out that there was an issue with including highly polymorphic loci and bi-allelic SNPs in the outlier tests, they were repeated including only Cryptochrome3 and OtsClock1b (each showing four alleles) along with the SNPs. However, this did not lead to identification of more outliers (data not shown).

The above outlier tests only consider allele frequencies, whereas functional variation at OtsClock1b consists of the number of polyQ repeats, that is, the length of alleles. At the scale of all populations (landlocked and anadromous), there was no significant association between mean allele length at OtsClock1b and latitude (Table 2; Figure S3a), and this was also the case at the scale of all anadromous populations from Greenland and at the scale of anadromous populations from Western Greenland, that is omitting the population SCOR-1 from Eastern Greenland (see Table 2). Across all anadromous populations from Greenland, there was also no significant association between mean allele length and both SST window start date, end date or duration (Table 2, Figure S3b-d). At the scale of anadromous populations from Western Greenland, there was, however, a positive association between mean allele length and both SST window start date or duration (Table 2 and Figure S3e-f), though we note that SST window start date and duration were strongly correlated and hence cannot be considered independent (y = −0.567x + 229.738, R^{2}_{adjusted} = .762, p = 1.38 \times 10^{-5}).
large-scale genetic differentiation among European landlocked char populations has been reported (Wilson et al., 2004), the present study represents a first assessment of genetic variation and structure at nuclear loci in anadromous Arctic char on a large geographical scale. Genetic variation at SNPs was clearly lower in the two landlocked populations than in the majority of anadromous populations, reflecting well-established patterns of variation observed across marine, anadromous and freshwater fish species and populations (Martinez et al., 2018; Ward, Woodwark, & Skibinski, 1994).

Focusing exclusively on SNP variation in anadromous populations in Western Greenland, the hierarchical AMOVA showed stronger differentiation among regional groups of populations as compared to differentiation among populations within groups. Along with the distinct clustering of populations according to geography in the DAPC analysis, the highly significant isolation by distance and the outcome of the spatial autocorrelation analysis, this provides evidence for a system connected by gene flow and with geographical distance as a major factor influencing genetic divergence. This could in principle represent a true hierarchical structure with distinct groups of local populations, or it could represent a continuous structure with isolation by distance, with the seemingly hierarchical structure reflecting an artefact due to gaps in the geographical coverage of sampling. The fact that strong isolation by distance was observed and points did not separate into different clusters (Figure 3a), which could otherwise indicate genetic breaks, favours the latter option. As a whole, the genetic structure of anadromous char populations along the Western Greenland coast is congruent with previous studies focusing on smaller geographical regions (Bernatchez, Dempson, & Martin, 1998; Christensen et al., 2018; Harris, Moore, Bajno, & Tallman, 2016; Harris, Moore, Galpern, Tallman, & Taylor, 2013; Moore et al., 2017; Moore, Harris, Tallman, & Taylor, 2013).

Christensen et al. (2018) analysed historical (DNA extracted from otoliths and scales from the 1950s) and contemporary samples from a subset of the anadromous populations included in this study (NUUK-1, NUUK-2, NUUK-4 and QAQO-2), and they found that the genetic structure was remarkably stable over time. Moreover, using a temporal method for estimating effective population size ($N_e$) and migration rate (m) (Wang & Whitlock, 2003), they found $N_e$ point estimates to exceed 500 in most populations and m to be at most 0.058. Based on the temporal stability, the estimated $N_e$ and m values and a model incorporating the relative importance of genetic drift, gene flow and strength of selection (Yeaman & Otto, 2011), it was suggested that anadromous Arctic char populations have the potential to be locally adapted (Christensen et al. (2018); see also Moore et al. (2013) and Santaquiteria, Svenning, and Praebel (2016)). This is certainly likely to be the case for populations distributed across the > 1,500 km geographical span along the Western Greenland coast, encompassing considerable climatic and other environmental variation. Climate change in the Arctic is in general expected to lead to a northward shift of climate regimes,
with southern populations being adapted to climate conditions that more northern populations will experience in the future, although the situation appears more complex for SST regimes and possible associated adaptation (see below). Does this mean that possible adaptive genetic variation could move across populations by gene flow, leading to future evolutionary rescue of populations.
maladapted to altered climatic conditions (Gonzalez et al., 2013)? The pronounced isolation by distance suggests that populations across the range are indeed connected. This is further supported by the genetic patch size of 450 km estimated by spatial autocorrelation analysis; although it is difficult to interpret this value directly in terms of gene flow, it does suggest connectivity among populations over long geographical distances. Hence, evolutionary rescue is possible, although the results do not inform about the rate at which beneficial variation for evolutionary rescue could disperse into increasingly maladapted populations affected by climate change.

### 4.2 Variation at phenology-related loci

The Arctic char populations of this study represented habitats showing strong variation in latitude and thereby photoperiod and sea surface temperature, the latter visualised by SST windows in Figure 5. Although it is often argued that Arctic char have only a short annual period available for foraging in the sea in some parts of their distribution range (Moore et al., 2017), in Greenland the time periods where sea surface temperature exceeded 2°C in fact varied from a few weeks to several months, leaving ample opportunity for local adaptation to this crucial environmental factor. Yet, the evidence for selection acting on the phenology-related loci was mixed.

The outlier tests applied (Beaumont & Nichols, 1996; Excoffier et al., 2009; de Villemereuil et al., 2015) suggested only one of the SNPs (Contig10740_78) to be a consistent high differentiation outlier, and none of the phenology-related candidate loci were indicated to be under divergent selection. It is possible that the choice of bi-allelic SNPs as supposedly neutral baseline loci was suboptimal, as two of the phenology-related loci showed twenty-four (Ots515NWFS) and seven (Cryptochrome2b2) alleles respectively. On the other hand, Cryptochrome3 and OtsClock1b each showed only four alleles and overall low heterozygosity within populations. Hence, using multiallelic microsatellite loci as a neutral background would not have been appropriate in such cases. Therefore, it cannot be ruled out entirely that some of the loci are in reality under selection, but that the outlier tests failed to detect this.

The tests incorporating allelic lengths at OtsClock1b, thereby reflecting functional polyQ repeat variation, showed no significant association between mean allelic length and latitude, as otherwise reported in Chinook and Chum salmon (O’Malley, Ford, et al., 2010; O’Malley et al., 2013). However, we did observe significant association between OtsClock1b mean allelic length and start date of SST window or total duration of the SST window, whereas no association was revealed for SST window end date. It is puzzling that the associations became nonsignificant when the geographically remote populations SCOR-1 from Eastern Greenland was included. One possibility may be due to phylogeographic complexity; mitochondrial DNA representing the two distinct Arctic and Atlantic phylogeographic lineages have previously been documented in Western Greenland, presumably reflecting postglacial secondary contact (Brunner et al., 2001; Moore et al., 2015). Preliminary results based on mitogenome sequencing suggest that SCOR-1 belongs exclusively to the Atlantic lineage, and hence, allele lengths at OtsClock1b might not be functionally equivalent to alleles from Western Greenland (where both the Arctic and Atlantic phylogeographic lineages are found). A second possibility is that the sea surface temperature regime in SCOR-1 is distinctly different and not comparable to those of Western Greenland populations, as the start date of the SST window is considerably later than in other populations (Figure 5, Table S4).

Under the assumption that the association between OtsClock1b mean allelic length and start date of SST windows represents a genuine biological signal, then this would suggest adaptation to emigrate from freshwater to the sea at the time that marine temperature

<table>
<thead>
<tr>
<th>Parameter tested</th>
<th>Geographical scale</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latitude</td>
<td>All populations</td>
<td>$y = 1.44x + 308.02$, $R^2_{\text{adjusted}} = .08, p = .129$</td>
</tr>
<tr>
<td>Latitude</td>
<td>Anadromous populations, Eastern and Western Greenland</td>
<td>$y = 1.38x + 311.32$, $R^2_{\text{adjusted}} = .06, p = .175$</td>
</tr>
<tr>
<td>Latitude</td>
<td>Anadromous populations, Western Greenland</td>
<td>$y = 1.62x + 296.84$, $R^2_{\text{adjusted}} = .11, p = .128$</td>
</tr>
<tr>
<td>SST window start date</td>
<td>Anadromous populations, Eastern and Western Greenland</td>
<td>$y = 0.29x + 359.18$, $R^2_{\text{adjusted}} = .17, p = .062$</td>
</tr>
<tr>
<td>SST window start date</td>
<td>Anadromous populations, Western Greenland</td>
<td>$y = 0.46x + 334.82$, $R^2_{\text{adjusted}} = .39, p = .007$</td>
</tr>
<tr>
<td>SST window end date</td>
<td>Anadromous populations, Eastern and Western Greenland</td>
<td>$y = -0.20x + 459.81$, $R^2_{\text{adjusted}} = -.01, p = .365$</td>
</tr>
<tr>
<td>SST window end date</td>
<td>Anadromous populations, Western Greenland</td>
<td>$y = -0.27x + 483.70$, $R^2_{\text{adjusted}} = .04, p = .238$</td>
</tr>
<tr>
<td>SST window duration</td>
<td>Anadromous populations, Eastern and Western Greenland</td>
<td>$y = -0.17x + 425.95$, $R^2_{\text{adjusted}} = .12, p = .100$</td>
</tr>
<tr>
<td>SST window duration</td>
<td>Anadromous populations, Western Greenland</td>
<td>$y = -0.26x + 441.42$, $R^2_{\text{adjusted}} = .308, p = .019$</td>
</tr>
</tbody>
</table>
regimes become favourable. Such adaptations would be highly important for making full use of the potential for foraging in the sea, a crucial factor in growth and survival (Jensen et al., 2018). Whereas there was also a significant association between mean allele length SST window duration, the strong correlation between start date and SST window duration raises questions about the specific parameter involved. The duration of SST window is defined by the start and end date of the window, and as there was no significant association between mean allele length and end date, then this would suggest that it is really the start date that is the parameter of biological significance.

It is somewhat surprising that no association was found with end date of SST window, as studies of other salmonids have documented association between OtsClock1b and run and/or spawning time variation (O’Malley et al., 2014, 2013; O’Malley, Ford, et al., 2010). However, most SST window end dates occurred later than the assumed time of spawning; in some cases (QAQO-1 and QAQO-2) as late as mid-November, whereas spawning is expected to take place no later than early October. The optimal time of spawning must be assumed to be primarily determined by temperature, water flow and other factors in the freshwater environments although conditions in the sea might also play a role, such as temperature affecting maturation. Hence, specific data on spawning time would be required for directly testing its association with OtsClock1b variation.

In total, the results did not show association between OtsClock1b allele length and latitude, but rather an association with SST regimes. Due to the influence of the West Greenland Current (Lloyd et al., 2007), SST regimes do not simply reflect latitude, but are generally highest in a broad region ranging from NUUK-1-5 to DISK-1 (see Figure 1). It is possible that for other traits and genes associated with selection in the freshwater environments, more clear-cut association with latitudinal variation would be found.

5 | CONCLUSIONS

The study documented strong genetic differentiation among Arctic char, including the most intensively sampled region along the Greenland West Coast. A significant pattern of isolation by distance was observed among Western Greenland anadromous populations, indicating connectivity and an absence of clear genetic breaks. At most phenology-related loci, no evidence for selection was observed, but in Western Greenland anadromous populations association was observed between mean allele length at OtsClock1b and the start date of the time window during which sea surface temperature exceeded 2°C, along with the duration of this time window. This suggests potentially important adaptations to geographical variation in sea surface temperatures and the optimal time of year for migrating to sea. At the same time, ongoing climate change is expected to affect sea surface temperature regimes, possibly causing current adaptations to become maladaptive in the future. The occurrence of gene flow among anadromous populations would facilitate redistribution of functionally important alleles at OtsClock1b across populations, for example from the populations DISK-1, KANG-1 and SISI-1 experiencing early onset of the SST window, towards northern populations like UUMM-1, UUMM-2 and QAAN that currently are subject to late onset of the SST window but may experience future earlier onset as a result of climate change. Hence, this could provide possibilities for evolutionary rescue in a rapidly changing environment, at least for phenological traits.

ACKNOWLEDGEMENTS

The authors thank Shenglin Liu, Rasmus Hedeholm, Lars Heilman, Anne-Laure Fercaud, John Fleng Steffensen, Terkel Broe Christensen, Jan Nielsen and Nynne Hjort Nielsen for assistance with collecting samples, Annie Brandstrup for technical assistance, two anonymous reviewers for comments and suggestions and The Danish Council for Independent Research, Natural Science for funding (grant no. 1323-00158A to Michael M. Hansen).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHORS’ CONTRIBUTION

Conceived and designed the investigation: MMH, RPAM, MWJ, LB, DJF, RN and KGO. Performed field and/or laboratory work: RPAM, MWJ, MMH, LB, DJF, KP, RN, BJ and JMP. Analysed the data: RPAM, MMH and MWJ. Contributed materials, reagents and/or analysis tools: MMH. Wrote the paper: RPAM, MMH and MWJ with contributions from LB, DJF, KP, KGO, RN, BJ and JMP.

DATA AVAILABILITY STATEMENT

Raw genotype data in Genepop format have been deposited in DRYAD https://doi.org/10.5061/dryad.sc30mr1 (Madsen et al., 2019).

ORCID

Michael M. Hansen https://orcid.org/0000-0001-5372-4828

REFERENCES


**SUPPORTING INFORMATION**

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

**How to cite this article:** Madsen RPA, Jacobsen MW, O’Malley KG, et al. Genetic population structure and variation at phenology-related loci in anadromous Arctic char (*Salvelinus alpinus*). *Ecol Freshw Fish*. 2020;29:170–183. https://doi.org/10.1111/eff.12504