Genetic diversity and connectivity within Mytilus spp. in the subarctic and Arctic

Mathiesen, Sofie Smedegaard; Thyrring, Jakob; Hansen, Jakob Hemmer; Berge, Jørgen; Sukhotin, Alexey; Leopold, Peter; Bekaert, Michaël; Sejr, Mikael Kristian; Eg Nielsen, Einar

Published in: Evolutionary Applications

Link to article, DOI: 10.1111/eva.12415

Publication date: 2017

Link back to DTU Orbit


General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Genetic diversity and connectivity within Mytilus spp. in the subarctic and Arctic

Sofie Smedegaard Mathiesen1,2 | Jakob Thyrring1 | Jakob Hemmer-Hansen2 | Jørgen Berge3,4 | Alexey Sukhotin5,6 | Peter Leopold3 | Michaël Bekaert7 | Mikael Kristian Sejr1,† | Einar Eg Nielsen2,†

Abstract
Climate changes in the Arctic are predicted to alter distributions of marine species. However, such changes are difficult to quantify because information on present species distribution and the genetic variation within species is lacking or poorly examined. Blue mussels, Mytilus spp., are ecosystem engineers in the coastal zone globally. To improve knowledge of distribution and genetic structure of the Mytilus edulis complex in the Arctic, we analyzed 81 SNPs in 534 Mytilus spp. individuals sampled at 13 sites to provide baseline data for distribution and genetic variation of Mytilus mussels in the European Arctic. Mytilus edulis was the most abundant species found with a clear genetic split between populations in Greenland and the Eastern Atlantic. Surprisingly, analyses revealed the presence of Mytilus trossulus in high Arctic NW Greenland (77°N) and Mytilus galloprovincialis or their hybrids in SW Greenland, Svalbard, and the Pechora Sea. Furthermore, a high degree of hybridization and introgression between species was observed. Our study highlights the importance of distinguishing between congener species, which can display local adaptation and suggests that information on dispersal routes and barriers is essential for accurate predictions of regional susceptibility to range expansions or invasions of boreal species in the Arctic.

KEYWORDS
arctic fauna, bivalves, climate change, glacial refugium, hybrid zone, Mytilus edulis, population structure, SNPs

1 INTRODUCTION
Nowhere else on Earth is the impact of climate change expected to be more severe than in the Arctic. Temperatures in the Arctic are estimated to increase by 4–7°C over the next century, with wide-ranging effects on Arctic species (ACIA 2004; IPCC 2014). This has caused shifts in species' abundances and distributions over the last decades (IPCC 2014; Poloczanska et al., 2013), and future temperature increases are believed to move species distribution limits toward the poles (ACIA 2004). Such effects, however, are nearly impossible to monitor and understand without proper baseline studies of the genetic variation within and between species (Brodersen & Seehausen, 2014). Almost all species investigated, including Arctic, have revealed genetically discrete populations that inhabit a specific subset of the species geographical and environmental range. These populations

†These authors jointly supervised this work.
can exhibit different adaptations and tolerance limits to specific environments (Limborg et al., 2012; Nielsen, Hemmer-Hansen, Larsen, & Bekkevold, 2009; Thyrring, Rysgaard, Blitch, & Sejr, 2015), which make it important to know their current distribution and the connectivity between populations and the processes governing the distribution of genetic variation. Through genetic analysis, it is possible to determine the level of genetic variability within both threatened and newly established populations, the origin of migrating individuals, direction of gene flow, and possible adaptive evolutionary changes associated with climate change (Brodersen & Seehausen, 2014; Hansen, Olivieri, Waller, & Nielsen, 2012; Laikre, Schwartz, Waples, & Ryman, 2010). All key factors needed to make predictions for the likely impact of climate change.

Bivalves of the genus Mytilus are frequently used as environmental indicators, as they are semi-sessile, have a relatively long life span, and are widely distributed in coastal regions in both Northern Hemisphere and Southern Hemisphere (Goldberg, 1986; Gosling, 2003; Rainbow, 1995; Thyrring, Juhl, Holmstrup, Blitch, & Sejr, 2015). Mytilus spp. are commercially and ecologically important species and often a dominant part of the intertidal and shallow subtidal fauna. Therefore, numerous studies of their responses to various stressors (e.g., temperature, salinity, pollutants) have been performed (Gosling, 2003; Jones, Lima, & Wethey, 2010; Mubiana, Qadah, Meys, & Blust, 2005; Sendergaard, Asmund, Johansen, & Riget, 2011; Wanamaker et al., 2007). Furthermore, Mytilus spp. have already demonstrated adaptations to different environments (Blitch, Sejr, & Høgslund, 2013; Thyrring, Rysgaard, et al., 2015) and a shift in their southern geographical range caused by increasing temperatures (Jones et al., 2010), making them an excellent model for inferring how species distributions might change in response to climate change. Additionally, Mytilus spp. have been the subjects of genetic studies for decades as the different species are morphologically difficult to distinguish. Consequently, the population structure of individual Mytilus species has been difficult to establish. Mytilus edulis L. 1758, Mytilus trossulus Gould 1850 and Mytilus galloprovincialis Lmk. 1819, all belong to the M. edulis species complex and are known to coexist and hybridize with conflicting patterns on the fitness for hybrids. Some studies did not observe any depressed fitness (Doherty, Brophy, & Gosling, 2009; Koehn, 1991; Riginos & Cunningham, 2005; Toro, Thompson, & Innes, 2006), while others (Gardner & Thompson, 2001; Toro, Innes, & Thompson, 2004; Toro, Thompson, & Innes, 2002; Tremblay & Landry, 2016) found a difference in fitness between parental types and hybrids and backcrosses. These findings and numerous studies on introgression between them (Fraisse, Belkhir, Welch, & Bierne, 2016; Roux et al., 2014) have challenged the isolation species concept (White, 1978); however, they are generally considered to be different species, as they remain ecological distinct despite semipermeable barriers for gene flow and introgression (Bierne, Borsa, et al., 2003; Fraisse, Roux, Welch, & Bierne, 2014; Saarman & Pogson, 2015). Mytilus trossulus is thought to have invaded the Arctic Ocean from the Pacific Ocean around 3.5 million years ago (mya) through the Bering Strait (Rawson & Hilbish, 1995, 1998; Vermeiji, 1991). As the Bering Strait closed during glacial periods, allopatric speciation resulted in the evolution of M. edulis in the Atlantic. Mytilus edulis has since spread to large parts of the Atlantic and due to apparent low gene flow (at least for some loci); M. edulis populations on each side of the Atlantic are genetically distinct (Riginos & Henzler, 2008; Riginos, Hickerson, Henzler, & Cunningham, 2004). Speciation between M. edulis and M. galloprovincialis most likely occurred through allopatric isolation approximately 2.5 mya (Quesada, Gallagher, Skibinski, & Skibinski, 1998; Rawson & Hilbish, 1995, 1998) with secondary contact and introgression occurring around 0.7 mya (Roux et al., 2014). Between interglacial periods 46,000 and 20,000 years ago, M. trossulus reinvaded the Arctic Ocean (Rawson & Harper, 2009). From here, it invaded both sides of the Atlantic founding M. trossulus/M. edulis hybrid zones along North American and European coasts (Riginos & Cunningham, 2005).

The geographical distribution and genetic population structure of Mytilus spp. have been intensively studied in boreal and temperate regions (Bierne, Borsa, et al., 2003; Hilbish, Carson, Plante, Weaver, & Gilg, 2002; Sarver & Zolt, 1993; Väinölä & Strelkov, 2011; Westerbom, Klipi, & Mustonen, 2002); however, little is known of their distribution and genetic population structure in the Arctic. In the subarctic and Arctic, M. edulis is considered the most abundant Mytilus species, and it has been recorded in Arctic regions of Russia, along the Norwegian coast, in Iceland and Greenland (Hummel, Colucci, Bogaards, & Strelkov, 2001; Riginos & Henzler, 2008; Sukhotin, Strelkov, Maximovich, & Hummel, 2007; Väinölä & Strelkov, 2011). In Greenland, Mytilus spp. populations are found all along the west coast, and southern populations from Tartoq and Narsarsuaq have been shown to be genetically distinct from European M. edulis displaying higher resemblance to Canadian and North American M. edulis populations (Riginos & Henzler, 2008). Few genetic analyses have been performed on Mytilus spp. in Greenland, and most studies have assumed these mussels to be M. edulis without genetic verification despite observations of variations in metabolic response to low temperatures between populations from NW and SW Greenland (Thyrring, Rysgaard, et al., 2015). Moreover, in 2004, subtidal M. edulis were discovered at the mouth of Isfjorden in Svalbard after 1,000 years of absence (Berge, Johnsen, Nilsen, Gulliksen, & Slagstad, 2005). These mussels were hypothesized to have been transported from Norway by the West Spitsbergen Current in 2002, but their origin has never been confirmed through genetic analysis. Mytilus trossulus is less common in Arctic waters. Väinölä and Strelkov (2011) found that M. trossulus had a scattered distribution in the White Sea and the Norwegian Sea, and Feder, Norton, and Geller (2003) found live M. trossulus in Arctic Alaska in the 1990s. Furthermore, Wenne, Bach, Zbawicka, Strand, and McDonald (2016) has recently reported a NW Greenlandic fjord at Maarmorilik (71°N) to be inhabited by M. edulis, M. trossulus, and their hybrids. Mytilus edulis and M. trossulus hybrid zones have also been found and studied on the European and N American Atlantic coasts. Riginos and Cunningham (2005) reviewed the literature on the subject to look at local adaptation and species segregation and found conflicting patterns of species segregation across the Atlantic. In the western Atlantic, M. trossulus was found on wave-exposed open coasts, whereas M. edulis appeared to dominate in sheltered areas of low salinity. However, European M. trossulus populations from the Baltic Sea appeared to be locally adapted
to the prevailing low salinities. The latter is in line with the findings of Wenne et al. (2016), who found a higher prevalence of *M. trossulus* in the inner Maarmorilik fjord compared with the more saline outer fjord. *Mytilus galloprovincialis* normally inhabits warmer waters, but in recent years the species and *M. galloprovincialis/M. edulis* hybrids have been observed along the Norwegian coast (Brooks & Farmen, 2013; Riginos & Henzler, 2008). This could be related to human activities like ship traffic in rural areas enabling faster invasion of waters otherwise not directly accessible to them (Anderson, Bilodeau, Gilg, & Hilbish, 2002; Geller, Carlton, & Powers, 1994). Furthermore, it has been demonstrated that *M. galloprovincialis* is capable of tolerating lower water temperatures (Inoue et al., 1997), highlighting the potential for this species to occur in the Arctic.

Most studies on *Mytilus* spp. have focused on a few allozymes, mtDNA markers, or microsatellites (Bierne, Daguin, Bonhomme, David, & Borsa, 2003; Brooks & Farmen, 2013; Feder et al., 2003; McDonald, Seed, & Koehn, 1991; Ouagajjou, Presa, Astorga, & Pérez, 2011; Presa, Perez, & Diz, 2002). However, in recent years the use of single nucleotide polymorphisms, SNPs, has become increasingly popular to answer questions about *Mytilus* spp. status, population structure, hybridization, and adaptive variation (Helyar et al., 2011; McGinnity, Cross, Crozier, & Prodohl, 2013) was used to calculate diversity in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveRsity (Keenan, 2015) was used to test each locus in each sample for departure from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium (LD) for each locus pair in each sample (10,000 dememorizations, 100 batches, and 5,000 iterations). Within samples, the program diveR
in 81 loci being retained for further analyses. As most SNP loci were developed from *M. edulis*, reliable scoring of *M. trossulus* individuals was not possible for three loci (174302_A, 67577_A, and 31051_A), and analyses concerning hybrid identification were performed for 78 SNP loci only.

### 2.5 Identification of hybrids

Based on the generated pairwise $F_{ST}$ estimates, the grouping of samples was visualized in a multidimensional scaling plot applying the cmdscale function in R (R Core Team, 2015). Additionally, a principal component analysis scatter plot was created in the R package Adegenet (Jombart, 2008; Jombart & Ahmed, 2011) to illustrate the genetic relationships among individuals across all samples. Structure v2.3.4, utilizing the Bayesian MCMC clustering approach (Pritchard, Stephens, & Donnelly, 2000) was used to visualize species integrity and identify possible hybridization among *Mytilus* spp. using a variable number of predefined clusters ($K$) for grouping individuals. This was also done to positively identify *M. edulis* individuals and subsequently create a reduced data set exclusively aimed at investigating population structure within this species. Considering the close genetic resemblance of *Mytilus* spp. and the assumed low gene flow between geographically distant samples (Riginos & Henzler, 2008), simulations were run for a number of predefined $K$ values. Based on an initial analysis of $K$ up to 18, we found the highest likelihoods for $K = 3-5$. Accordingly, we used this as the basis to identify the major groupings within the species complex. For all simulations, a burn-in of 10,000 iterations was used followed by 100,000 MCMC repetitions. To evaluate the power of designating individuals as pure or hybrids, we followed the procedure described in Nielsen, Hansen, Ruzzante, Meldrup, and Grønkjær (2003) using the program Hybridlab (Nielsen, Bach, & Kotlicki, 2006). Briefly, we simulated 1,000 individuals of each of the following classes: parentals, F1/F2, and backcrosses. This was done based on the allele frequencies of the reference samples of *M. trossulus*, *M. galloprovincialis*, and *M. edulis* samples identified by initial Structure runs to likely consist exclusively of *M. edulis* and hybrids (see results section), we chose to only include individuals of *M. trossulus* and *M. galloprovincialis*.

### 2.6 Population structure of *Mytilus edulis*

A reduced data set was used to assess the population structure in *M. edulis*. Based on the results from the analysis of simulated parentals and hybrids (see results section), we chose to only include individuals

**TABLE 1** Summary of sample information including sampling location and year, sampling code indicating location and possibly size, estimates of expected ($H_e$), and observed ($H_o$) heterozygosities and the allelic richness

<table>
<thead>
<tr>
<th>Country</th>
<th>Location</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Sampling year</th>
<th>Code</th>
<th>N</th>
<th>Habitat type</th>
<th>$H_e$</th>
<th>$H_o$</th>
<th>Allelic richness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greenland</td>
<td>Qaanaaq, 15–30 mm</td>
<td>77.4650</td>
<td>−69.2403</td>
<td>2014</td>
<td>QAS</td>
<td>30</td>
<td>Intertidal zone</td>
<td>0.09</td>
<td>0.06</td>
<td>1.40</td>
</tr>
<tr>
<td>Greenland</td>
<td>Qaanaaq, &gt;50 mm</td>
<td>77.4650</td>
<td>−69.2403</td>
<td>2014</td>
<td>QAL</td>
<td>30</td>
<td>Intertidal zone</td>
<td>0.07</td>
<td>0.04</td>
<td>1.31</td>
</tr>
<tr>
<td>Greenland</td>
<td>Upernavik</td>
<td>72.7939</td>
<td>−56.1028</td>
<td>2014</td>
<td>UPE</td>
<td>43</td>
<td>Intertidal zone</td>
<td>0.25</td>
<td>0.14</td>
<td>1.73</td>
</tr>
<tr>
<td>Greenland</td>
<td>Nuuk, 15–30 mm</td>
<td>64.1968</td>
<td>−51.7104</td>
<td>2014</td>
<td>NUS</td>
<td>30</td>
<td>Intertidal zone</td>
<td>0.25</td>
<td>0.23</td>
<td>1.64</td>
</tr>
<tr>
<td>Greenland</td>
<td>Nuuk, &gt;50 mm</td>
<td>64.1968</td>
<td>−51.7104</td>
<td>2014</td>
<td>NUL</td>
<td>24</td>
<td>Intertidal zone</td>
<td>0.24</td>
<td>0.21</td>
<td>1.69</td>
</tr>
<tr>
<td>Greenland</td>
<td>Kobbeforder</td>
<td>64.1367</td>
<td>−51.3909</td>
<td>2014</td>
<td>KOB</td>
<td>30</td>
<td>Intertidal zone</td>
<td>0.25</td>
<td>0.22</td>
<td>1.65</td>
</tr>
<tr>
<td>Iceland</td>
<td>Iceland</td>
<td>62.0261</td>
<td>−22.1594</td>
<td>2014</td>
<td>ICE</td>
<td>45</td>
<td>Intertidal zone</td>
<td>0.27</td>
<td>0.27</td>
<td>1.72</td>
</tr>
<tr>
<td>Iceland</td>
<td>Lofoten</td>
<td>68.3380</td>
<td>13.8780</td>
<td>2014</td>
<td>LOF</td>
<td>45</td>
<td>Intertidal zone</td>
<td>0.26</td>
<td>0.24</td>
<td>1.76</td>
</tr>
<tr>
<td>Iceland</td>
<td>Tromsø, 15–30 mm</td>
<td>69.8278</td>
<td>18.9226</td>
<td>2014</td>
<td>TRS</td>
<td>10</td>
<td>Intertidal zone</td>
<td>0.27</td>
<td>0.29</td>
<td>1.76</td>
</tr>
<tr>
<td>Iceland</td>
<td>Tromsø, &gt;50 mm</td>
<td>69.8278</td>
<td>18.9226</td>
<td>2014</td>
<td>TRL</td>
<td>30</td>
<td>Subtidal zone</td>
<td>0.28</td>
<td>0.23</td>
<td>1.64</td>
</tr>
<tr>
<td>Norway</td>
<td>Kongsfjorden</td>
<td>79.1123</td>
<td>11.1362</td>
<td>2012</td>
<td>SV1</td>
<td>21</td>
<td>Subtidal zone</td>
<td>0.26</td>
<td>0.23</td>
<td>1.73</td>
</tr>
<tr>
<td>Norway</td>
<td>Kongsfjorden</td>
<td>79.1123</td>
<td>11.1362</td>
<td>2013</td>
<td>SV2</td>
<td>13</td>
<td>Subtidal zone</td>
<td>0.29</td>
<td>0.24</td>
<td>1.79</td>
</tr>
<tr>
<td>Norway</td>
<td>Kongsfjorden</td>
<td>79.1123</td>
<td>11.1362</td>
<td>2014</td>
<td>SV3</td>
<td>13</td>
<td>Subtidal zone</td>
<td>0.26</td>
<td>0.25</td>
<td>1.70</td>
</tr>
<tr>
<td>Norway</td>
<td>Adventfjorden</td>
<td>78.2381</td>
<td>15.6026</td>
<td>2014</td>
<td>SV4</td>
<td>10</td>
<td>Subtidal zone</td>
<td>0.29</td>
<td>0.24</td>
<td>1.78</td>
</tr>
<tr>
<td>Russia</td>
<td>Pechora Sea, west</td>
<td>69.3563</td>
<td>58.8393</td>
<td>2014</td>
<td>PSW</td>
<td>18</td>
<td>Subtidal zone</td>
<td>0.29</td>
<td>0.29</td>
<td>1.75</td>
</tr>
<tr>
<td>Russia</td>
<td>Pechora Sea, east</td>
<td>69.3204</td>
<td>58.7566</td>
<td>2014</td>
<td>PSE</td>
<td>27</td>
<td>Subtidal zone</td>
<td>0.29</td>
<td>0.29</td>
<td>1.77</td>
</tr>
<tr>
<td>Russia</td>
<td>White Sea, Kandalaksha Bay</td>
<td>66.3372</td>
<td>33.6494</td>
<td>2014</td>
<td>WS1</td>
<td>45</td>
<td>Subtidal zone</td>
<td>0.32</td>
<td>0.24</td>
<td>1.86</td>
</tr>
<tr>
<td>Russia</td>
<td>White Sea, Onega Bay</td>
<td>64.2079</td>
<td>36.6187</td>
<td>2014</td>
<td>WS2</td>
<td>45</td>
<td>Intertidal zone</td>
<td>0.29</td>
<td>0.27</td>
<td>1.77</td>
</tr>
<tr>
<td>Russia</td>
<td>Magadan, Okhotsk Sea³</td>
<td>80.07</td>
<td>15.6026</td>
<td>2007</td>
<td>MTR</td>
<td>15</td>
<td>Subtidal zone</td>
<td>0.11</td>
<td>0.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>

³These samples were only used for reference i.e. a representative of *Mytilus trossulus* and *Mytilus galloprovincialis*.
with admixture proportions below 0.2, as estimated by Structure. This was done to avoid extensive influence of hybridization on estimates of population divergence, but at the same time allowing for statistical uncertainty regarding whether individuals were pure *M. edulis* individuals or not. No significant differentiation was found between sampled mussels of different size classes from the same location (Tromsø, Nuuk, and Qaanaaq). Consequently, they were pooled prior to downstream analyses of population structure. Pairwise $F_{ST}$ estimates were generated with Genepop 4.2, while Structure v2.3.4 was used to estimate the most likely number of genetic clusters. A burn-in period of 50,000 iterations was chosen followed by 100,000 MCMC repetitions for $K$ values 2–4. A hierarchical AMOVA was conducted in Arlequin v.3.5.2.2 (Excoffier & Lischer, 2010) to infer the proportion of genetic variance distributed among the different *M. edulis* clusters and among samples within the clusters detected by Structure (see results section). To visualize the genetically based grouping of *M. edulis* population samples, a multidimensional scaling plot was generated, while a principal component analysis, PCA, was used to illustrate the relationships among *M. edulis* individuals in general and specifically for the Norwegian, Svalbard, and Russian samples to infer the likely origin of Svalbard mussels. The PCA scatter plots were generated in R, using the cmdscale function and the package Adegenet.

### 2.7 | Outlier analysis

To identify loci potentially under selection in the "*Mytilus edulis*" data set, the joint distribution of $F_{ST}$ and heterozygosity under a hierarchical island model of population structure was examined using Arlequin v.3.5.2.2 (Excoffier & Lischer, 2010) based on the method in Excoffier, Hofer, and Foll (2009). Accounting for the hierarchical population structure reduces the probability of false discoveries (Excoffier et al., 2009). Samples were grouped according to the genetic clustering analyses: (i) Greenlandic samples, (ii) Samples from Norway, the Svalbard archipelago, and Russian waters, and (iii) the Icelandic sample (see the section under Results subsection "Population structure of *M. edulis*"). The analytical settings for generating 95% and 99% confidence intervals were 20,000 simulations, 100 demes per group, and 10 groups. Loci outside the 95% quantile were considered possible subjects to selection, as these loci deviate more than could be expected under a model of neutral population structure. From this analysis, an exclusive

---

**FIGURE 1** Map showing the different sampling locations and the proportion of three different *Mytilus* species and inferred hybrids at each location. For explanation of sample identification codes, see Table 1. Unidentified individuals denote apparent hybrids of all three *Mytilus* spp.
“outlier” data set and a “neutral” data set were created to test the importance of outlier loci for defining the inferred population structure of *M. edulis*; that is, the true connectivity among populations based on neutral processes (drift and migration) could be obscured by loci under divergent selection. For both data sets, overall and pairwise $F_{ST}$ estimates were generated in Genepop 4.2, while Structure v2.3.4 with $K = 2$ (using settings as above) was used to investigate whether the population structure found in *M. edulis* based on all loci could be identified from both the “outlier” and “neutral” data sets, or whether they displayed contrasting patterns.

3 | RESULTS

3.1 | Summary statistics

Three loci (159069_A, 171383_A, and 170478_A) deviated significantly from HWE in ten samples or more, and they were discarded from further analyses.

In total, 73,206 pairwise tests for LD between loci within samples were performed of which 3,194 tests were significant (4.36%). On average, 154 of 3,485 tests were significant within samples (range 0–986 significant tests). Only three SNP pairs displayed significant LD in more than five samples: 137120_A x BM8E (significant in six samples), and 175018_A x 40154_A (significant in eight samples), and 175018_A x 241544_A (significant in 12 samples). Subsequently, one locus from each of the coupled SNP pairs was discarded (BM8E, 241544_A and 40154_A) to eliminate effects of linkage on downstream analyses.

Allelic richness ranged from 1.00 to 1.86 (Table 1), with the lowest values in the Qaanaaq samples (QAS and QAL). The levels of $H_e$ and $H_o$ (Table 1) ranged from 0.14 to 0.32 in all samples except for QAS and QAL, which had particularly low values ranging from 0.04 to 0.09. In all samples, $H_e$ was close to $H_e$ except for the Upernavik sample (UPE), where $H_e$ and $H_0$ was 0.25 and 0.14, respectively, and one of the White Sea samples (WS1) with $H_e$ and $H_o$ of 0.32 and 0.24. In the reduced data set, with samples consisting only of inferred *M. edulis*, individuals provide estimates of allelic richness between 1.57 and 1.67 (Appendix 1) and $H_e/H_o$ values ranging between 0.22 and 0.30.

The overall $F_{ST}$ across samples was 0.273. The pairwise $F_{ST}$ values ranged between 0 and 0.860 (Table S2) with the highest pairwise $F_{ST}$ value between the *M. trossulus* and *M. galloprovincialis* reference samples. Further, high values were found between *M. galloprovincialis* and the N Greenland samples from Qaanaaq; QAS and QAL (0.738 and 0.774) and between *M. trossulus* and all other samples except the three N Greenland samples (QAS, QAL, and UPE).

3.2 | Identification of hybrids

The multidimensional scaling plot (Fig. 2A) visualizes the genetic differentiation among all samples including the reference samples for *M. trossulus* (MTR) and *M. galloprovincialis* (MGA). The majority of samples clustered together in a “*M. edulis*” cluster. However, the QAS and QAL samples clustered with the *M. trossulus* reference sample, while samples UPE, WS1, and Lofoten (LOF) were located between the three main “species” clusters. UPE and WS1 appeared to be distributed between the *M. trossulus* and *M. edulis* clusters, while LOF was situated between the “*M. edulis*” and “*M. galloprovincialis*” clusters. The clustering of QAS and QAL with MTR, and the inferred separation of UPE, WS1, and LOF from the “*M. edulis*” cluster were further supported by the principal component analysis scatter plot of individual genotypes (Fig. 3A). Most individuals clustered together as a “*M. edulis*” cluster, except for individuals from the UPE and WS1 samples, which appeared to contain individuals distributed between the “*M. trossulus*” and “*M. edulis*” clusters, suggesting that these individuals may be hybrids. The *M. galloprovincialis* reference sample clustered in the proximity of the “*M. edulis*” samples in the multidimensional scaling plot; however, a clear separation between the *M. edulis* and *M. galloprovincialis* samples was still apparent (Fig. 2A).

The Structure clustering analysis for $K = 4$ (Fig. 4, for $K = 3$ and 5 see Appendix 2) showed that the clusters make biologically sense as they corresponded to *M. trossulus*, *M. galloprovincialis*, Greenlandic *M. edulis*, and other *M. edulis*. This configuration also allowed the identification of *M. edulis/M. galloprovincialis* or *M. edulis/M. trossulus* hybrids. This was supported by the Structure analysis including the simulated parentals and hybrids, which showed relatively narrow 95% confidence intervals for the simulated *M. edulis* parentals regardless of their geographical origin (0.89–0.99 *M. edulis* ancestry for Greenland and 0.92–0.99 for the other *M. edulis*, see Appendix 3). Likewise, the simulated *M. trossulus*, *M. galloprovincialis* parentals suggested very high power for identifying *M. edulis/M. galloprovincialis* or *M. edulis/M. trossulus* hybrids. However, as only two relatively small samples of *M. trossulus* and *M. galloprovincialis* provided the foundation for the simulations, we conservatively chose an admixture level of 20% as the cutoff point between *Mytilus* spp. parentals and *M. edulis/M. galloprovincialis* or *M. edulis/M. trossulus* hybrids. This was done in order to allow for uncertainty caused by population structure and missing genotypes within the samples of real individuals. This approach enabled the construction of an inclusive “*Mytilus edulis*” data set. When using $K = 3$ the analysis was unable to split the samples into the three a priori defined species groups (Appendix 2B).

The inferred proportion (using the 20% criterion) of the different *Mytilus* species and hybrids in each of the geographical samples (Fig. 1) show that *M. edulis* is the most common species within the sampled subarctic and Arctic populations, where pure *M. edulis* specimens constitute approximately 66% of all sampled individuals. Pure *M. edulis* were present in all samples except for QAS and QAL, which were mainly *M. trossulus* (87%–90%), with few individuals (3–4) showing evidence of *M. edulis* hybridization. Only two samples, UPE and WS1, contained both pure *M. edulis* and *M. trossulus* individuals. The UPE sample contained approximately 51% *M. trossulus* and 33% *M. edulis* and 14% *M. edulis/M. trossulus* hybrids, while the WS1 sample was comprised of 9% *M. trossulus*, 80% *M. edulis*, 9% *M. edulis/M. trossulus* hybrids, and 2% *M. edulis/M. galloprovincialis* hybrids. The distribution of *M. galloprovincialis* individuals is mainly restricted to samples from the Norwegian coast and the Svalbard archipelago (34 and 4 individuals, respectively). A single apparent *M. galloprovincialis* individual was found in the sample of large mussels from Nuuk (NUL).
The LOF sample contained the highest number of *M. galloprovincialis* observed—64% and further 11% *M. edulis* and 22% *M. edulis*/*M. galloprovincialis* hybrids. In cases where more than a few hybrids were found, the distribution of admixture estimates of real individuals was compared to the simulated hybrids. In all cases, different classes of hybrids (F1, F2, and backcrosses) were suggested. However, as explained above the comparison of real and simulated individuals should be interpreted with caution.
3.3 | Population structure of *Mytilus edulis*

The overall $F_{ST}$ for all samples identified as *M. edulis* was 0.048. Pairwise $F_{ST}$ values ranged from 0 to 0.113 with the highest values between the Greenlandic samples and the Norwegian, Svalbard, and Russian samples (Table S3). The lowest $F_{ST}$ values were found between geographically proximate samples such as the two White Sea samples (WS1 and WS2) and the two sampling sites in Svalbard (SVA and SV4). For sites with samples of different size classes, $F_{ST}$ estimates ranged between 0.001 for the Nuuk samples (NUS and NUL) and 0.018 for the Tromsø samples (TRS and TRL) (Supplementary Table S3). The low $F_{ST}$ for Nuuk samples indicates short-term temporal stability of genetic population structure. The higher $F_{ST}$ estimate for Tromsø mussels was not significant, thus allowing the pooling of size classes for the downstream analyses.

The cluster analysis of the "*Mytilus edulis*" data set ($K = 2–4$) showed a clear clustering of samples, essentially separating the Greenlandic samples from the other samples (Fig. 5). The likelihood of $K = 2$ was highest splitting the *M. edulis* samples into two groups; the Greenlandic samples versus the Norwegian, Svalbard, and Russian samples and identifying the Icelandic sample a mixture of eastern and western Atlantic gene pools (Fig. 5A). The plots for $K = 3$ and $K = 4$ added no additional biologically sensible information.

The hierarchical AMOVA for the three groups (Greenlandic, Icelandic, and Norwegian/Svalbard/Russian) provided an estimated variance of 5.78% among groups and 0.43% among samples within groups. The multidimensional scaling plot of population samples (Fig. 2B) and the principal component analysis scatter plot of individual genotypes (Fig. 3B) further supported the population structure of *M. edulis* inferred by Structure with three groups: (i) Greenlandic samples, (ii) Norwegian, Svalbard, and Russian samples, and (iii) the Icelandic sample found between the two main clusters inferred by axis 1.

The principal component analysis scatter plot including only Norwegian, Svalbard, and Russian samples (Fig. 3C) did not provide a clear separation of individuals as these individuals were scattered with no apparent pattern.
3.4 | Outlier analysis

The outlier tests identified six loci as $F_{ST}$ outliers, with six loci significant at the 5% level and three at the 1% level. All of these outliers are high $F_{ST}$ outliers (Fig. 6) indicating diversifying selection (Beaumont & Nichols, 1996), although a few of them could represent the upper tail of the neutral $F_{ST}$ distribution. Also, a strong genetic cline as observed here is known to sometimes overestimate the number of loci under diversifying selection (Strand, Williams, Oleksiak, & Sotka, 2012). Furthermore, introgression between *Mytilus* spp. has been found to cause high $F_{ST}$ outliers (Gosset and Bierne 2012). Pairwise $F_{ST}$ values ranged from 0 to 0.059 for the “neutral” data set and from 0 to 0.474 for the “outlier” data set (Tables S4 and S5). The Structure analyses for both the “neutral” and “outlier” data set also supported the initial population structure separating Greenlandic samples from the Norwegian, Svalbard, and Russian samples and with the Icelandic sample of admixed origin (Appendices 4 and 5).

4 | DISCUSSION

4.1 | Distribution of *Mytilus* spp. in the Arctic

Baseline information of species distribution and their genetic composition is imperative in order to quantify the impacts of climate change on species distribution ranges, biodiversity, and the effects of hybridization between species and populations (Gardner, Zbawicka, Westfall, & Wenne, 2016). Molecular genetic knowledge is a key measure to identify the distribution of invasive congener species (Geller, Darling, & Carlton, 2010), which may cause cascading ecosystem effects. Despite congener species appearing morphologically similar, interspecific variation in ecology and physiology may impact population fitness (Fly & Hilbish, 2013; Fraisse et al., 2016; Somero, 2005). In the Arctic, baselines studies on genetic variation and species abundance are largely absent but urgently needed (Bluhm et al., 2011; Wassmann, Duarte, Agusti, & Sejr, 2011). Pioneer work should therefore focus on keystone model species (such as *Mytilus*), because of their disproportionally large effect on their environment.

*Mytilus* spp. were found pan-Arctic (although only one individual of *M. galloprovincialis* was identified in Greenland). Generally, *M. edulis* was the most common species making up approximately 66% of all sampled individuals. The biogeographic structures of the three *Mytilus* spp. reflect the major current systems of the region. Pure populations were mainly found in regions (such as W Greenland and the Pechora Sea) with a lower influence of Pacific and Atlantic water, than other sampling sites. Northwards currents from boreal waters facilitate larval dispersal from southern populations (Berge et al., 2005; Renaud, Sejr, Bluhm, Sirenko, & Ellingsen, 2015). For instance, the northward flowing current regimes (such as the Norwegian Current) allows non-Arctic species to extend their range into the Arctic from the Atlantic or Pacific Ocean (Bluhm et al., 2011; Fetzer & Arntz, 2008; Sirenko & Gagaev, 2007). Ocean currents also explain why *Mytilus* spp. remain absent in NE Greenland despite the environmental resemblance of NW Greenland with regard to temperatures and ice conditions (Sejr, Blicher, & Rysgaard, 2009). In general, the NE Greenland shelf

---

**FIGURE 6** $F_{ST}$ outlier analyses in Arlequin v3.5.1.3 utilizing the hierarchical island model. Black solid dots denote loci, and gray dashed and dotted lines indicate 95% and 99% confidence intervals, respectively. Loci outside the 95% and 99% confidence intervals are suggested to be under selection.
is considered biogeographically different from the rest of Greenland (Piepenburg et al., 2011). The absence of Mytilus mussels in NE Greenland is likely a result of dispersal barriers due to the lack of an downstream source population, as the East Greenland Current flows from north to south, exemplifying how outflow shelves may respond slowly to climatic changes (Renaud et al., 2015). This is further supported by the presence of Mytilus mussels in SE Greenland, at Tasilaq (Ammassalik, 65°N) (Ockelmann, 1958), which is influenced by a branch of the Irminger Current from the Atlantic Ocean.

The present study highlights the need for further genetic studies in the region as a M. trossulus population was found in the most northern sampled region of NW Greenland (77°N) with M. edulis populations residing in SW Greenland (64°N). This discovery was unexpected, as a seemingly established M. trossulus population has not been found in the high Arctic prior to this study. Several possible mechanisms could explain the presence of M. trossulus in Qaanaaq and Upernavik. First, these populations could have survived in a refugium near NW Greenland during the last glacial period. Glacial refugia are known from North Atlantic temperate regions and evidence suggests that M. edulis may have survived north of the ice margin (Maggs et al., 2008; Riginos & Henzler, 2008). Second, there could be a contemporary spread of M. trossulus from the Pacific Ocean. Jones et al. (2003) found that waters around NW Greenland contained high levels of phosphate indicating Pacific water being transported into this area. Also, there are a few reports of live M. trossulus in Arctic Alaska and Canada (Feder et al., 2003), so the spread of planktonic larvae from the Canadian Arctic could be possible. A third scenario could be that M. trossulus spread to Arctic Greenland from the East coast of Canada. However, as the West Greenland Current moves along the coast from south to north, and Mytilus mussels are expected to disperse with rather than against currents, this scenario seems unlikely (McQuaid & Phillips, 2000). Finally, Mytilus spp. are known to disperse by human activities and can survive long distances and fluctuating temperatures (Lee & Chown, 2007). Qaanaaq is situated less than 150 km from the US Thule Air Base, which receives supplies by US ships; this is providing an alternative dispersal route of M. trossulus from the north Pacific.

The invasive M. galloprovincialis appeared widespread from Greenland to the Pechora Sea. In Norway, M. galloprovincialis appears common along the coastline (Brooks & Farmen, 2013), and the discovery of M. galloprovincialis in Svalbard suggests colonization by ocean currents as hypothesized by Borge et al. (2005) or ship traffic from the Norwegian mainland (Ware et al., 2014).

4.2 Mytilus hybrid zones in the Arctic

Most sampling locations displayed varying degrees of hybridization and introgression between different Mytilus spp. and only four locations contained apparently pure populations (Fig. 1). Introgression can affect a population’s fitness and vulnerability to climate change. In the study region, hybrid zones were found in Norway, Svalbard, and Greenland, with the highest abundance of the invasive M. galloprovincialis found along the Norwegian coast, especially in Lofoten (68°N) further supporting the findings by Brooks and Farmen (2013) and Riginos and Henzler (2008). Additionally, a surprisingly high amount of M. galloprovincialis was found at Svalbard. We also found evidence of limited introgression of M. galloprovincialis in the Russian and Icelandic samples, and the ecological consequences of invasive mussels in these regions need to be studied further. In the White Sea, M. trossulus individuals were only recorded in one of two locations. This small-scale regional variation in species composition was also observed by Väänölä and Strelkov (2011), who also found M. trossulus and M. edulis/M. trossulus hybrids but to a much lesser extent than M. edulis. It is believed that the expansion of M. trossulus in the White Sea is most likely facilitated by ships (Väänölä & Strelkov, 2011). This explains the fact that populations dominated by M. trossulus are confined to sites with harbors and seaports, while M. edulis inhabit all the coastline of the White Sea, where the substrates are appropriate. In the present study, the sample WS1 that contained M. trossulus and their hybrids were collected directly in the area of the White Sea Biological Station Kartesh, which has a regular ship connection with Chupa, a small town in Kandalaksha Bay. Recently, M. trossulus was found in the Chupa harbor (Katolikova, Khaitov, Vänölä, Gantsevich, & Strelkov, 2016), where ship traffic from the Barents Sea has been relatively intensive. In contrast, the WS2 site with pure M. edulis in the sample is located on an uninhabited island Kondostrov in the Omegna Bay, which is far from the towns with intensive ship traffic.

4.3 Population structure of Mytilus edulis

The genetic structure of the M. edulis populations in this study revealed a significant split between M. edulis samples from each side of the Atlantic, with Icelandic M. edulis appearing as an admixture of the two gene pools. This divergence of W and E Atlantic populations is in line with the findings of Riginos and Henzler (2008) and Waltari and Hickerson (2013), who suggested that M. edulis survived in a W Atlantic glacial refugium. Furthermore, Riginos et al. (2004) found low gene flow across the Atlantic, providing an explanation for the continuing divergence of M. edulis populations from W and E Atlantic coasts. These studies primarily looked at mitochondrial DNA, but their results are strongly supported by the SNP analysis presented here. This, however, contrasts to the meta-population analysis of polychaete and echinoderm populations in the Arctic showing high gene flow between populations (Hardy et al., 2011). This difference in gene flow patterns between different species with long planktonic larval stage further highlights the necessity of understanding the population structure within species to best conserve biodiversity in the Arctic.

In general, FST values between samples from Norway, Svalbard, and Russia and the Icelandic sample are lower than between the Icelandic sample and Greenlandic samples. Śmietanka, Burzyński, Hummel, and Wenne (2014) suggested a single glacial Atlantic refugium founding European M. edulis. However, our studied sample from Iceland suggests the population to consist of individuals of mixed ancestry. Further analyses of their origin/history could be elucidated by conducting additional analysis of samples from both sides of the Atlantic. Considering that the major North Atlantic Current reaches Iceland from the east, it is perhaps more likely that Iceland would be
recruiting spat from East Atlantic populations. This is also inferred by Riginos and Henzler (2009), who found postcolonization gene flow from northern Europe to Iceland.

The outlier tests identified six loci as \( F_{ST} \) outliers at the 5% significance levels. All of these outliers are high \( F_{ST} \) outliers (Fig. 6) indicating diversifying selection (Beaumont & Nichols, 1996). However, a strong genetic cline as observed here is known to sometimes overestimate the number of loci under diversifying selection (Strand et al., 2012). Furthermore, introgression between Mytilus spp. have been found to cause high \( F_{ST} \) outliers (Gosset & Bierne, 2012), and this result should be interpreted with some caution. Still, we find that the pattern of population structure is the same for the “neutral” and the “outlier” data sets (Appendices 4 and 5), suggesting that patterns of neutral population structure is correlated with adaptive evolution in response to divergent local environmental conditions. Temperature influences the large-scale geographical distribution of species (Sunday, Bates, & Dulvy, 2011); however, on a local scale other factors including predation, the presence of sea ice, suitable habitats, water current, and salinity can influence the distribution of intertidal species (Høgslund, Sejr, Wiktor, Blicher, & Wegeberg, 2014; Kroeker et al., 2016; Paine, 1974), and these conditions are very different between W Greenland and the other sampling sites (Rayner et al., 2003). Still, the high divergence between samples from the Eastern Atlantic and Greenland cannot be explained alone by loci subject to selection. \( F_{ST} \) values for the “neutral” data set are still high (Table S4) suggesting a high degree of isolation between groups. This isolation in turn may have facilitated local adaptation at this rather large geographical scale. For more specific insights on the environmental factors responsible for local adaptation, the geographical scale, and its genomewide significance, a more elaborate sampling design is warranted including more regional samples and a higher degree of genomic coverage.

### 4.4 | Implications for conservation of marine species in the face of climate change

The effects of global warming increase the spread and associated threat of nonindigenous species across the globe (Gardner et al., 2016; Hellmann, Byers, Bierwagen, & Dukes, 2008; Saarman & Pogson, 2015). A study by Wisz et al. (2015) predicted that continued warming of the Arctic could open the Bering Strait and thus facilitate a Pacific–Arctic exchange of nonindigenous species, which could have adverse impact on Arctic biodiversity. Moreover, human activities are short-cutting natural dispersal barriers for nonindigenous species (Carlton & Geller, 1993), posing a global risk of spreading these to novel regions. In this regard, especially ship traffic facilitates dispersal (e.g. in ballast water and hull fouling; Chan, MacIsaac, & Bailey, 2015; Geller et al., 1994; Ware et al., 2014). Such intrusions of nonindigenous species into the Arctic have already occurred (e.g. Pacific king crabs *Paralithodes camtschaticus* and bluefin tuna *Thunnus thynnus*; CAFF 2013; MacKenzie, Payne, Boje, Hoyer, & Siegstad, 2014; Oug, Cochrane, Sundet, Norling, & Nilsson, 2011), and Saarman and Pogson (2015) found that the nonindigenous *M. galloprovincialis* pose an ecological threat to *M. trossulus* along the Californian coast as it had displaced and continues to displace the native *M. trossulus*. The surprisingly broad distribution of *M. galloprovincialis* in the Arctic therefore highlights the benefit of using genetic tools and stresses the need for developing measures to detect and identify nonindigenous species and pathways of introduction, to understand and reduce the threat of invasive species in the Arctic.

Prior to the current investigation, multiple studies have assumed *Mytilus* mussels in the Arctic to be exclusively *M. edulis* (Berge et al., 2005; Hansen, Hanken, Nielsen, Nielsen, & Thomsen, 2011; Jensen, 1905; Strand & Asmund, 2003). The identification of three *Mytilus* spp. across the Arctic has implications for ecological and ecotoxicological research in the region. *Mytilus* mussels are extensively used in biological monitoring programs (Wenne et al., 2016). However, interspecific differences in physiology and responses to environmental pollutants have been reported (Brooks, Farmen, Heier, Blanco-Rayon, & Izagirre, 2015; Fly & Hilbish, 2013), and thus, the lack of genetic knowledge could seriously affect the conclusions of ongoing biological monitoring. We therefore emphasize the importance of applying genetic tools to document species status, when conducting ecological, ecotoxicological, and physiological studies.

Moreover, assuming that the distribution and genetic connectivity between regions observed in this study is to be a first approximation representative for benthic invertebrates in general, several important observations were made related to quantifying changes in species distribution in a warmer Arctic. A number of congener species exists, which display different responses to changes in temperature. The genetic connectivity and inferred gene flow are closely linked to major ocean currents, which means that predicting range changes purely based on future climate predictions without considering dispersal potential or barriers can be misleading. In fact, changes in ocean currents and thereby in supply of potential colonizers may be a more important driver of change than warming per se. This has previously been demonstrated by the species changes observed during the large northward expansion of Atlantic water in the Barents Sea and along the W Greenland coast in the 1930s (Drinkwater, 2006). Genetically isolated areas like outflow shelves without upstream source populations (such as, NE Greenland) appear to be especially vulnerable to human vectors (such as shipping) as the absence of several species here likely reflects lacking postglacial invasion rather than adverse climatic conditions. Finally, NW Greenland *M. trossulus* populations with an affinity to the Pacific suggest that exchange of species from the Pacific across the Arctic and into the Atlantic is taking place. However, all of these factors should be further validated through urgently needed studies documenting current distribution and genetic composition of marine species in the Arctic.

### ACKNOWLEDGEMENTS

We wish to thank Dr. Risto Väinölä and Prof. Paulino Martínez Portela for providing mussel tissue for reference samples. Dorte Meldrup is acknowledged for laboratory support. Sampling in Nuuk was conducted as part of the Greenland Ecosystem Monitoring (GEM) program, and we also appreciate Snorri Gunnarson and Dr. Dorte
Krause-Jensen for collecting mussels. The project was funded by the 15th of June Foundation, the Program of Russian Academy of Sciences “Fundamental Research to the Development of Arctic”, the Framcentre flagship Fjord and Coast project “Life at the Edge”, and a Norwegian Research Council project (project nr 225044). This work is a contribution to the Arctic Science Partnership (ASP), asp-net.org.

DATA ARCHIVING STATEMENT


CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES


Toro, J. E., Thompson, R. J., & Innes, D. J. (2002). Reproductive isolation and reproductive output in two sympatric mussel species (Mytilus edulis, M. trossulus) and their hybrids from Newfoundland. Marine Biology, 141, 897–909.

Toro, J. E., Thompson, R. J., & Innes, D. J. (2006). Fertilization success and early survival in pure and hybrid larvae of Mytilus edulis (Linnaeus, 1758) and M. trossulus (Gould, 1850) from laboratory crosses. Aquaculture Research, 37, 1703–1708.


SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.
APPENDIX 1 $H_e$, $H_o$ and allelic richness for the "Mytilus edulis" data set.
Estimates of expected ($H_e$) and observed ($H_o$) heterozygosities and the allelic richness for the samples only including $M. edulis$.

<table>
<thead>
<tr>
<th>Code</th>
<th>$H_e$</th>
<th>$H_o$</th>
<th>Allelic richness</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPE</td>
<td>0.24</td>
<td>0.24</td>
<td>1.54</td>
</tr>
<tr>
<td>NUS</td>
<td>0.25</td>
<td>0.23</td>
<td>1.53</td>
</tr>
<tr>
<td>NUL</td>
<td>0.23</td>
<td>0.22</td>
<td>1.54</td>
</tr>
<tr>
<td>KOB</td>
<td>0.25</td>
<td>0.22</td>
<td>1.54</td>
</tr>
<tr>
<td>ICE</td>
<td>0.28</td>
<td>0.27</td>
<td>1.60</td>
</tr>
<tr>
<td>LOF</td>
<td>0.26</td>
<td>0.28</td>
<td>1.58</td>
</tr>
<tr>
<td>TRS</td>
<td>0.27</td>
<td>0.29</td>
<td>1.64</td>
</tr>
<tr>
<td>TRL</td>
<td>0.28</td>
<td>0.24</td>
<td>1.50</td>
</tr>
<tr>
<td>SV1</td>
<td>0.27</td>
<td>0.25</td>
<td>1.59</td>
</tr>
<tr>
<td>SV2</td>
<td>0.27</td>
<td>0.27</td>
<td>1.62</td>
</tr>
<tr>
<td>SV3</td>
<td>0.26</td>
<td>0.25</td>
<td>1.58</td>
</tr>
<tr>
<td>SV4</td>
<td>0.27</td>
<td>0.25</td>
<td>1.60</td>
</tr>
<tr>
<td>PSW</td>
<td>0.30</td>
<td>0.29</td>
<td>1.66</td>
</tr>
<tr>
<td>PSE</td>
<td>0.30</td>
<td>0.29</td>
<td>1.67</td>
</tr>
<tr>
<td>WS1</td>
<td>0.28</td>
<td>0.25</td>
<td>1.62</td>
</tr>
<tr>
<td>WS2</td>
<td>0.29</td>
<td>0.27</td>
<td>1.66</td>
</tr>
</tbody>
</table>

APPENDIX 2 Structure analysis for the full data set.
APPENDIX 3 Confidence intervals
95% confidence intervals for the inferred ancestry for simulated parentals and hybrids analyzed with Structure (K = 4). Ancestry estimates for the two inferred *Mytilus edulis* clusters (Greenland and Eastern Atlantic) are pooled. See text for explanation.

<table>
<thead>
<tr>
<th>Simulated individuals</th>
<th><em>Mytilus edulis</em></th>
<th><em>Mytilus trossulus</em></th>
<th><em>Mytilus galloprovincialis</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Greenland</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. edulis</td>
<td>0.89–0.99</td>
<td>0.00–0.02</td>
<td>0.01–0.10</td>
</tr>
<tr>
<td>M. trossulus</td>
<td>0.00–0.01</td>
<td>0.99–1.00</td>
<td>0.00–0.01</td>
</tr>
<tr>
<td>M. edulis/M. trossulus F1</td>
<td>0.34–0.54</td>
<td>0.44–0.56</td>
<td>0.01–0.10</td>
</tr>
<tr>
<td>M. edulis/M. trossulus F2</td>
<td>0.36–0.59</td>
<td>0.39–0.60</td>
<td>0.01–0.09</td>
</tr>
<tr>
<td>M. edulis/M. trossulus backcross M. edulis</td>
<td>0.60–0.82</td>
<td>0.15–0.34</td>
<td>0.01–0.10</td>
</tr>
<tr>
<td>M. edulis/M. trossulus backcross M. trossulus</td>
<td>0.14–0.31</td>
<td>0.67–0.82</td>
<td>0.01–0.07</td>
</tr>
<tr>
<td><strong>Eastern Atlantic</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. edulis</td>
<td>0.92–0.99</td>
<td>0.00–0.02</td>
<td>0.01–0.07</td>
</tr>
<tr>
<td>M. trossulus</td>
<td>0.00–0.01</td>
<td>0.99–1.00</td>
<td>0.00–0.01</td>
</tr>
<tr>
<td>M. galloprovincialis</td>
<td>0.01–0.02</td>
<td>0.00–0.01</td>
<td>0.97–0.99</td>
</tr>
<tr>
<td>M. edulis/M. trossulus F1</td>
<td>0.43–0.55</td>
<td>0.43–0.54</td>
<td>0.01–0.06</td>
</tr>
<tr>
<td>M. edulis/M. trossulus F2</td>
<td>0.38–0.60</td>
<td>0.37–0.59</td>
<td>0.01–0.09</td>
</tr>
<tr>
<td>M. edulis/M. trossulus backcross M. edulis</td>
<td>0.70–0.76</td>
<td>0.20–0.25</td>
<td>0.04–0.05</td>
</tr>
<tr>
<td>M. edulis/M. trossulus backcross M. trossulus</td>
<td>0.17–0.33</td>
<td>0.65–0.80</td>
<td>0.01–0.08</td>
</tr>
<tr>
<td>M. edulis/M. galloprovincialis F1</td>
<td>0.37–0.68</td>
<td>0.00–0.02</td>
<td>0.31–0.62</td>
</tr>
<tr>
<td>M. edulis/M. galloprovincialis F2</td>
<td>0.32–0.73</td>
<td>0.00–0.03</td>
<td>0.26–0.67</td>
</tr>
<tr>
<td>M. edulis/M. galloprovincialis backcross M. edulis</td>
<td>0.61–0.96</td>
<td>0.00–0.02</td>
<td>0.03–0.39</td>
</tr>
<tr>
<td>M. edulis/M. galloprovincialis backcross M. galloprovincialis</td>
<td>0.14–0.39</td>
<td>0.00–0.02</td>
<td>0.60–0.86</td>
</tr>
</tbody>
</table>

APPENDIX 4 Structure analysis for the “Mytilus edulis” “neutral” data set.
Results from clustering analysis (K = 2) for the “Mytilus edulis” data set only including SNPs not documented as outliers. Samples are 1: UPE, 2: NUU comprising of NUS and NUL, 3: KOB, 4: ICE, 5: LOF, 6: TRO comprising of TRS and TRL, 7: SVA comprising of SV1, SV2 and SV3, 8: SV4, 9: PSW, 10: PSE, 11: WS1, and 12: WS2. For explanation of sample identification codes, see Table 1.
APPENDIX 5 Structure analysis for the "Mytilus edulis" "outlier" data set.
Results from clustering analysis (K = 2) for the "Mytilus edulis" data set only including the six outlier SNPs. Samples are 1: UPE, 2: NUU comprising of NUS and NUL, 3: KOB, 4: ICE, 5: LOF, 6: TRO comprising of TRS and TRL, 7: SVA comprising of SV1, SV2, and SV3, 8: SV4, 9: PSW, 10: PSE, 11: WS1, and 12: WS2. For explanation of sample identification codes, see Table 1.