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

Sofie Smedegaard Mathiesen¹,² (SSM), Jakob Thyrring¹ (JT), Jakob Hemmer-Hansen² (JHH), Jørgen Berge³,⁴ (JB), Alexey Sukhotin⁵,⁶ (AS), Peter Leopold³ (PL), Michaël Bekaert⁷ (MB), Mikael Kristian Sejr¹‡ (MKS) and Einar Eg Nielsen²‡ (EEN)

¹Arctic Research Centre, Department of Bioscience, Aarhus University, Ny Munkegade 114, building 1540, DK-8000 Aarhus C, Denmark

²Section for Marine Living Resources, National Institute of Aquatic Resources, Technical University of Denmark, Vejlsovej 39, 8600 Silkeborg, Denmark

³Faculty of Biosciences, Fisheries and Economics, UiT The Arctic University of Norway, P.O. Box 6050 Langnes, 9037 Tromsø, Norway

⁴The University Centre in Svalbard, P.O. Box 156, N-9171 Longyearbyen, Norway

⁵White Sea Biological Station, Zoological Institute of Russian Academy of Sciences, Universitetskaya nab. 1, 199034 St. Petersburg, Russia

⁶Invertebrate Zoology Department, St. Petersburg State University, Universitetskaya nab. 7/9, 199034 St. Petersburg, Russia

⁷Institute of Aquaculture, University of Stirling FK9 4LA, Stirling, Scotland, UK

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/eva.12415

This article is protected by copyright. All rights reserved.
Running title: Genetic diversity of Mytilus in the Arctic

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. In order 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 M. trossulus in high Arctic NW Greenland (77°N) and M. 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 are 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
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°C to 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 (Poloczanska et al. 2013; IPCC 2014), 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 and 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 can exhibit different adaptations and tolerance limits to specific environments (Nielsen et al. 2009; Limborg et al. 2012; Thyrring et al. 2015b), which makes 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 (Laikre et al. 2010; Hansen et al. 2012; Brodersen and Seehausen 2014). 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 and southern hemispheres (Goldberg 1986; Rainbow 1995; Gosling 2003; Thyrring et al. 2015a). *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, etc.) have been performed (Gosling 2003; Mubiana et al. 2005; Wanamaker et al. 2007; Jones et al. 2010; Søndergaard et al. 2011). Furthermore, *Mytilus* spp. have already demonstrated adaptations to different environments (Blicher et al. 2013; Thyrring et al. 2015b) and a shift in their southern geographical range caused by increasing temperatures (Jones et al. 2015b).
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, *M. trossulus* Gould 1850 and *M. galloprovincialis* Lmk. 1819, all belong to the *Mytilus 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 (Koehn 1991; Riginos and Cunningham 2005; Toro et al. 2006; Doherty et al. 2009), while others (Gardner and Thompson 2001; Toro et al. 2002; Toro et al. 2004; Tremblay and Landry 2016) found a difference in fitness between parental types and hybrids and backcrosses. These findings and numerous studies on introgression between them (Roux et al. 2014; Fraïsse et al. 2016) has challenged the isolation species concept (White 1978), however they are generally considered to be different species, as they remain ecological distinct despite semi-permeable barriers for gene flow and introgression (Bierne et al. 2003a; Fraïsse et al. 2014; Saarman and 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 (Vermeij 1991; Rawson and Hilbish 1995, 1998). 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); *Mytilus edulis* populations on each side of the Atlantic are genetically distinct (Riginos et al. 2004; Riginos and Henzler 2008). Speciation between *M. edulis* and *M. galloprovincialis* most likely occurred through allopatric isolation approximately 2.5 mya (Rawson and Hilbish 1995; Quesada et al. 1998; Rawson and Hilbish 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 and 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 and Cunningham 2005).
The geographical distribution and genetic population structure of *Mytilus* spp. have been intensively studied in boreal and temperate regions (Sarver and Foltz 1993; Hilbish et al. 2002; Westerbom et al. 2002; Bierne et al. 2003a; Väinölä and Strelkov 2011); 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 et al. 2001; Sukhotin et al. 2007; Riginos and Henzler 2008; Väinölä and 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 and 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 et al. 2015b). Moreover, in 2004, subtidal *M. edulis* were discovered at the mouth of Isfjorden in Svalbard after 1,000 years of absence (Berge et al. 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 et al. (2003) found live *M. trossulus* in Arctic Alaska in the 1990s. Furthermore, Wenne et al. (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 to 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 (Riginos and
Henzler 2008; Brooks and Farmen 2013). This could be related to human activities like ship traffic in
rural areas enabling faster invasion of waters otherwise not directly accessible to them (Geller et al.
1994; Anderson et al. 2002). Furthermore, it has been demonstrated that *M. galloprovincialis* is
capable of tolerating low 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
(McDonald et al. 1991; Presa et al. 2002; Bierne et al. 2003b; Feder et al. 2003; Ouagajjou et al. 2011;
Brooks and Farmen 2013). 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; Zbawicka et al. 2012; Zbawicka et
al. 2014; Saarman and Pogson 2015).

Utilizing 81 nuclear SNPs, we examined the distribution of *Mytilus* spp. in subarctic and Arctic
regions ranging from the eastern Baffin Bay to the Pechora Sea with a special emphasis on the
spatiotemporal population structure of *M. edulis*. We further aimed at identifying the source
population or populations for the newly discovered *M. edulis* population in Svalbard and whether the
observed differences in temperature response found in W Greenland *Mytilus* populations (Thyrring et
al. 2015b) could be caused by genetically based local adaptation.
Materials and methods

Study sites and sampling

Nineteen *Mytilus* spp. samples, consisting of 509 individuals in total, were collected from thirteen subarctic and Arctic sites (Table 1 and Fig. 1). Our primary focus was to assure broad geographical coverage and to sample regions where specific hypothesis regarding origin had been generated. The aim was to collect between 30 and 50 individuals from each site. However, as sampling Arctic regions is associated with high logistical costs, we had to rely on already available samples and numbers in some instances. Three samples were collected along the Norwegian NW coast at Tromsø (TRS and TRL) and Lofoten (LOF) and four samples were collected from the Svalbard archipelago (SV1, SV2, SV3 and SV4). Four samples were obtained from the Russian Arctic: two from the White Sea (WS1, WS2) and two from the southeast part of the Barents Sea: Pechora Sea to the east (PSE) and to the west (PSW) of Dolgiy Island. Further, one sample was collected in Iceland: south of Reykjavik (SRI), and six samples were collected in western Greenland: Nuuk (NUS and NUL), Kobbefjord (KOB), Upernavik (UPE) and Qaanaaq (QAS and QAL). From Tromsø, Nuuk and Qaanaaq, mussels of different size classes were collected as size can be used as a proxy for age class and hence indicate possible short-term genetic change over time; TRS, NUS and QAS were smaller mussels in the size range 15–30 mm in length, while TRL, NUL and QAL being mussels larger than 50 mm.

Samples were stored at -19°C (TRS, TRL, SRI, KOB, UPE, QAS and QAL) or in 96% ethanol at 4°C (LOF, SV1, SV2, SV3, SV4, WS1, WS2, PSW and PSE). Measurements of shell length, width and height of frozen mussels were conducted at the laboratory facilities at DTU Aqua in Silkeborg, Denmark, whereas the measurements of ethanol preserved specimens were carried out at the sampling locations.

Reference samples of *M. trossulus* and *M. galloprovincialis* were provided from the Sea of Okhotsk, Russia, and from around Galician Rías in NW Spain, respectively, in order to evaluate the species status of the sampled mussels and to identify potential hybrids.
**DNA extraction**

A minimum of 30 mg (wet weight) of mantle tissue was dissected from each mussel, and DNA was extracted using the Omega EZNA Tissue DNA kit (Omega Bio-Tek, Norcross, GA, USA) according to the manufacturer’s instructions for tissue. DNA content in the extracts were verified on a NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA, USA).

**SNP genotyping**

SNP genotyping was conducted using the Fluidigm Biomark™ HD System. 96.96 Dynamic Array IFCs were read on a Real-Time PCR system after amplification and scored using Fluidigm SNP Genotyping Analysis software. The samples were genotyped for a panel of 96 SNPs; 19 from previous publications on *Mytilus* spp. genetic structure (Zbawicka et al. 2012; Zbawicka et al. 2014) and 77 new SNPs originated from RAD sequencing of genomic *M. edulis* DNA (EBI Sequence Read Archive (SRA) study ERP006912) at the University of Stirling (Supplementary Table S1).

**Summary statistics**

Loci with more than 25% missing data across all samples were discarded. Genepop 4.2 (Raymond and Rousset 1995; Rousset 2008) 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 et al. 2013) was used to calculate allelic richness and estimate expected ($H_e$) and observed ($H_o$) heterozygosities. This was done for both the full data set and for a reduced data set consisting exclusively of inferred *M. edulis* individuals (see explanation in section on *M. edulis* population structure below). Overall and pairwise $F_{ST}$ values for all samples were estimated in Genepop 4.2. This initial sorting and discarding of SNP loci resulted in 81 loci being retained for further analyses. As most SNP loci were developed from *M. edulis*, reliable scoring of *M. trossulus*...
Identification of hybrids

Based on the generated pairwise $F_{ST}$ estimates, the grouping of samples was visualized in a multi-dimensional 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 and Ahmed 2011) to illustrate the genetic relationships among individuals across all samples. Structure v2.3.4, utilizing the Bayesian MCMC clustering approach, (Pritchard et al. 2000) was used to visualize species integrity and identify possible hybridization among *Mytilus* spp. using a variable number of pre-defined clusters (K) for grouping individuals. This was also done in order 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 and Henzler 2008), simulations were run for a number of pre-defined 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 in order 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. In order to evaluate the power of designating individuals as pure or hybrids, we followed the procedure described in Nielsen et al. (2003) using the program Hybridlab (Nielsen et al. 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* individuals (NUS, KOB and WS2). Separate simulations were conducted for *M. edulis* samples from Greenland (NUS, KOB) and the Eastern Atlantic (WS2). The simulated and
real individuals were included in a common Structure run (K = 4) and 95% confidence intervals for the inferred ancestry of the simulated individuals were recorded and compared to the real individuals.

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 with admixture proportions below 0.2, as estimated by Structure. This was done in order 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<sub>ST</sub> 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 and 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 multi-dimensional 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.
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 and Lischer 2010) based on the method in Excoffier et al. (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: (1) Greenlandic samples, (2) Samples from Norway, the Svalbard archipelago and Russian waters, and (3) the Icelandic sample (see the section under Results sub-section Population structure of Mytilus 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 ‘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*. I.e. 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 if they displayed contrasting patterns.

Results

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 6 samples); 100078_A x 40154_A (significant in 8 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 $H_o$ was close to $H_e$ except for the Upernavik sample (UPE), where $H_e$ and $H_o$ 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 provided 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 (Supplementary 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).

**Identification of hybrids**

The multi-dimensional 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 multi-dimensional 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 A2) showed that the clusters make biological 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 A3). 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 exclusive ‘*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 A2.b).
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.

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 (Supplementary 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 vs. 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 multi-dimensional 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: (1) Greenlandic samples, (2) Norwegian, Svalbard and Russian samples, and (3) 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.
**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 and 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 et al. 2012). Furthermore, introgression between *Mytilus* spp. have 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 (Supplementary Table 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 A4 and A5).

**Discussion**

*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 et al. 2016). Molecular genetic knowledge is a key measure to identify the distribution of invasive congener species (Geller et al. 2010), which may cause cascading ecosystem effects. Despite congener species appearing morphologically similar, interspecific variation in ecology and physiology may impact population fitness (Somero 2005; Fly and Hilbish 2013; Fraïsse et al. 2016). In the Arctic, baselines studies on genetic variation and species abundance are largely absent but urgently needed (Bluhm et al. 2011; Wassmann et al. 2011). Pioneer work should therefore focus on keystone model species (such as *Mytilus*), because of their disproportionally large effect on their environment.

This article is protected by copyright. All rights reserved.
Mytilus spp. were found pan-Arctic (though 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 larvae dispersal from southern populations (Berge et al. 2005; Renaud et al. 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 (Sirenko and Gagaev 2007; Fetzer and Arntz 2008; Bluhm et al. 2011). Ocean currents also explain why Mytilus spp. remain absent in NE Greenland despite the environmental resemblance of NW Greenland with regards to temperatures and ice conditions (Sejr et al. 2009). In general, the NE Greenland shelf 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 a 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 Tasiilaq (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 suggest that M. edulis may have survived north of the ice margin (Maggs et al. 2008; Riginos and 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 and Phillips 2000). Finally, *Mytilus* spp. are known to disperse by human activities and can survive long distances and fluctuating temperatures (Lee and 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 and Farmen 2013), and the discovery of *M. galloprovincialis* in Svalbard suggests colonization by ocean currents as hypothesized by Berge et al. (2005) or ship traffic from the Norwegian mainland (Ware et al. 2014).

*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 needs 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 Väinölä 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 (Väinölä and 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 et al. 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 Onega Bay, which is far from the towns with intensive ship traffic.

*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 is, however, contrasting 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 highlight the necessity of understanding the population structure within species to best conserve biodiversity in the Arctic.
In general, $F_{ST}$ values between samples from Norway, Svalbard and Russia and the Icelandic sample are lower than between the Icelandic sample and Greenlandic samples. Śmietanka et al. (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 post-colonization 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 and 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 and Bierne 2013), 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 A4 and A5), 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 et al. 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 (Paine 1974; Høgslund et al. 2014; Kroeker et al. 2016), 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 (Supplementary 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 genome wide significance, a more
elaborate sampling design is warranted including more regional samples and a higher degree of genomic coverage.

**Implications for conservation of marine species in the face of climate change**

The effects of global warming increase the spread and associated threat of non-indigenous species across the globe (Hellmann et al. 2008; Saarman and Pogson 2015; Gardner et al. 2016). 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 non-indigenous species, which could have adverse impact on Arctic biodiversity. Moreover, human activities are short-cutting natural dispersal barriers for non-indigenous species (Carlton and 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] (Geller et al. 1994; Ware et al. 2014; Chan et al. 2015). Such intrusions of non-indigenous species into the Arctic have already occurred [e.g. Pacific king crabs *Paralithodes camtschaticus* and bluefin tuna *Thunnus thynnus*] (Oug et al. 2011; CAFF 2013; MacKenzie et al. 2014), and Saarman and Pogson (2015) found that the non-indigenous *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 highlight the benefit of using genetic tools and stresses the need for developing measures to detect and identify non-indigenous 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* (Jensen 1905; Strand and Asmund 2003; Berge et al. 2005; Hansen et al. 2011). The identification of three *Mytilus* spp. across the Arctic have 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 (Fly and Hilbish 2013; Brooks et al. 2015), 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 is 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) since the absence of several species here likely reflects lacking post-glacial 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 Martinez 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

This article is protected by copyright. All rights reserved.
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**

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.438h3

**Conflict of Interest**

The authors declare no conflict of interest.

**Literature cited**


Chan, F. T., H. J. MacIsaac, and S. A. Bailey. 2015. Relative importance of vessel hull fouling and ballast water as transport vectors of nonindigenous species to the Canadian Arctic. Canadian Journal of Fisheries and Aquatic Sciences **72**:1230-1242.


This article is protected by copyright. All rights reserved.


Rousset, F. 2008. GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Molecular Ecology Resources 8:103-106.


This article is protected by copyright. All rights reserved.


Table 1 Summary of sample information including sampling location and year, sampling code indicating location and possibly size, estimates of expected (H<sub>e</sub>) and observed (H<sub>o</sub>) 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&lt;sub&gt;e&lt;/sub&gt;</th>
<th>H&lt;sub&gt;o&lt;/sub&gt;</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></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></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></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></td>
<td>Nuuk, &gt;50 mm</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>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kobbeffjord</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>Norway</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></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></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>Svalbard</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></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></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></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 of Dolgiy Island</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></td>
<td>Pechora Sea, east of Dolgiy Island</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>Location</td>
<td>Latitude</td>
<td>Longitude</td>
<td>Year</td>
<td>Code</td>
<td>Zone</td>
<td>Subspecific 1</td>
<td>Subspecific 2</td>
<td>Subspecific 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>------</td>
<td>------</td>
<td>-----------</td>
<td>---------------</td>
<td>---------------</td>
<td>---------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White Sea, Kandalaksha Bay</td>
<td>66.3372</td>
<td>33.6494</td>
<td>2014</td>
<td>WS1</td>
<td>Subtidal zone</td>
<td>0.32</td>
<td>0.24</td>
<td>1.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White Sea, Onega Bay</td>
<td>64.2079</td>
<td>36.6187</td>
<td>2014</td>
<td>WS2</td>
<td>Intertidal zone</td>
<td>0.29</td>
<td>0.27</td>
<td>1.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Magadan, Okhotsk Sea*</td>
<td></td>
<td></td>
<td>2007</td>
<td>MTR</td>
<td></td>
<td>0.11</td>
<td>0.01</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spain Galician Rias*</td>
<td></td>
<td></td>
<td>2009</td>
<td>MGA</td>
<td></td>
<td>0.15</td>
<td>0.11</td>
<td>1.27</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* These samples were only used for reference i.e. a representative of *M. trossulus* and *M. galloprovincialis*. 
**Figure legends**

**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.

**Figure 2** Multi-dimensional scaling plot of a) all samples and b) designated *M. edulis* samples based on pairwise genetic distances among samples. For explanation of sample identification codes see Table 1. Further codes: NUU comprise of NUS and NUL, TRO comprise of TRS and TRL, and SVA comprise of SV1, SV2 and SV3.

**Figure 3** Principal component scatter plot of individual genotypes for a) all samples b) *M. edulis* samples and c) Norwegian, Svalbard and Russian samples. For explanation of sample identification codes see Table 1. Further codes: NUU comprise of NUS and NUL, TRO comprise of TRS and TRL, and SVA comprise of SV1, SV2 and SV3.

**Figure 4** Results of Structure clustering analyses for the full data set for K = 4. Samples are 1: QAS, 2: QAL, 3: UPE, 4: NUS, 5: NUL, 6: KOB, 7: ICE, 8: LOF, 9: TRS, 10: TRL, 11: SV1, 12: SV2, 13: SV3, 14: SV4, 15: PSW, 16: PSE, 17: WS1, 18: WS2, 19: MTR, and 20: MGA. For explanation of sample identification codes see Table 1.

**Figure 5** Results from clustering analyses of the *Mytilus edulis* data set with: a) K = 2, b) K = 3 and c) K = 4. 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.

**Figure 6** F^*_ST_* outlier analyses in Arlequin v3.5.1.3 utilizing the hierarchical island model. Black solid dots denotes loci, and grey 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.
Appendices

**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 analysed with Structure (K = 4). Ancestry estimates for the two inferred *M. edulis* clusters (Greenland and Eastern Atlantic) are pooled. See text for explanation.

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