UiT The Arctic University of Norway

Faculty of Biosciences, Fisheries and Economics

The Norwegian College of Fishery Science

Unravelling population genetic structure of *Arctogadus glacialis* in Northeast Greenland with whole genome sequencing

Rasmus Kjær Koch

FSK-3860 Master's thesis in Fisheries and Aquaculture, May 2024

Abstract

In the light of the current climate change with rapidly increasing temperatures, reducing ice cover and lower albedo, the physico-chemical conditions in the Arctic fluctuate more than any other place on the planet. This has a severe effect on Arctic marine ecosystems and the species they are composed by. To what extent remains unknown for several species such as the high-Arctic cryopelagic gadid *Arctogadus glacialis* (Peters, 1874), also known as ice cod. Due to its elusiveness, little is known about its biology, which is important knowledge for understanding the implications of climate change on the species. This project used a unique collection of *A. glacialis* from Northeast Greenland to investigate population genetic structure between and within fjord and shelf areas. Whole genome sequencing and population structure analyses of 110 individuals revealed overall and significant population genetic structure between the two types of habitats. Furthermore, the results revealed genetic differentiation within fjord and shelf areas, respectively, but the structuring was inconsistent throughout the analyses. Based on the literature, this was the first study to investigate and detect population genetic structure in *A. glacialis* using whole genome sequencing. The population genetic structure between fjord and shelf areas may be explained by the post-glacial colonization of the fjords, where the individuals were later physically constrained as sill formation and sea level began to increase. Moreover, local adaptations to their respective environments, whether it was fjord or shelf, may also be associated to the observed population genetic structure. Future work is needed on the ancestral origin of *A. glacialis* and adaptive divergence, which would give implications of climate change's effect on the species.

Table of Contents

Preface **1**

Introduction **2**

Materials and methods **10**

DNA extraction and quality assessment 11

Sample preparation for whole genome sequencing 14

Bioinformatic analysis 14

Population genetic structure analysis 16

Results **18**

Post-filtering- and sequencing results 18

Population genetic structure analysis 19

Discussion **26**

Population genetic structure between shelf and fjord areas 26

Population genetic structure within shelf and fjord areas, respectively 28

Methodical discussion 30

Concluding remarks and future perspectives 32

Acknowledgements **33**

References **34**

Appendices **42**

Appendix 1. Purification of Total DNA from Animal Tissues (Spin-Column Protocol) 42

Appendix 2. Purification of Total DNA from Animal Tissues (DNeasy 96 Protocol) 45

Appendix 3. Manual for Qubit 1X ds DNA BR Assay with a Qubit 4 Fluorometer 50

Appendix 4. Sequencing requirements 50

Appendix 5. Additional PCA plot 1 51

Appendix 6. Additional PCA plot 2 52

Appendix 7. DAPC scatter plot 53

Appendix 8. Admixture of all K values 54

Web links for workflow **56**

Preface

This project was based on collections and sampling from the TUNU Programme (UiT). One and a half year ago, when I first got in contact with my main supervisor Kim Præbel, I did not know what my thesis subject should be. Suddenly, Kim introduced me to a project about genetics in a high-Arctic cod, which sparked my interest immediately. Coming from the same surroundings every day for four years at the University of Copenhagen, it was also a unique opportunity to come to northern Norway and experience a new environment. Both in terms of the nature and a new network. During my education in Copenhagen, Arctic biology has been one of my favorite subjects to learn about, especially due to the impacts of climate change. However, genetics never interested me to the same extent, and it was therefore a gamble to write a whole thesis about that. During this project, I have gotten a different view on how to use genetics to solve biological issues, which I can take with me later in my career and for that, I am thankful. I will also express my thanks to my main supervisor Kim Præbel, who has been a great and an engaging mentor throughout this project. Also, a big thanks to Shripathi Bhat, whom I have spent much time with during my data analysis. When things did not go to plan or if I had any questions, he always took his time to help and guide me in the right direction. I will also express my thanks to Kristel Berg and Andrea Iselin Elvheim for their assistance in the lab and my co-supervisor Arve Lynghammar for helping me in the final stage of the thesis.

Introduction

The temperature in the Arctic is increasing two to three times faster than the global average (Wassmann et al., 2011). This process is accelerating as thawing of ice and snow leads to a lower albedo and therefore a higher absorption of sunlight by the oceans, which results in rapidly increasing temperatures. The increasing temperatures and precipitation also lead to increased freshwater runoff into the Arctic Ocean, which enhances stratification, lowers salinity and increases the input of nutrients (Wassmann et al., 2011). These events have a severe effect on the biodiversity and species richness in Arctic marine ecosystems (Hollowed et al., 2013) but to what extend still remains unknown despite increased interest and scientific effort the recent years. Marine species are sentinels when studying the relationship between the environment and the genome (Nielsen et al., 2009), as they experience fluctuations in e.g. temperature, oxygen, salinity and pollution. Because the environmental conditions in the Arctic fluctuate more than any other place on the planet (Pettitt-Wade et al., 2021), they are even better model organisms in that context. Arctic species that are ice-associated at some point during their life cycle such as the ice cod *Arctogadus glacialis* (Peters, 1874) have been suggested to be highly influenced by environmental changes due to physico-chemical gradients such as reducing ice cover and salinity changes (Hollowed et al., 2013).

Distribution and general biology of A. glacialis

Arctogadus glacialis has a circumpolar distribution and is mainly found in shelf areas in coastal waters and fjords of Northeast (NE) Greenland, but also appear in the Central Arctic Ocean, East Siberian Sea and northern Canada (Karamushko et al., 2022; Mecklenburg et al., 2018; Nielsen & Jensen, 1967). The distribution in the water column also varies, where it previously has been caught in shallow waters on the East Siberian Sea shelf from ice stations and ships on the outer shelf to the slope of the Arctic Ocean. It has also been caught by trawling in the Beaufort Seas (Frost & Lowry, 1983). However, *A. glacialis* is most common in deep waters over the outer continental shelf and slope at depths ranging from 5–930 m (Jordan et al., 2003). In the European Arctic, it is most abundant at 300–400 m (Aschan et al., 2009), however, Aschan et al. (2009) also concluded that the shelf areas were sampled more than deep water localities exceeding 700 m depth, which according to the authors, may have impacted

the results. *Arctogadus glacialis* has also been observed in offshore localities under drift sea ice and in river mouths (Mecklenburg et al., 2018). Moreover, it is most common near the sea bed (Aschan et al., 2009) but if the depth exceeds approximately 800 m, the ice cod is a bit further up in the water column (Mecklenburg et al., 2018).

Arctogadus glacialis is characterized as a cryopelagic species as it is associated with sea ice in some parts of its life history (Møller et al., 2002) but it is not known which role sea ice has on its biology. The ice cod is usually found in cold waters ranging from -0.6–1.5°C (Ghigliotti et al., 2020) but has also been caught in water temperatures at 2.5°C and 3.4°C (Aschan et al., 2009), which indicates that *A. glacialis* is not constrained to Arctic waters.

The ice cod has been found in e.g. narwhal, bearded seal, harp seal, Greenland halibut and ringed seal stomachs (Pettitt-Wade et al., 2021). Because it is preyed upon by Arctic marine mammals and fishes and that it is coupled with lower trophic levels through their diet, *A. glacialis* serve as a fundamental link (together with the phylogenetically close counterpart, polar cod *Boregadus saida*) between benthic/pelagic food webs and higher/lower trophic levels in the Arctic (Pettitt-Wade et al., 2021).

Spawning

The reproductive biology and spawning of the ice cod have not been thoroughly described/understood and the available literature is contradicting. Firstly, it has been stated that the ice cod spawns during winter under the ice, but which particular months are not specified further (Süfke et al., 1998). In that study, 280 individuals of *A. glacialis* were sampled with Agassiz trawl and bottom trawl at eight different stations in the NE Water Polynya outside NE Greenland in July-August 1990. Besides examining diet, body size and sex ratio, gonadosomatic index (GSI) was calculated, which can be used as a measurement for maturity of ovaries (Süfke et al., 1998). Based on low GSI measurements in female specimens collected in August, the authors concluded that *A. glacialis* spawns in winter, whereas other authors have suggested that spawning takes place during summer. This was based on trawling of fry in the East Siberian Sea during October and findings of females with mature ovaries (Aschan et al., 2009; Jordan et al., 2001), which suggests that *A. glacialis* has different spawning time in different areas. The ice cod's distribution through different life stages has not only been linked to the season but have also been associated with the depth. It has been suggested by Jordan et al.

(2003) and Aschan et al. (2009) that 1) *A. glacialis* spawns inshore in shallow waters, 2) juvenile development takes place offshore in deeper waters (due to reports of medium sized specimens at offshore localities), 3) that adults are present in the entire water column. Süfke et al. (1998) also suggested that small specimens in the NE Water Polynya outside NE Greenland with a standard length below 10 cm were most abundant at the shallowest station, whereas specimens over 20 cm dominated at the deep stations (Süfke et al., 1998). However, the median standard length was not significant with increasing depth (Süfke et al., 1998).

The diet of A. glacialis

Although knowledge about the general biology of *A. glacialis* is limited, their diet has been investigated by a number of authors. Firstly, the ice cod has been shown to feed on small crustaceans such as calanoid copepods, amphipods, mysids, ostracods and chaetognaths (Süfke et al., 1998). In the previously mentioned study from NE Water Polynya, copepods and amphipods were found in the stomachs of *A. glacialis* at all of the eight stations but copepods were most abundant overall. Christiansen et al. (2012) have further suggested that the ice cod's diet depends on life stage and depth, as they found that the main food source in shallow waters below 250 m depth for small ice cods (below 12 cm in total body length) was copepods. However, in deeper waters where larger specimens dominated, amphipods and mysids were more abundant in their stomachs. Further south from the NE Water Polynya, the diet has been examined in two northeastern Greenland fjords, Tyrolerfjord and Dove Bugt (Christiansen et al., 2012) for both *A. glacialis* and *B. saida*. These two species coexist and will therefore compete for some of the same food resources. In the two fjords, crustaceans, annelids, teleosts and echinoderms were found in the stomachs of either *B. saida* or *A. glacialis*. Annelids were exclusively consumed by the ice cod in both fjords, and echinoderms were completely absent. Furthermore, the ice cod did not feed on teleosts in Dove Bugt. Although the authors also suggested intraspecific competition for food, benthic prey were exclusively consumed by the ice cod, whereas the polar cod consumed pelagic prey (Christiansen et al., 2012).

Furthermore, Walters (1961) attempted to map the horizontal and vertical movements that *A. glacialis* makes in order to catch prey. Based on surveys from ice drift stations from the Chukchi Rise, the author hypothesized that the ice cod moved across shallow waters in mid-

winter towards pelagic feeding grounds (Walters, 1961). However, the author was not able to completely verify the hypothesis. It has also been suggested that *A. glacialis*, together with *B. saida*, undergo vertical movements in the late larval and early juvenile stage during summer (Bouchard et al., 2016). More specifically, the two species may show a diel movement, where most larvae inhabit 0–5 m below the sea surface at night and 5–10 m below the sea surface in daytime. This diel movement has been hypothesized to limit the predation mortality due to the avoidance of visual predators in daytime (Bouchard et al., 2016).

Population genetics

Population genetic structure is defined as a systematic variation in allele frequencies through space and is influenced by genetic drift, gene flow, mutation and natural selection (Jones & Wang, 2012). Genetic drift is a mechanism where the genetic composition/allele frequency in a population changes randomly from one generation to another (Hedrick, 2000). An individual can then receive two copied genes from the same gene in a previous generation, which is referred to as autozygosity. On the other hand, if the genes are not identical, it is called allozygosity, which covers both homozygosity and heterozygosity. Moreover, genetic drift is restricted to small populations. Gene flow refers to alleles moving through space due to migration or movement of gametes from one population to another. This can be limited by physical barriers such as rivers or mountains, or biological factors as e. g. high predation rates or low dispersal capacity of a species (Jones & Wang, 2012).
Mutation refers to changes in the DNA sequence and introduces new alleles to a population and lastly, natural selection causes changes in allele frequencies due to adaptation. In marine systems, gene flow via gametes and migration over large distances are some of the key contributors to low genetic divergence (Pálsson et al., 2009).

A cornerstone in population genetics is the Hardy-Weinberg (HW) principle, which predicts how gene frequencies is inherited from one generation to another based on different assumptions (Alghamdi & Padmanabhan, 2014; Hedrick, 2000). A population will be in HW equilibrium given a number of assumptions such as; 1) random mating, 2) no natural selection, 3) no genetic drift (therefore, a large population size is necessary), 4) no gene flow or migration, 5) no mutation and 6) the locus must be autosomal (Kliman, 2016). If one or more of these

assumptions are not met, the given population is not in HW equilibrium. The formula is as follows:

$$p^2 + 2pq + q^2 = 1$$

Where p is the genotype frequency of allele A, and q is the genotype frequency of allele a (Kliman, 2016). p^2 and q^2 therefore represent dominant and recessive homozygous genotype frequencies, and $2pq$ represent heterozygous genotype frequencies. If a population is not in HW equilibrium and therefore do not follow the assumptions mentioned above, the equation can be used for detecting which evolutionary force is involved. Another central component in population genetics is linkage equilibrium (LD), which is modulated by recombination rates and is affected by e.g. cross-over events during meiosis. LD means "nonrandom association of alleles at two or more loci" (Slatkin, 2008) and is also affected by natural selection, genetic drift and mutation.

Such principles from population genetics are the foundation for population genomics, which explains genetic variation and molecular evolution in a species. Population genomics is defined as discrimination between locus-specific (selection) and genome-wide effects (genetic drift and mutation) (Nielsen et al., 2009). Furthermore, it is used to analyze genomic variation in and between populations. In population genetic studies regarding pelagic marine fishes, it is common that the genetic diversity even at large geographical distances is little or non-existent, which could be due to a lack of physical barriers compared to terrestrial environments. This may result in high dispersal and gene flow. Together with higher dispersal rates when spawning and increased migration capabilities, it can be challenging to detect population genetic structure in marine fishes (Pálsson et al., 2009).

First era in population genomics

Since population genomics was applied for the first time in 1966 (Lewontin & Hubby, 1966), genetic methods have constantly evolved and still is, creating different genetic eras.
The first era was based on allozyme electrophoresis, which was initiated with a description of

genetic variation in *Drosophila* and humans (Harris, 1966; Lewontin & Hubby, 1966). Allozyme electrophoresis is a similar method to gel electrophoresis of DNA, but instead represent separation of proteins by molecular weight (Casillas & Barbadilla, 2017). Allozymes have different migratory patterns due to differences in molecular sizes and electrical charges of the proteins. Therefore, it is exclusively the protein-coding sections of the DNA that can be detected. However, not all of these changes can be detected in an electrophoretic gel (Berta et al., 2015), as the changes in amino acids have to affect the movement of a protein in a gel. Moreover, it can generally be difficult to interpret the bands after their migration if the bands are weak. Also, even if some alleles show the same migration pattern and have the same electric charge, it is not certain they are alike. Convergent evolution where two non-related species evolve the same traits but have different genotypes (Kocher & Stepien, 1997) have been suggested to be an explanation to that. In an electrophoretic gel, allozymes is detected by looking for staining of the gel after the protein migration. The allozymes and the different mobility patterns can be identified by migratory differences of the bands for homozygotes and multiple bands for heterozygotes (Berta et al., 2015).

Later, allozyme electrophoresis was replaced by the nucleotide sequence era, which was initiated by applying restriction enzymes to make restriction mapping, which is a method that breaks an unknown DNA segment and afterwards separates the DNA fragments by using gel electrophoresis (Casillas & Barbadilla, 2017; Saraswathy & Ramalingam, 2011). The restriction mapping and nucleotide site diversity made it possible to detect genetic diversity in DNA sequences. The allozymes were also replaced by e.g. microsatellites, amplified fragment length polymorphisms (AFLPs), single nucleotide polymorphisms (SNPs) and mtDNA (Nielsen et al., 2009).

Current era

The current era in population genomics is dominated by next generation sequencing (NGS), which makes it possible to sequence whole genomes and analyze millions-billions of short sequence reads instead of only analyzing specific parts of the DNA as seen in the past (Casillas & Barbadilla, 2017). One of those methods is Pool-seq, which is a cost-effective method that pools individuals e.g. from the same population instead of sequencing and library prepping every individual (Casillas & Barbadilla, 2017; Rellstab et al., 2015). Pool-seq is especially precise for revealing differences in allele frequencies in large population samples, even

though not all chromosomes are analyzed. Those that are analyzed, are often not sequenced more than one time, which brings down the cost.

In a review by Schlötterer et al. (2014), Pool-seq and whole genome sequencing (WGS) of individuals were compared in terms of the accuracy of allele frequency estimates, which was based on different factors. The model in Fig. 1 assumed that the same number of sequence reads were used in both methods. The more individuals in a pool, the more precise was the estimation of allele frequencies. A pool containing just 50 individuals showed more precise allele frequencies estimates than WGS in some cases (Fig. 1a) (Schlötterer et al., 2014). However, when the number of individuals sequenced separately in WGS was the same as the pool size of 50 individuals in Pool-seq, WGS provided more accurate estimates of allele frequencies. Furthermore, Pool-seq was more accurate than WGS when the coverage per sequenced individual increased (Fig. 1b). However, in most simulations the pool sizes were significantly higher than the amount of individuals used in WGS, which had an impact on which method was the most precise.

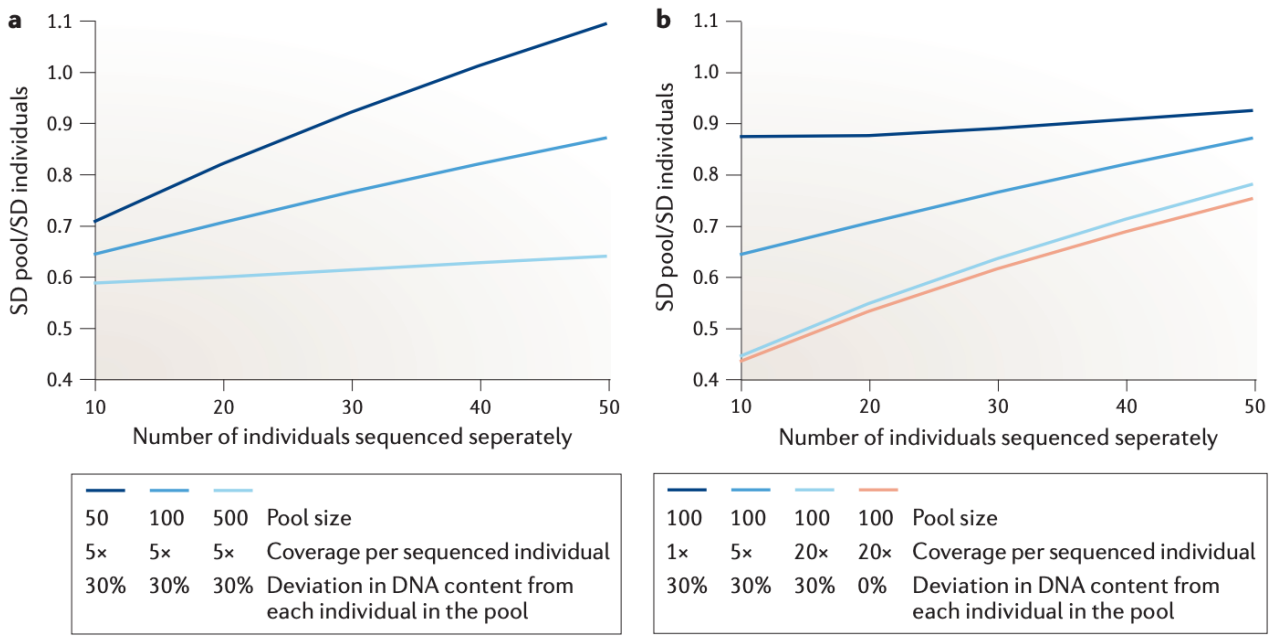


Fig. 1. Comparison of cost-effectiveness of Pool-seq and WGS. Y-axis indicates the standard deviation of allele frequency estimates. When this value is below 1, Pool-seq is the most precise method. X-axis indicates the number of individuals sequenced using WGS (Schlötterer et al., 2014).

Another cost-effective alternative is low-coverage whole genome sequencing (lcWGS) (Lou et al., 2021). This method has a high width of coverage, as it provides information of the entire genome, and because the information is being spread across the whole genome of separate individuals, but at low coverage at each site. It is still possible to obtain individual genotypes with this method, but the certainty of the allele frequency estimates is low due to the low coverage (Lou et al., 2021).

Whole genome sequencing identifies individual genotypes in a whole population (Lou et al., 2021). WGS with a high sequencing coverage of each individual genome is without question the most precise when it comes to high-quality population-wide data, but it is an expensive alternative, especially for species without a reference genome. Pool-seq and lcWGS are efficient and cheaper alternatives, especially when it comes to examining genetic relationships at a population scale (Lou et al., 2021; Schlötterer et al., 2014). Pool-seq and lcWGS require high quality reference genomes of the target species or a closely related species (Lou et al., 2021) but due to the cost-effectiveness, advances in NGS technology, initiatives such as Earth Biogenome Project (Lewin et al., 2018), reference genomes accumulate at a higher rate (Schlötterer et al., 2014).

Since *A. glacialis* lives in cold and remote Arctic waters (Mecklenburg et al., 2018), little is still known about its biology and population genetic structure. This is important knowledge for understanding which impact a warming Arctic and thawing sea ice have on the ice cod and its surrounding ecosystems. Microsatellites have previously been used to study population differentiation and species identification in *B. saida* and *A. glacialis* (Nelson et al., 2013). Moreover, mitochondrial variation (mtDNA) has been studied for the ice cod and two other Arctic gadoids *Boreogadus saida* and *Gadus ogac* (Pálsson et al., 2009). Until now, it has been challenging to detect population structure between different populations for the ice cod, as the focus regarding Arctic gadids has mostly been on its co-existing counterpart *B. saida*.

This present project uses WGS to assess the allele frequency variation in a unique collection of *A. glacialis*, which have been obtained from eight different fjord and shelf localities in NE Greenland. The aim of this project is to investigate if *A. glacialis* displays population genetic structure between two different habitat types (fjord vs. shelf). Population genetic structure within fjord localities and shelf localities, respectively, will also be investigated.

Materials and methods

Between 2013–2022, genetic samples of 417 individuals were obtained from eight different localities (Belgicabanken, Besselfjord mid, Besselfjord mouth, Besselfjord shelf, Bredefjord, Dove Bugt, Moskusoksefjord and Tyrolerfjord) in NE Greenland (Fig. 2), collected through the TUNU Programme (UiT) (Christiansen, 2012).

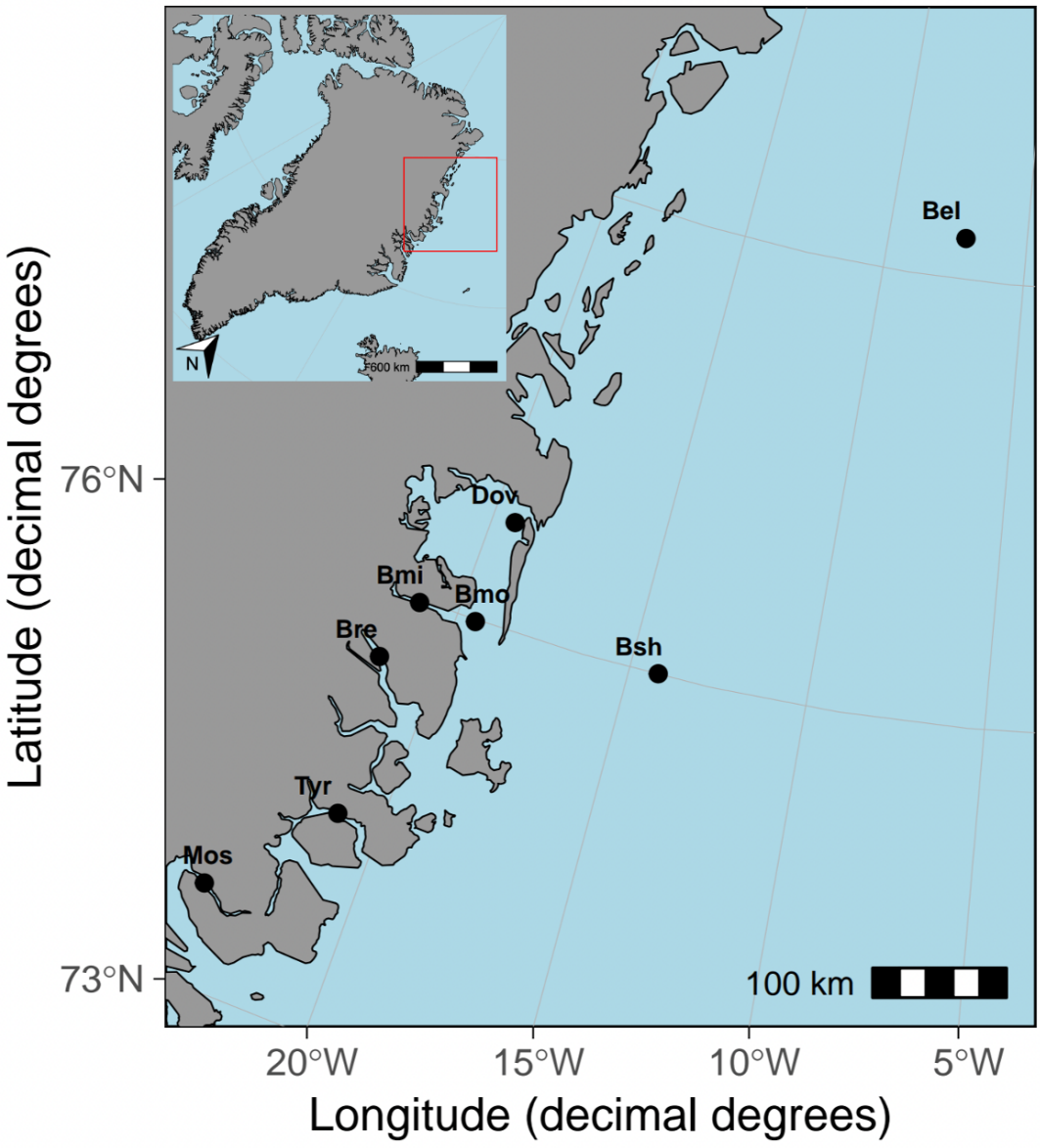


Fig. 2. Overview of the eight different sampling localities in NE Greenland. Bel = Belgicabanken, Bsh = Besselfjord shelf, Dov = Dove Bugt, Bmo = Besselfjord mouth, Bmi = Besselfjord mid, Bre = Bredefjord, Tyr = Tyrolerfjord, Mos = Moskusoksefjord. The inserted small map illustrates Greenland, where the area of the localities is highlighted.

In 2013, specimens of *A. glacialis* were caught from Moskusoksefjord, Bredefjord and Tyrolerfjord. In 2017, it was Belgicabanken and the three Besselfjord localities: Besselfjord mid, Besselfjord mouth and Besselfjord shelf. Lastly in 2022, catches in Dove Bugt and Bredefjord were obtained (Table 1). Note that samples from Bredefjord were carried out in 2013 and 2022, respectively.

Out of the eight localities listed in Table 1, three of them were characterized as shelf areas (Belgicabanken, Besselfjord mouth and Besselfjord shelf). If a locality was geographically connected to the open sea and/or located close to a shelf, a locality was suggested to be a shelf area. Belgicabanken and Besselfjord shelf were two known offshore (Fig. 2) shelf areas (Budéus et al., 1997; Pers. com. Kim Præbel). Besselfjord mouth was an inshore locality but connected to the open sea, as the locality was sampled on the east side of a major sill. The sill was located at the entrance of Besselfjord, which ranged from 50–200 m below the sea surface (Zoller et al., 2023).

The rest of the localities were characterized as fjord areas (Besselfjord mid, Bredefjord, Dove Bugt, Moskusoksefjord and Tyrolerfjord). If sill(s) were present between the inner part of a fjord and the open sea and/or if the sampling was made relatively far into a fjord, a locality was suggested to be a fjord area. Besselfjord mid was sampled on the west/inner side of the above-mentioned sill in Besselfjord. Bredefjord was sampled far into the fjord (Fig. 2). Dove Bugt was a bay (Olsen et al., 2020) but was characterized as a fjord in this study due to sampling far into the bay. Moskusoksefjord contained a sill (Olsen et al., 2022) and the same applied for Tyrolerfjord, as specimens in Tyrolerfjord were caught between two sills ranging up to 60 m below the sea surface (Ribeiro et al., 2017).

DNA extraction and quality assessment

DNA extractions from fin and muscle tissue were carried out from 182 of the 417 individuals. A total of 26 individuals per station were selected except Belgicabanken with 14 individuals and Dove Bugt with 13 individuals due to too small amounts of tissue in the rest of the samples. For Bredefjord 2013, 14 individuals were selected, whereas for Bredefjord 2022, 12 individuals were selected (Table 1). Since these were later merged, the sum of individuals extracted from Bredefjord was 26. The thought behind the number of extracted individuals was to leave room for samples with low DNA quality, which then could be left out from sequencing without having to extract DNA a second time.

*Table 1. Overview of localities, the year they were sampled, geographic positions, habitat type (fjord vs. shelf), number of extractions and number of samples sent for sequencing (N = 119). Bredefjord 2013 and Bredefjord 2022 were the same locality but sampled in two different years. SR 1 = shipment round one, SR 2 = shipment round two, Seq = samples sent for sequencing. * Indicates the location of replicates.*

Locality	Year	Position	Habitat	Extracted	SR 1	SR 2	Seq
Belgicabanken*	2017	79.28° N -07.22° E	Shelf	14	2	7	9
Besselfjord_mid	2017	75.98° N -21.09° E	Fjord	26	13	3	16
Besselfjord_mouth	2017	75.99° N -19.46° E	Shelf	26	8	8	16
Besselfjord_shelf	2017	76.01° N -14.19° E	Shelf	26	4	11	15
Bredefjord	2013	75.54° N -21.58° E	Fjord	14	4	5	9
Bredefjord	2022	75.54° N -21.58° E	Fjord	12	3	3	6
Dove_Bugt*	2022	76.70° N -19.33° E	Fjord	13	7	6	13
Moskusoksefjord	2013	73.69° N -23.47° E	Fjord	26	10	6	16
Tyrolerfjord*	2013	74.46° N -21.10° E	Fjord	26	8	11	19

The protocol, Purification of Total DNA from Animal Tissues (Spin-Column Protocol) from Qiagen (Appendix 1) was utilized for the first 18 individuals to get familiar with the protocol before making plate extractions. The first deviation from the protocol happened after cutting up to 25 mg tissue, where the tissue was dried for ethanol before it was transferred to a 1.5 ml microcentrifuge tube, which otherwise would have overestimated the weight of the tissue. The drying for ethanol was not listed in the protocol.

Another deviation was the adding of 4 µl RNase A (100 mg/ml), which was an optional step before step 3 (Appendix 1). This was done for all samples to avoid RNA contamination later in DNA quality assessment. Four different approaches in step 7 and 8 were also tested:

1) 200 µl Buffer AE acclimated to room temperature was added at once onto the DNeasy membrane without repeating the elution.

2) Buffer AE was heated beforehand to 56°C, but 100 µl Buffer AE was added. This was repeated as recommended in step 8, giving a total of 200 µl Buffer AE.
3) Buffer AE was heated beforehand to 56°C, but 100 µl Buffer AE was added at once without repeating the elution in step 8.
4) Buffer AE was heated beforehand to 56°C, and 50 µl Buffer AE was added. This was repeated as recommended in step 8, giving a total of 100 µl Buffer AE.
Afterwards, plate extractions were made for the rest of the samples by following the DNeasy 96 Protocol (Appendix 2). In step 16 the elution was made with 200 µl Buffer AE as listed in the protocol, but the elution was not repeated with another 200 µl AE as recommended in step 17.

To quality check the DNA extractions, gel electrophoresis, Qubit and NanoDrop were carried out. For gel electrophoresis, 0.4 g Agarose was added to a flask containing 50 ml TAE Buffer and was afterwards heated until the Agarose gel dissolved. 5 µl gel-red was then added to the mixture, poured into a vial after mixing and left to set. Meanwhile, the ladder mixture was made with a total volume of 6 µl between dye, ladder and water, whereas the sample mixture also had a total volume of 6 µl but between sample, dye and water. The amount of ladder, water and sample was adjusted if the ladder or sample concentration was too high. After the mixtures were added to the gel vials, the gel ran for 15 minutes at 200 V or 30–40 minutes at 140 V.

Qubit 1X dsDNA BR Assay Kit using a Qubit 4 Fluorometer (Web link 1) was applied to measure DNA concentration in the samples. The user manual provided by the external provider Thermo Fisher Scientific was followed (Appendix 3).

Lastly, spectrophotometric quantification of 1 µl of every sample was made using NanoDrop to measure absorbance ratios at 260/280 nm and 260/230 nm. An absorbance ratio of approximately 1.8 is optimal for nucleic acids at 260/280 nm (Web link 2). However, if the value is significantly lower than 1.8, it may indicate e.g. protein contamination or a DNA concentration lower than 10 ng/µl. At 260/230 nm the absorbance ratio should range between 2.0 and 2.2. A low ratio may indicate the presence of a contaminant absorbing at 230 nm or less, or if the solution of the blank measurement is not the same as in the sample (Web link 2). In this case, Buffer AE was used both as the blank measurement and in the sample solution.

Sample preparation for whole genome sequencing

Before shipping the samples for sequencing to the external service provider NOVOGENE, Co, Ltd., some requirements needed to be met. Qubit ≥ 200 ng, volume ≥ 20 µl, concentration ≥ 10 ng/µl, 260/280 = 1.8–2.0. Furthermore, no degradation or contamination should be present, when samples were assessed on an Agarose gel (Appendix 4). Those samples that met these requirements were sent in two rounds for next generation sequencing (using PE150 chemistry on a NovoSeq 6000) at the external service provider aiming for 10X genome coverage. As no genome size estimate were available for *A. glacialis*, the genome size estimate of 0.88 Gb for the closely related *Boreogadus saida* was used for planning the sequencing (Hardie & Hebert, 2003). Among these samples, two replicates were sent for one individual from Dove Bugt and Belgicabanken, respectively, and two replicates for each of two individuals from Tyrolerfjord. The replicates were split up in the two different shipments to test for batch effect and human error.

A total of 119 samples (111 different individuals) were sent for WGS (Table 1). A total volume of 20–30 µl of each sample was transferred to 1.5 ml Eppendorff low bind tubes, wrapped in parafilm and shipped with blue ice.

Bioinformatic analysis

The demultiplexed raw reads were downloaded from NOVOGENE, Co, Ltd. and further quality filtered for uncalled bases (Ns) and presence of sequencing adapters and sequence length. Cutadapt v3.5 (Martin, 2011) in paired-end mode with following options was used: Upper limit of Ns in a sequence (*--max-n 8*), minimum sequence length in base pair (*-m 75*) and minimum base quality (*-q 20*).

The quality trimmed paired-end reads were then mapped to an existing genome of Atlantic cod (*Gadus morhua*) (gadMor3.0, GCA_902167405.1) using MEM algorithm employed in Burrows-Wheeler Aligner (Li & Durbin, 2009). Further SAMtools view and SAMtools sort (Li et al., 2009) were used to filter alignments with mapping score less than 20 (*-q 20*) out. The resultant BAM files were used for SNP calling and genotyping in ANGSD v0.940 (Korneliussen et al., 2014) with following modifications: Removal of bad alignments, PCR duplicates (*-remove bads 1*), keeping only alignments where read pairs aligned in proper direction (*-only proper pairs 1*), minimum mapping quality of 20 (*-minMapQ 20*) and presence

of SNP genotype in at least 101 (*-minInd 101*). Afterwards, SAMtools genotype likelihood model (Li, 2011) was used to call the SNPs. Major and minor alleles from genotype likelihood were inferred, considering only those polymorphic sites that presented a SNP *P* value ≤ 1e-6 and genotype depth of 9X (*-geno_minDepth 9*), minor allele frequency (MAF) of 5% (*-minMaf 0.05*) and output genotypes into basic BIM Collaboration Format (BCF) (*-bcf 1*) and beagle format (*-doGlf 2*). The basic BCF format was used (genotypes with minimum 9X coverage) for all downstream analysis.

The BCF file generated by ANGSD was further converted to Variant Call Format (VCF) using BCFtools (Li, 2011). Then basic quality statistics for VCF file such as missing data per individual, missing data per SNP loci, mean depth of coverage per SNP loci and mean depth of coverage per individual across all the SNPs loci were calculated using VCFtools (Danecek et al., 2011). Based on these outputs and some additional filtering criteria (see below), the SNPs were filtered. Removal of indels (*--remove-indels*), MAF of 0.05, maximum non missing data of 80%, (means genotype should be called in at least 80% of individuals), minimum mean depth of 11 (*--min-meanDP*) and maximum mean depth of 63 (*--max-meanDP*) (twice the mean depth). Finally, the SNPs were filtered for linkage disequilibrium. Plink v1.9 (Purcell et al., 2007) was utilized to prune the SNPs that were in linkage (r^2) (*--indep-pairwise 50 5 0.2*). If two SNPs were under linkage disequilibrium, the SNP with lowest MAF would be discarded.

The quality filtered VCF data was divided into all SNPs and F_{st} based data set. The all SNPs data set consisted of all the quality filtered (see above) SNPs, whereas the F_{st} based data set was generated using locus-wise F_{st} calculated in outFLANK v0.2. This was done to remove the SNPs with low information content. The VCF file was converted into outFLANK compatible genotype matrix using vcfR package (Knaus & Grünwald, 2017) in R (Web link 3). The function MakeDiploidFSTMat in outFLANK v0.2 (Whitlock & Lotterhos, 2015) was then used to calculate locus-wise F_{st} (calculated across sampling localities). R generic function quantile was used to output 90th percentile for F_{st} distribution/10% highest F_{st} values (F_{st} 90%) and for second data set generation. The F_{st} 90% data set was the basis of all results in population structure analysis except Appendix 5 (full data set) and 6 (F_{st} above 0).

Replicate analysis was performed using VCFtools and plink v1.9. Firstly, replicate sample specific VCF file was generated. The resulting VCF file was then processed with plink v1.9 to generate eigenval- and eigenvec files for later Principal Component Analysis (PCA). PCA were made in plink v1.9 using *--pca* and *--allow-extra-chr option*. PCA plots were made in R using tidyverse package (Wickham et al., 2019).

Population genetic structure analysis

Pairwise F_{st} between localities and the corresponding *P* values were calculated using StAMPP package (Pembleton et al., 2013) in R. Number of bootstraps were set to 100, confidence interval to 95% and number of cores set to six. To test if Bredefjord 2013 and Bredefjord 2022 could be merged, the *P* value between the two was assessed, which was followed by a calculation of the Bonferroni coefficient to increase the confidence.

PCA plots were generated with PCA output and tidyverse package in R. Both PCA figures were based on F_{st} 90% but further PCA plots of $F_{st.}$ above 0 (Appendix 6) and a PCA containing all SNPs were also included (Appendix 5) for comparison purposes. Discriminant Analysis of Principal Components (DAPC) were made with adegenet (Jombart, 2008) package in R and both analyses were based on 10 principal components (PC).

To support the results from PCA and DAPC, a neighbor-joining (NJ) tree with Nei's Distance were carried out in R using the poppr package (Kamvar et al., 2014). The tree was made with 1,000 bootstraps and a cutoff value of 50 to specify when the program should return the bootstrap values on the nodes. It was afterwards edited in Figtree (Web link 4) for cosmetic purposes.

Along with PCA, Admixture analysis was also made to identify the genetic groups. The F_{st} 90% VCF file was converted to plink format using plink v1.9. Chromosome names were then changed to numbers. Admixture v1.3.0 with *--cv 9* option was used. K = 1–9 was tested and the value with lowest cross-validation (CV) scores was accepted as putative number of genetic clusters represented in the dataset.

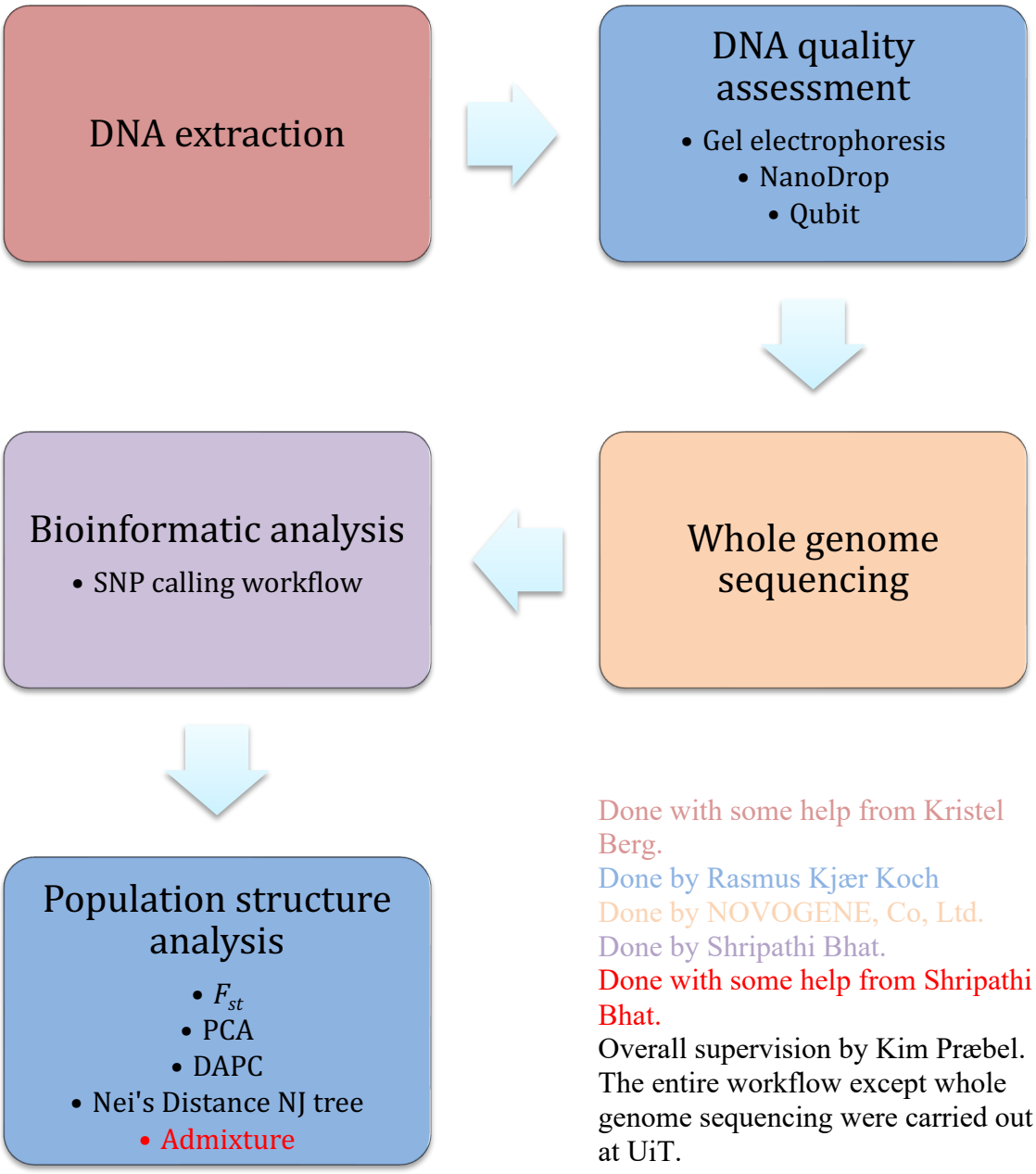


Fig. 3. Workflow of the study.

Results

Post-filtering- and sequencing results

Sequencing of 159 samples (111 different individuals), which included re-sequenced samples and eight replicates, yielded an average of 27,646,256 paired-end reads with a standard deviation of 13,340,125. After the filtering, a total of 542,520 SNPs were left and one individual with 99% missing data was removed. Also, only one sample from each replicate pair was kept in the final dataset (based on lowest missing value), which resulted in a total of 110 individuals that were used for further population structure analysis (Table 2).

Locus-wise global F_{st} analysis showed a minimum F_{st} of -0.055, mean F_{st} = 0.001 and maximum F_{st} = 0.314. Subsequently, a quantile analysis was made, which showed F_{st} 90% = 0.021. This indicated that the 10% highest F_{st} values were above F_{st} 0.021 and gave a total of 54,252 SNPs, which were the basis of all population genetic structure analyses (except Appendix 5 and 6).

Bredefjord 2022 and Bredefjord 2013 were merged to Bredefjord, as the *P* value of the pairwise F_{st} analysis was 0.05 and the Bonferroni coefficient was also 0.05. Therefore, there was no significant difference between Bredefjord 2022 and 2013.

Table 2. Number of individuals per station left post-filtering- and replicate removal (N = 110). Three localities were characterized as shelf areas, and five localities were characterized as fjord areas.

Locality	Total individuals	Habitat
Belgicabanken	7	Shelf
Besselfjord_mid	16	Fjord
Besselfjord_mouth	16	Shelf
Besselfjord_shelf	15	Shelf
Bredefjord	15	Fjord
Dove_Bugt	11	Fjord
Moskusoksefjord	15	Fjord
Tyrolerfjord	15	Fjord

Population genetic structure analysis

Pairwise F_{st} values between the eight localities were plotted as a correlation matrix heatmap (Fig. 4). F_{st} values ranged from 0.026–0.061 and had a mean value of 0.039. Belgicabanken displayed the highest F_{st} values when compared to the other localities, as values ranged from 0.041–0.061. When compared to fjord areas such as Tyrolerfjord, Dove Bugt and Besselfjord mid, the highest values were shown with F_{st} 0.052, 0.058, 0.061, respectively. However, when compared to the two other shelf localities Besselfjord shelf and Besselfjord mouth, F_{st} values were lower, showing values of 0.041 and 0.044 but when looking at all the pairwise F_{st} values, these differences were still higher than the average F_{st}. Furthermore, the comparison between Besselfjord mouth and Besselfjord shelf displayed low differentiation (F_{st} 0.027). The same went for Bredefjord compared to Besselfjord mouth (F_{st} 0.026) and Tyrolerfjord compared to Moskusoksefjord (F_{st} 0.028). *P* values were also calculated for each of them, but since all *P* values were 0.000, they were left out of Fig. 4. With a *P* value of 0.000, there was a significant difference in F_{st} values between all the localities.

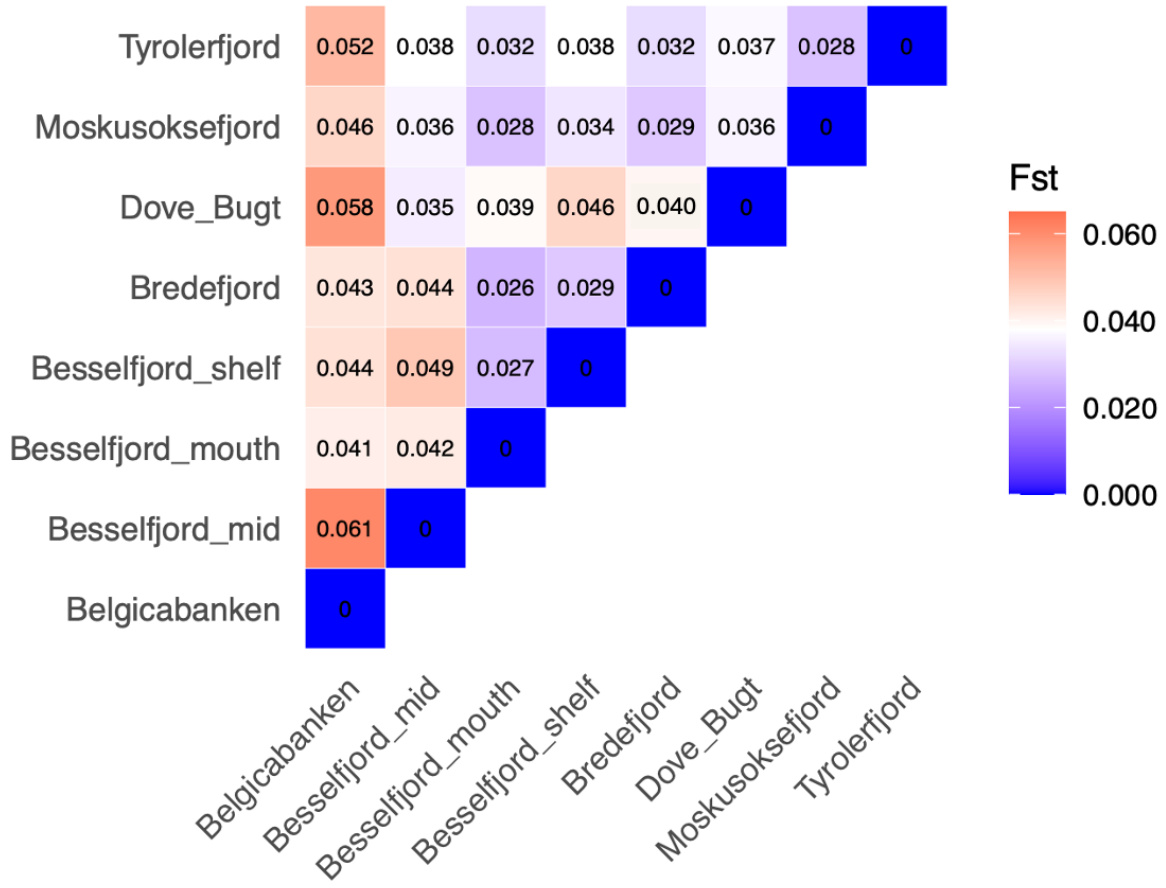


Fig. 4. Correlation matrix heatmap of pairwise F_{st} values. Comparisons between the same two localities were set to 0 (blue). Scale bar of F_{st} values was set to a middle value of 0.038 (white) and a maximum F_{st} value of 0.065 (red).

The genetic similarities, which were observed within the shelf areas in pairwise F_{st} analysis, were also displayed in PCA. In the following PCA plot (Fig. 5) the two offshore localities Belgicabanken and Besselfjord shelf clustered tightly together, which were followed by a cluster of most individuals from Besselfjord mouth. However, two individuals from Besselfjord mouth were genetically different from the rest of the individuals from Besselfjord mouth, as they mixed with Moskusoksefjord and Dove Bugt. Individuals from Besselfjord mid were from the same fjord as Besselfjord mouth but had been caught further in the fjord when comparing these two localities. However, despite the short geographical distance between these two, Besselfjord mid clustered far away from Besselfjord mouth.

Further down, Bredefjord showed a similarity in allele frequencies with individuals from Besselfjord mouth and Moskusoksefjord, which indicated that individuals within Bredefjord were very likely to be genetically differentiated. Moskusoksefjord showed the same trend, as they were also mixed with Dove Bugt. Dove Bugt revealed the highest genetic divergence within the locality. Lastly, Tyrolerfjord showed a differentiation from the other localities, but the genomic distance within the locality was still noticeable.

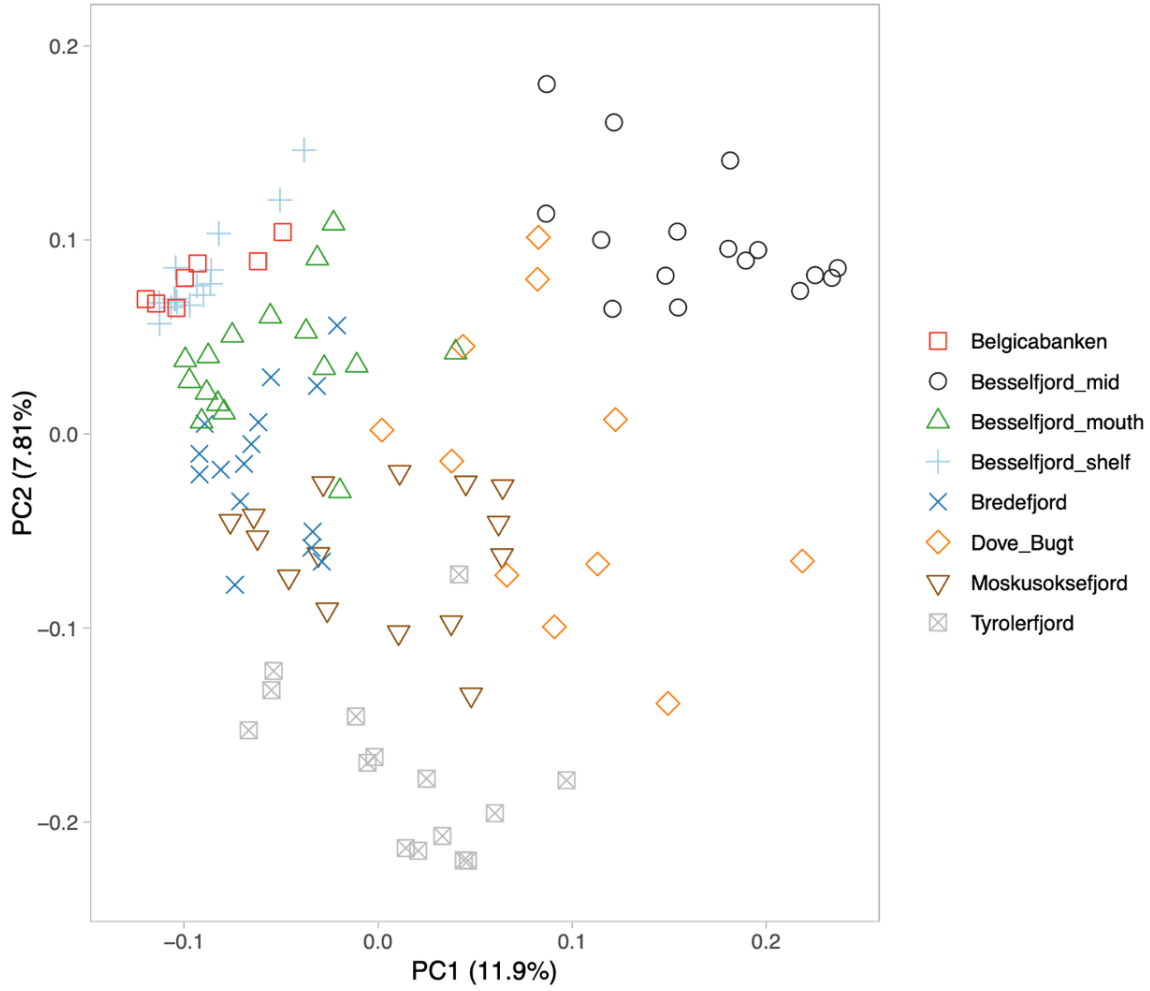


Fig. 5. PCA plot. The plot shows the variation in allele frequencies in the eight different localities, with PC1 explaining 11.9% of the variation in allele frequencies in the data set and PC2 explaining 7.81%.

When only assessing fjord and shelf areas, the PCA showed a clear cluster of the shelf areas compared to the fjords (Fig. 6). The plot was identical to the PCA in Fig. 5 but by highlighting the two different habitats, the genetic difference between individuals from the fjords and shelf areas became even more clear, even though few individuals were not following this trend.

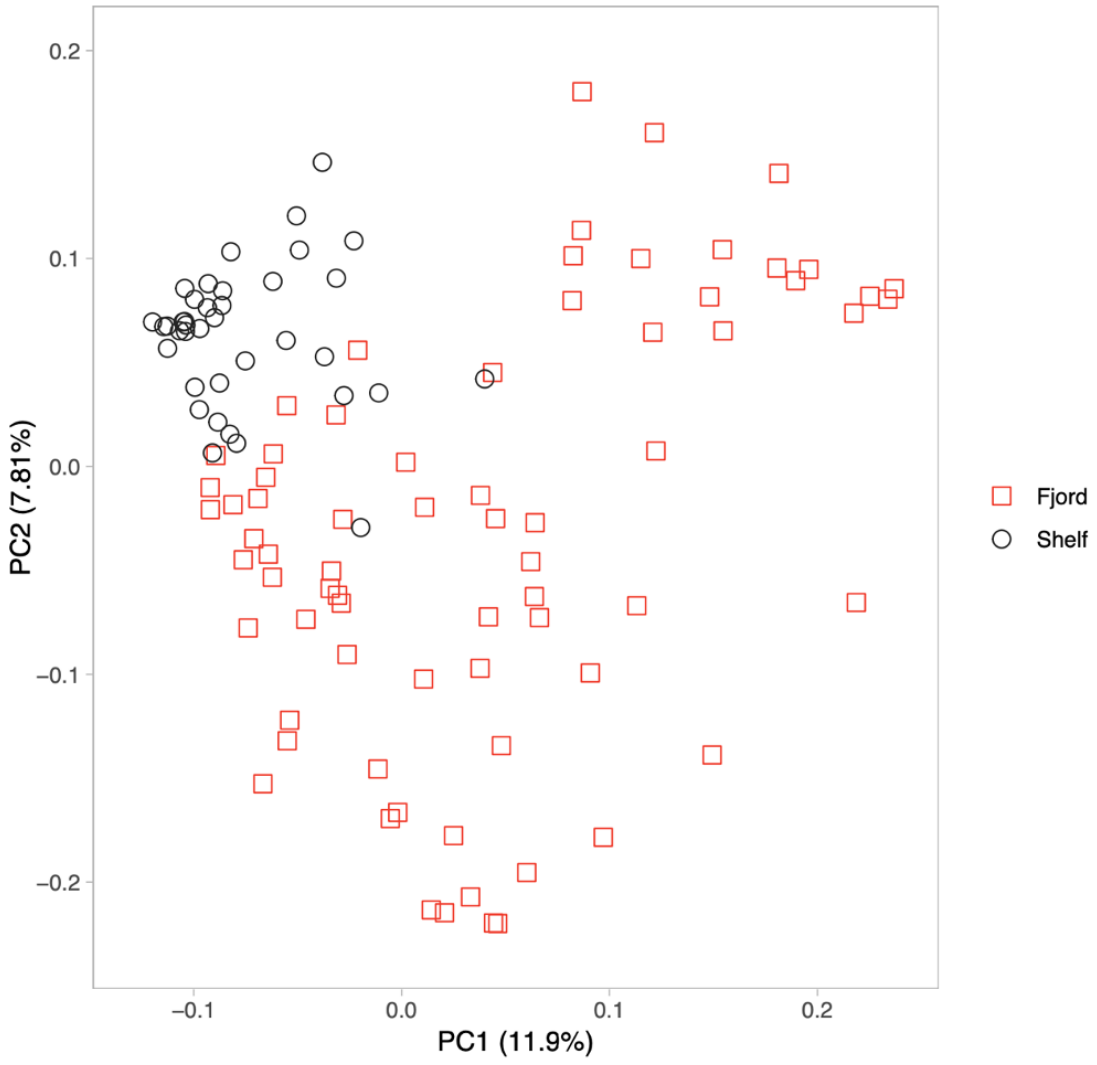


Fig. 6. PCA plot showing the genetic variation between the two habitat types; Fjord vs. shelf. Belgicabanken, Besselfjord shelf and Besselfjord mouth were grouped as shelf (red squares), whereas Besselfjord mid, Bredefjord, Dove Bugt, Moskusoksefjord and Tyrolerfjord were grouped as fjord (black circles).

As there were only two groups in the data input (fjord vs. shelf), it was not possible to get a scatter plot in Fig. 7 as in Appendix 7. The alternative was a density plot (Fig. 7), where in this case the genetic differentiation between fjord and shelf localities was compared. Individuals from fjord localities showed a genetic difference from the shelf localities.

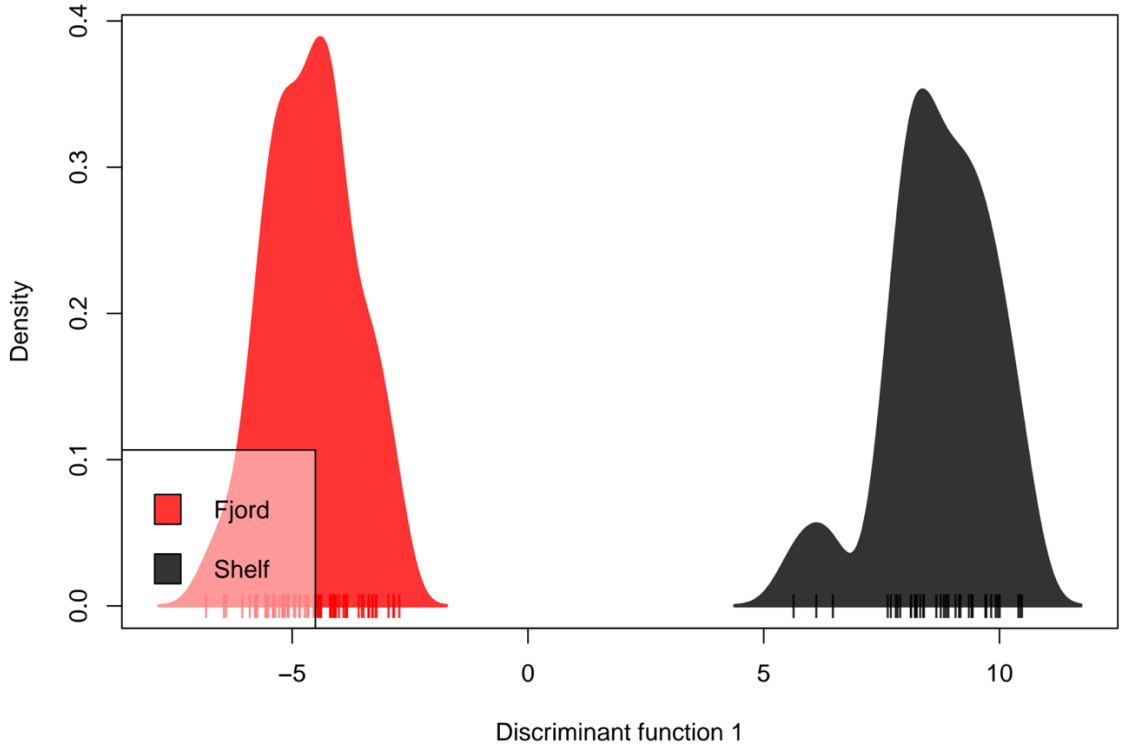


Fig. 7. DAPC density plot. Fjord localities are marked in red, whereas shelf localities are marked in black. 10 PCs were retained, which explained 64.69% of the total allele frequency variation.

In order to support the results in PCA and DAPC, a Nei's Distance NJ tree was made, which illustrated a clear difference between shelf and fjord localities (Fig. 8). Within the shelf area, Besselfjord shelf and Belgicabanken clustered together with a high bootstrap support (100%) and were followed by Besselfjord mouth (71.1%). Within the fjord area the bootstraps were all 100%. Furthermore, Dove Bugt and Besselfjord mid grouped together, which were more differentiated with Moskusoksefjord than with Tyrolerfjord, but the overall trend was the difference between fjord and shelf areas.

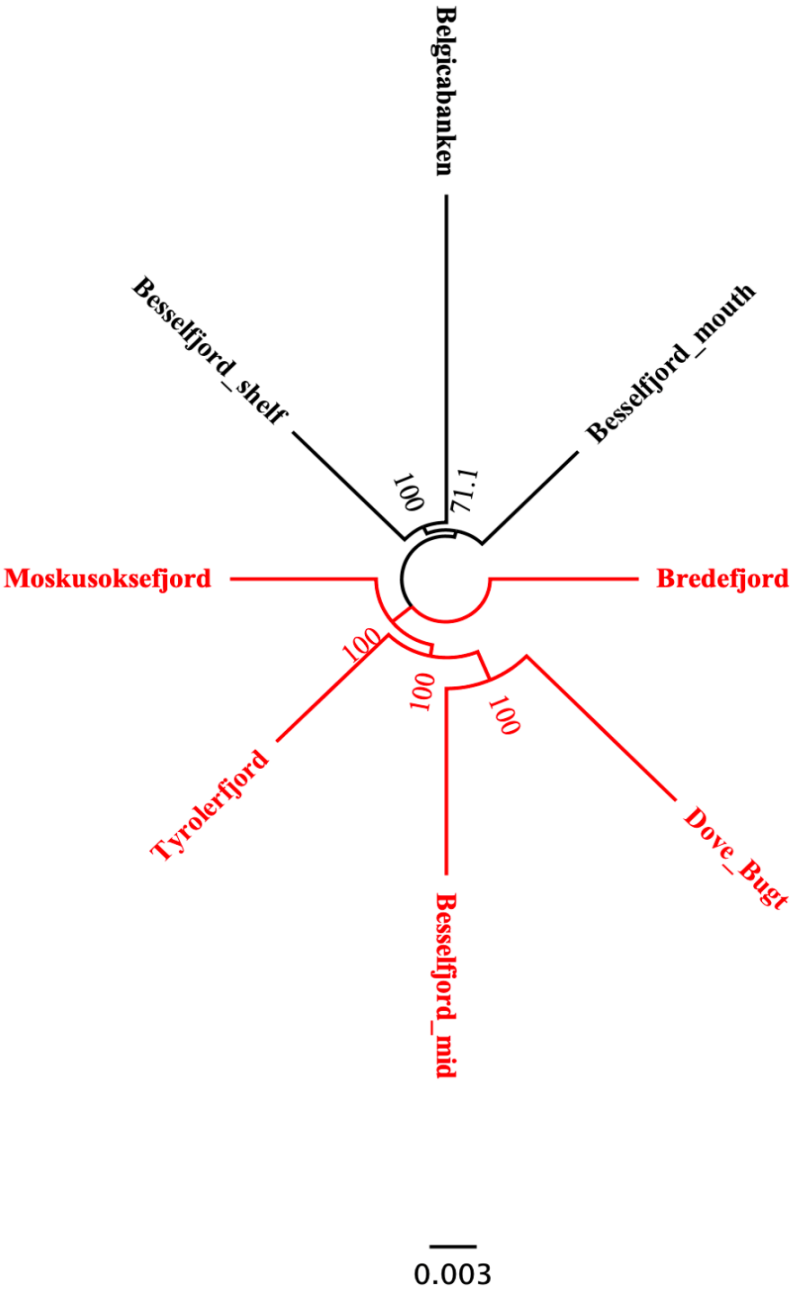


Fig. 8. NJ tree of all the locations using Nei's Distance. Bootstrap values are based on 1,000 iterations and scale bar indicates changes in Nei's Distance. Fjords and shelf areas are highlighted and grouped in red and black, respectively.

Admixture analysis revealed that the most likely K was K = 1 (Fig. 9), as it had the lowest CV value of 0.43, which indicated no population structure. However, at K = 2 the CV value was 0,00013 higher, indicating that K = 2 was also likely. At K = 2, Besselfjord mid showed genetic differentiation from the other localities, and shelf localities showed high genetic

similarities within them and with Bredefjord. Tyrolerfjord, Dove Bugt and Moksusoksefjord revealed a mix of the two clusters, and in K = 3 some of the properties were replaced with genetic material from Tyrolerfjord, especially in the other fjords. Here, Tyrolerfjord showed high genetic divergence from the rest of the localities. In K = 4, Moskusoksefjord, Tyrolerfjord, Besselfjord mid and shelf localities + Bredefjord revealed population genetic structure between them, while Dove Bugt displayed mixed properties. When assessing higher K values such as K = 7 and K = 8 (Appendix 8), all localities displayed genetic differentiation.

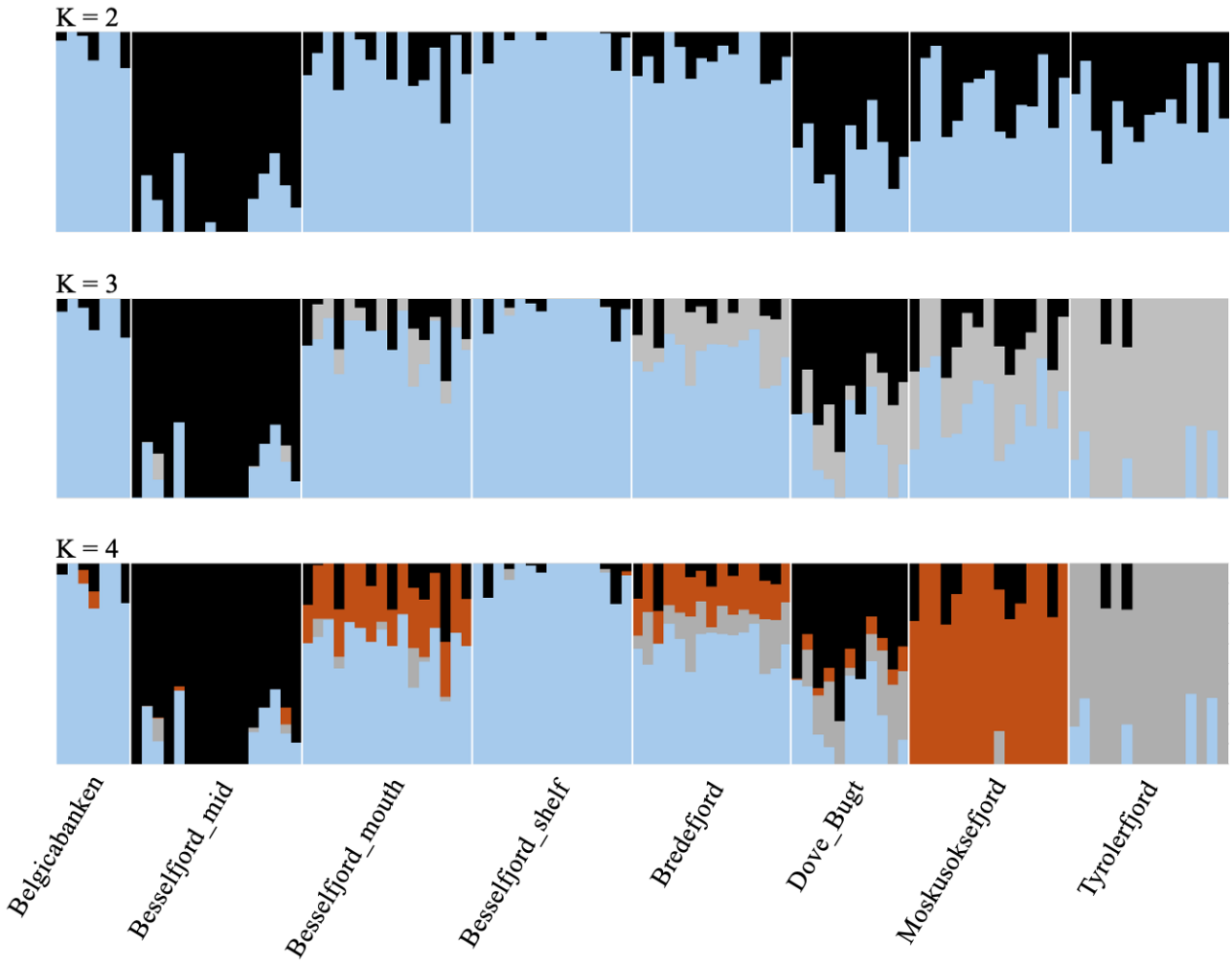


Fig. 9. Admixture analysis showing the spatial distribution of genetic clusters for K = 2, K = 3, K = 4. Color codes match with those in PCA (black = Besselfjord mid, blue = Besselfjord shelf, brown = Moskusoksefjord, grey = Tyrolerfjord.

Discussion

The aim of this study was to investigate population genetic structure for *A. glacialis* between two habitat types (fjord and shelf areas) in NE Greenland, and if population structure was displayed within each of the two habitats. The results revealed an overall genetic differentiation between the two habitat types, which was the main objective in this study.
Within the fjords, population structure analyses revealed three different conclusions. Firstly, significant genetic differences between all fjord localities were found in pairwise F_{st} analysis. Secondly, PCA and Admixture at K = 3 indicated genetic structure between a few localities and the rest. Lastly, no overall genetic difference was observed in DAPC density plot and Nei's Distance NJ tree, as the fjords grouped together. Population structure analyses within the three shelf localities revealed that they were genetically similar, except the pairwise F_{st} analysis, which showed significant genetic differences.

Population genetic structure between shelf and fjord areas
Even though the results suggested some genetic differences within fjord and shelf areas, respectively, the differences were most pronounced between the two habitats, which was firstly due to clustering of shelf localities in PCA. This was even clearer in the DAPC density plot and further supported in Nei's Distance NJ tree, which showed a clear separation of the two habitat types. The observed population structure between fjord and shelf areas may be related to the post-glacial history of *A. glacialis,* where some individuals from an ancestral population colonized the fjords. Later, they became isolated from the shelf populations due to formation of sills and changes in sea level, which has been suggested for *B. saida* (Madsen et al., 2016). In Madsen et al. (2016), it was suggested that the population genetic structure between fjord and shelf areas in NE Greenland and West Svalbard was related to an ancestral population, which were present at lower latitudes during Last Glacial Maximum (LGM) ca 21 kilo annum (ka) (Bigg et al., 2008). At that time, NE Greenland fjords were fully glaciated (Zoller et al., 2023) and shelf habitats in the North Atlantic were lost due to a 120–135 m decrease in sea level (Bigg et al., 2008). After the deglaciation, *B. saida* was then hypothesized to colonize the Arctic fjords, which would explain the observed genetic structure between the two different habitats.

Such post-glacial expansion has also been suggested for another cod species, as the Northwest Atlantic *G. morhua* may have expanded to the Canadian Arctic between 8–5 ka (Hardie et al., 2006).

Zoller et al. (2023) made an overview of the deglaciation events from their findings and previous publications from marine and terrestrial data. It showed that Dove Bugt and Besselfjord, that were sampled in this study, deglaciated 8.6–12.8 ka, thus indicating that a possible post-glaciation colonization of *A. glacialis* might have taken place shortly after this event. However, deglaciation events have not been determined for the other six localities sampled in this study but due to the relatively short distance between the localities, it might also apply for the rest of the localities, but that remains hypothetical. Since there are no records of the ice cod from before, during or shortly after LGM, it is not possible to determine where the ancestral population expanded from, and how they inhabited the fjord and shelf areas, which could explain the population genetic structure. However, the suggestion that an ancestral population colonized the fjords and later became isolated from the shelf areas, might not be the only explanation for the genetic structure.

The genetic structure might also be due to an adaptation to the conditions in their respective fjord and therefore makes migration to the shelf areas outside the fjord unlikely. Such scenario has been suggested between two flounders, (*Platichthys flesus* and *Platichthys stellatus*) as the authors concluded that sea temperature possibly served as a barrier to gene flow (Borsa et al., 1997). Temperature gradients have previously been studied near one of the fjords that were sampled in this project. Temperature and oxygen measurements in August 2014 taken near Tyrolerfjord showed first and foremost a difference in temperature between the inner part of Tyrolerfjord and the shelf outside the fjord (Ribeiro et al., 2017). In the inner part of the fjord, the temperature was approximately -1°C and uniform throughout the water column until the first few m of the surface water, which was 10°C. In the shelf area, the temperature was higher near the seabed from 0–1°C compared to the inner fjord.
Furthermore, oxygen saturation measurements in the inner fjord fluctuated between undersaturated (70%) levels near the seabed to slightly oversaturated levels in the upper few m of the surface, whereas oxygen levels in the shelf area were close to saturation in the entire water column (Ribeiro et al., 2017). If *A. glacialis* in fjord vs. shelf areas has adapted to these or potentially other environmental conditions, it would possibly explain the observed population genetic structure. However, the measurements from Ribeiro et al. (2017) were a snapshot of the conditions that particular year, and the conditions in Tyrolerfjord may not be a proxy for

the conditions in rest of the localities. Adaptive divergence has previously been hypothesized for the closely related *B. saida* (Madsen et al., 2016) and in other marine fishes such as the European flounder (*Platichthys flesus*) in NE Atlantic (Hemmer-Hansen et al., 2007). Based on microsatellite data, Hemmer-Hansen et al. (2007) suggested that local adaptation was a driver for population genetic structure between two geographically close populations. Also, selection combined with low gene flow have been found as drivers of genetic divergence in *G. morhua* between East and West Greenland (Pampoulie et al., 2011). Temperature and salinity have also been suggested to drive population structure for turbot (*Scophthalmus maximus*) (Nielsen et al., 2004) and European hake (*Merluccius merluccius*) (Cimmaruta et al., 2005). However, no study has yet shown if adaptation to living in fjords or shelf areas is the driver of population genetic structure for *A. glacialis*, but based on findings in other marine fishes, it may be a reasonable explanation.

Although an overall population structure was observed in this study between fjord and shelf areas, Bredefjord showed signs of low genetic divergence when compared to the suggested shelf area Besselfjord mouth. This was shown in pairwise F_{st}, Admixture analysis at low K values and PCA, to some extent. Since Besselfjord mouth was not commonly known as a true offshore/shelf area compared to Besselfjord shelf and Belgicabanken, it may explain the similarities with Bredefjord. On the other hand, Besselfjord mouth was not exactly a fjord either due to the close connection with the open sea. For future work, the locality could be categorized as a bay, together with Dove Bugt, which would result in three different habitat types: Fjord, shelf and bay. Then the population genetic structure between shelf areas and the rest of the localities would likely have been more substantial. This claim is based on the results in PCA and Admixture.

Population genetic structure within shelf and fjord areas, respectively

Pairwise F_{st} comparisons showed a significant genetic difference between all eight localities, however, some trends in differentiation were observed. Belgicabanken for instance, showed the highest genetic differentiation from the other localities, but the comparisons with Besselfjord mouth and Besselfjord shelf showed lower differentiation (F_{st} 0.041 and F_{st} 0.044, respectively) compared to most of the other localities. Furthermore, the pairwise comparison between Besselfjord mouth and Besselfjord shelf showed one of the lowest values (F_{st} 0.027). The relatively low genetic differentiation between Belgicabanken, Besselfjord shelf and

Besselfjord mouth fitted well with the fact that they all were categorized as shelf areas. This was also supported in PCA, Nei's Distance NJ tree and DAPC density plot. The clustering of the shelf localities may be due to a recent divergence and the individuals would therefore have less time to substantially differentiate themselves from each other. Since little is known about *A. glacialis*' spawning grounds, another explanation might be that the shelf areas share the same spawning area, which increases gene flow between them. However, significant *P* values for the pairwise comparisons and Admixture at e.g. K = 7 and K = 8 (Appendix 8) still indicated genetic divergence within the shelf areas.

When comparing the fjord localities with each other, most of the pairwise F_{st} values were below average but still significant. The genetic differences and similarities between these localities were also displayed in PCA. Here, the most conspicuous observation was the high differentiation within Dove Bugt, which has previously been suggested as a spawning area for *A. glacialis* (pers. com. Kim Præbel) and is thus a product of individuals coming from other fjords that have different genetic signatures. Bredefjord also showed genetic divergence within the locality. Bredefjord has a relatively flat and deep seafloor (above 500 m depth) (Arndt et al., 2015; Web link 5) and further out towards the open sea, it is more shallow but has a wide area. This wide entrance of the fjord may enable individuals originating from some of the other localities to use Bredefjord as a feeding area, thus gene flow with other fjord localities could be a factor. Moskusoksefjord also showed genetic differentiation within the locality, but this locality was likely physically constrained due to a sill (Olsen et al., 2022). Since *A. glacialis* is most abundant at 300–400 m depth in the European Arctic (Aschan et al., 2009), it is likely that gene flow between Moskusoksefjord and the other localities is non-existent, when such sill is present. An explanation for the wide genetic divergence within Moskusoksefjord may be that alleles from an ancestral population, that entered the fjords after the previous glacial period, were widely distributed in Moskusoksefjord and therefore still had some genetic similarities with individuals from the other fjords.
Moskusoksefjord not only showed divergence within the locality but also mixing with other localities such as Bredefjord and Dove Bugt. Here, parallel evolution could be a factor if the alleles originating from an ancestral population had adaptive traits, and if the selection pressure for the three localities is the same (due to similar stressors in their environments) for those genomic areas included in this study. Parallel evolution has previously been found in European whitefish (*Coregonus lavaretus*), as the authors found parallel morphological divergence between three populations in northern Norway (Siwertsson et al., 2013). Parallelism in

eco-morphology for cichlid fishes has also been found (Elmer et al., 2014). Since signatures of adaptive divergence was not examined in this study, it is hypothetical.

Moreover, Tyrolerfjord showed genetic differences from the other fjords, as they clustered in PCA and Admixture at K = 3. The individuals from Tyrolerfjord were caught between two sills ranging up to 60 m below the sea surface (Ribeiro et al., 2017), which likely serve as barriers and may thus inhibit gene flow with other populations. The same went for Besselfjord mid, which showed the largest differentiation from the other localities because they, like Tyrolerfjord, were constrained by sills (Zoller et al., 2023). Even though Tyrolerfjord and Besselfjord mid revealed genetic divergence from the other fjords in PCA and Admixture at K = 3, no overall population genetic structure within the fjords was observed in DAPC density plot and NJ tree, as all fjords grouped together.

Furthermore, Besselfjord mid and Dove Bugt grouped together in the NJ tree, but this genetic similarity was also not supported in the other population structure analyses. The grouping was likely due to the results in the PCA, where two individuals from Dove Bugt had similar allele frequencies as the individuals in Besselfjord mid. The inconsistencies in the results within shelf and fjord areas, respectively, may be due to differences in performance of the population genetic software based on the data set.

Methodical discussion

In this study, a good-quality genome from *G. morhua* was used as a reference genome. However, it is ideal to use a species-specific reference genome since mapping to a genome from another species might reduce estimates of variability and therefore affect downstream analysis (Lou et al., 2021).

To remove the SNPs with low information content, only F_{st} 90% were chosen, which meant that pairwise F_{st}, PCA, DAPC, Nei's Distance NJ tree and Admixture were based on 54,252 SNPs, which had the 10% highest F_{st} values. F_{st} 99% could also have been utilized instead and would likely have shown larger genetic differentiations between the localities. However, the preference is subjective, but it is important to keep most of the genetic differentiation without filtering out too many SNPs.

The global locus-wise F_{st} values and the pairwise comparisons may seem low, as F_{st} in theory ranges from 0–1, where F_{st} 0 is no differentiation and F_{st} 1 is the maximum differentiation

(Wright, 1951). In general, marine organisms tend to show low differentiation between populations due to a higher dispersal of gametes and less physical barriers compared to that in terrestrial organisms (Pálsson et al., 2009).

PCA is a great tool for multivariate analyses that easily and quickly can determine genetic structures in large datasets, but it has some limitations (Jombart et al., 2010). When studying population structures, it is important to define the number of clusters, which is not a part of the PCA workflow. Moreover, the PCA plots, used in this study, only utilized the first two PCs. It is possible to make PCA with the first and the third PC instead, but the allele frequency variation in two PCs is only a fraction of the total variation in the data set. DAPC however, uses several PCs and K-means clustering to determine groups of individuals (Jombart et al., 2010).
In this study, DAPC was therefore also made for all individuals (Appendix 7), which retained the 10 first PCs, thus retaining a higher allele frequency variation than in the PCA. The results of this analysis were not coherent with the other results in this study, as e.g. Besselfjord shelf showed similarities with Moskusoksefjord, and Belgicabanken were grouped next to Tyrolerfjord (Appendix 7). It indicated that those eight extra PCs contained very different allele frequencies than the two largest PCs used in PCA. It may also be that the eight PCs had allele frequency variations that were not represented in the rest of the PCs, and by including even more PCs than eight, this "noise" would be evened out. Moreover, the 10 PCs retained in DAPC still only covered 64.69% of the total variance, so theoretically, it is ideal to include all PCs to get the full variation. However, it is not standard procedure to include all PCs, as there are risks of overfitting the discriminant functions (Jombart & Collins, 2015). Future work for DAPC related to *A. glacialis* is therefore needed.

Admixture showed that the most likely K was K = 1 (due to the lowest CV value), thus indicating no population genetic structure. Since the CV value for K = 2 did not deviate much from the value at K = 1, K = 2 could also be most likely, which indicated population genetic structure between Besselfjord mid and the rest of the localities. However, this result was still not coherent with the PCA, as PCA indicated four clusters with Besselfjord mid, shelf localities (Belgicabanken, Besselfjord shelf, Besselfjord mouth) + most individuals from Bredefjord, Tyrolerfjord, the rest of the localities. The K value in Admixture analysis is often underestimated and assuming a true K is always incorrect (Lawson et al., 2018). Admixture should therefore be considered as a supplement to other genetic analyses.

Concluding remarks and future perspectives

Whole genome sequencing of *A. glacialis* revealed overall population genetic structure between fjord and shelf areas in NE Greenland. This may be due to their post-glacial colonization where those individuals, that inhabited the fjords, were later physically constrained due to sill formation and increased sea level. The genetic differentiation may also be due to adaptations to their respective habitat whether it is shelf or fjord. However, Bredefjord did show genetic similarities with the suggested shelf area Besselfjord mouth. Since Besselfjord mouth was not a proper fjord or shelf locality, future work on introducing another habitat type "bay" would likely enhance genetic differentiation between shelf areas and the rest of the localities. Population structure analysis within shelf and fjord areas, respectively, revealed population genetic structure in some analyses and no population genetic structure in other analyses. This could be interpreted as different population genetic software perform differently based on the dataset. Further studies should be conducted to investigate both the genetic basis of local adaptations to similar stressors and environmental conditions using the top 1% SNPs but also the performance of genetic software.

Since these explanations are hypothetical, future work on whether population genetic structure is driven by adaptive divergence in *A. glacialis* and studying the species' ancestral origin are needed. Such knowledge is key for understanding the impact climate change and future habitat loss has on the species.

Acknowledgements

This project was based on collections and sampling from the TUNU Programme (UiT) and because of the hard work of the people involved in that program, I got the best prerequisites to make this thesis. Also, because of the ERASMUS grant, I was able to be an exchange student for 10 months and experience a new environment.

References

Alghamdi, J., & Padmanabhan, S. (2014). Fundamentals of Complex Trait Genetics and Association Studies. Handbook of Pharmacogenomics and Stratified Medicine (1st ed.). Academic Press, 235–257.

Arndt, J. E., Jokat, W., Dorschel, B., Myklebust, R., Dowdeswell, J. A., & Evans, J. (2015). A new bathymetry of the Northeast Greenland continental shelf: Constraints on glacial and other processes. Geochemistry, Geophysics, Geosystems, 16(10), 3733–3753.

Aschan, M., Karamushko, O. V., Byrkjedal, I., Wienerroither, R., Borkin, I. V., & Christiansen, J. S. (2009). Records of the gadoid fish *Arctogadus glacialis* (Peters, 1874) in the European Arctic. Polar Biology, 32(7), 963–970.

Berta, A., Kovacs, K. M., & Sumich, J. L. (2015). Marine Mammals (3rd ed.). Academic Press, 533–595.

Bigg, G. R., Cunningham, C. W., Ottersen, G., Pogson, G. H., Wadley, M. R., & Williamson, P. (2008). Ice-age survival of Atlantic cod: Agreement between palaeoecology models and genetics. The Royal Society, 275(1631), 163–173.

Borsa, P., Blanquer, A., & Berrebi, P. (1997). Genetic structure of the flounders *Platichthys flesus* and *P. stellatus* at different geographic scales. Marine Biology, 233–246.

Bouchard, C., Mollard, S., Suzuki, K., Robert, D., & Fortier, L. (2016). Contrasting the early life histories of sympatric Arctic gadids *Boreogadus saida* and *Arctogadus glacialis* in the Canadian Beaufort Sea. Polar Biology, 39(6), 1005–1022.

Budéus, G., Schneider, W., & Kattner, G. (1997). Distribution and exchange of water masses in the Northeast Water Polynya (Greenland Sea). Journal of Marine Systems, 123–138.

Casillas, S., & Barbadilla, A. (2017). Molecular population genetics. Genetics, 205(3), 1003–1035.

Christiansen, J. S. (2012). The TUNU-Programme: Euro-Arctic Marine Fishes - Diversity and Adaptation. In Adaptation and Evolution in Marine Environments. Springer Berlin Heidelberg, 35–50.

Christiansen, J. S., Hop, H., Nilssen, E. M., & Joensen, J. (2012). Trophic ecology of sympatric Arctic gadoids, *Arctogadus glacialis* (Peters, 1872) and *Boreogadus saida* (Lepechin, 1774), in NE Greenland. Polar Biology, 35(8), 1247–1257.

Cimmaruta, R., Bondanelli, P., & Nascetti, G. (2005). Genetic structure and environmental heterogeneity in the European hake (*Merluccius merluccius*). Molecular Ecology, 14(8), 2577–2591.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. Bioinformatics, 27(15), 2156–2158.

Elmer, K. R., Fan, S., Kusche, H., Luise Spreitzer, M., Kautt, A. F., Franchini, P., & Meyer, A. (2014). Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. Nature Communications, 5.

Frost, K. J., & Lowry, L. F. (1983). Demersal fishes and invertebrates trawled in the northeastern Chukchi and western Beaufort seas. National Oceanic and Atmospheric Administration Technical Report.

Ghigliotti, L., Christiansen, J. S., Carlig, E., Di Blasi, D., & Pisano, E. (2020). Latitudinal cline in chromosome numbers of ice cod *A. glacialis* (gadidae) from Northeast Greenland. Genes, 11(12), 1–13.

Hardie, D. C., Gillett, R. M., & Hutchings, J. A. (2006). The effects of isolation and colonization history on the genetic structure of marine-relict populations of Atlantic cod (*Gadus morhua*) in the Canadian Arctic. Canadian Journal of Fisheries and Aquatic Sciences, 63(8), 1830–1839.

Hardie, D. C., & Hebert, P. D. N. (2003). The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. Genome, 46(4), 683–706.

Harris, H. (1966). Enzyme polymorphisms in man. The Royal Society, 164(995), 298–310.

Hedrick, P. W. (2000). Genetics of populations (2nd ed.). Jones and Bartlett Publishers.

Hemmer-Hansen, J., Nielsen, E. E., Frydenberg, J., & Loeschcke, V. (2007). Adaptive divergence in a high gene flow environment: Hsc70 variation in the European flounder (*Platichthys flesus L.*). Heredity, 99(6), 592–600.

Hollowed, A. B., Planque, B., & Loeng, H. (2013). Potential movement of fish and shellfish stocks from the sub-Arctic to the Arctic Ocean. Fisheries Oceanography, 22(5), 355–370.

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics, 24(11), 1403–1405.

Jombart, T., & Collins, C. (2015). A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0.0. Imperial College London.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genetics, 11.

Jones, O. R., & Wang, J. (2012). A comparison of four methods for detecting weak genetic structure from marker data. Ecology and Evolution, 2(5), 1048–1055.

Jordan, A. D., Jungersen, M., & Steffensen, J. F. (2001). Oxygen consumption of East Siberian cod: No support for the metabolic cold adaptation theory. Journal of Fish Biology, 59(4), 818–823.

Jordan, A. D., Møller, P. R., & Nielsen, J. G. (2003). Revision of the Arctic cod genus *Arctogadus*. Journal of Fish Biology, 62(6), 1339–1352.

Kamvar, Z. N., Tabima, J. F., & Gŕunwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ, 2014 (1), 1–14.

Karamushko, O. V., Lynghammar, A., & Christiansen, J. S. (2022). Ice Cod *Arctogadus glacialis* (Peters, 1874) in Northeast Greenland - A First Sketch of Spatial Occurrence and Abundance. Diversity, 14(11).

Kliman, R. M. (2016). Encyclopedia of Evolutionary Biology (1st ed.). Academic Press, 208–211.

Knaus, B. J., & Grünwald, N. J. (2017). Vcfr: a package to manipulate and visualize variant call format data in R. Molecular Ecology Resources, 17(1), 44–53.

Kocher, T. D., & Stepien, C. A. (1997). Molecular Systematics of Fishes (C. A. Stepien & T. D. Kocher (eds.); 1st ed.). Academic Press, 1–11.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics.

Lawson, D. J., van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. Nature Communications, 9(1).

Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., Durbin, R., Edwards, S. V, Forest, E., Thomas, M., Gilbert, P., Goldstein, M. M., Grigoriev, I. V, Hackett, K. J., Haussler, D., Jarvis, E. D., Johnson, W. E., Patrinos, A., Richards, S., …, & Zhang, G. (2018). Earth BioGenome Project: Sequencing life for the future of life. Proceedings of the National Academy of Sciences, 115(17), 4325–4333.

Lewontin, R. C., & Hubby, J. L. (1966). A molecular approach to the study of genic heterozygosity in natural populations. 2. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics, 595–609.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics, 27(21), 2987–2993.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25(14), 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078–2079.

Lou, R. N., Jacobs, A., Wilder, A. P., & Therkildsen, N. O. (2021). A beginner's guide to low-coverage whole genome sequencing for population genomics. Molecular Ecology, 30(23), 5966–5993.

Madsen, M. L., Nelson, R. J., Fevolden, S. E., Christiansen, J. S., & Præbel, K. (2016). Population genetic analysis of Euro-Arctic polar cod *Boreogadus saida* suggests fjord and oceanic structuring. Polar Biology, 39(6), 969–980.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal.

Mecklenburg, C. W., Lynghammar, A., Johannesen, E., Byrkjedal., I., Christiansen, J. S., Dolgov, A. V., Karamushko, O. V., Mecklenburg, T. A., Møller, P. R., Steinke, D., & Wienerroither, R. M. (2018). Marine Fishes of the Arctic region. Volume 1. Conservation of Arctic Flora and Fauna.

Møller, P. R., Jordan, A. D., Gravlund, P., & Steffensen, J. F. (2002). Phylogenetic position of the cryopelagic codfish genus *Arctogadus Drjagin*, 1932 based on partial mitochondrial cytochrome b sequences. Polar Biology, 25(5), 342–349.

Nelson, R. J., Bouchard, C., Madsen, M., Praebel, K., Rondeau, E., von Schalburg, K., Leong, J. S., Jantzen, S., Sandwith, Z., Puckett, S., Messmer, A., Fevolden, S. E., & Koop, B. F. (2013). Microsatellite loci for genetic analysis of the Arctic gadids *Boreogadus saida* and *Arctogadus glacialis*. Conservation Genetics Resources, 5(2), 445–448.

Nielsen, J. G., & Jensen, J. M. (1967). Revision of the Arctic cod genus *Arctogadus* (Pisces: Gadidae). Meddelelser om Grønland 184, 1–26.

Nielsen, E. E., Hemmer-Hansen, J., Larsen, P. F., & Bekkevold, D. (2009). Population genomics of marine fishes: Identifying adaptive variation in space and time. Molecular Ecology, 18(15), 3128–3150.

Nielsen, E. E., Nielsen, P. H., Meldrup, D., & Hansen, M. M. (2004). Genetic population structure of turbot (*Scophthalmus maximus*) supports the presence of multiple hybrid zones for marine fishes in the transition zone between the Baltic Sea and the North Sea. Molecular Ecology, 13(3), 585–595.

Olsen, I. L., Arne Rydningen, T., Forwick, M., Sverre Laberg, J., & Husum, K. (2020). Last glacial ice sheet dynamics offshore NE Greenland - A case study from Store Koldewey Trough. Cryosphere, 14(12), 4475–4494.

Olsen, I. L., Laberg, J. S., Forwick, M., Rydningen, T. A., & Husum, K. (2022). Late Weichselian and Holocene behavior of the Greenland Ice Sheet in the Kejser Franz Josef Fjord system, NE Greenland. Quaternary Science Reviews, 284.

Pálsson, S., Källman, T., Paulsen, J., & Árnason, E. (2009). An assessment of mitochondrial variation in Arctic gadoids. Polar Biology, 32(3), 471–479.

Pampoulie, C., Daníelsdóttir, A. K., Storr-Paulsen, M., Hovgård, H., Hjörleifsson, E., & Steinarsson, B. A. (2011). Neutral and nonneutral genetic markers revealed the presence of inshore and offshore stock components of Atlantic cod in Greenland waters. Transactions of the American Fisheries Society, 140(2), 307–319.

Pembleton, L. W., Cogan, N. O. I., & Forster, J. W. (2013). StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Molecular Ecology Resources, 13(5), 946–952.

Pettitt-Wade, H., Loseto, L. L., Majewski, A., & Hussey, N. E. (2021). Cod movement ecology in a warming world: Circumpolar Arctic gadids. Fish and Fisheries, 22(3), 562–591.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81(3), 559–575.

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. Molecular Ecology, 24(17), 4348–4370.

Ribeiro, S., Sejr, M. K., Limoges, A., Heikkilä, M., Andersen, T. J., Tallberg, P., Weckström, K., Husum, K., Forwick, M., Dalsgaard, T., Massé, G., Seidenkrantz, M. S., & Rysgaard, S. (2017). Sea ice and primary production proxies in surface sediments from a High Arctic Greenland fjord: Spatial distribution and implications for palaeoenvironmental studies. Ambio, 46, 106–118.

Saraswathy, N., & Ramalingam, P. (2011). Concepts and Techniques in Genomics and Proteomics (1st ed.). Woodhead Publishing.

Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals-mining genome-wide polymorphism data without big funding. Nature Reviews Genetics, 15(11), 749–763.

Siwertsson, A., Knudsen, R., Adams, C. E., Præbel, K., & Amundsen, P. A. (2013). Parallel and non-parallel morphological divergence among foraging specialists in European whitefish (*Coregonus lavaretus*). Ecology and Evolution, 3(6), 1590–1602.

Slatkin, M. (2008). Linkage disequilibrium - Understanding the evolutionary past and mapping the medical future. Nature Reviews Genetics (Vol. 9, Issue 6, 477–485).

Süfke, L., Piepenburg, D., & Von Dorrien, C. F. (1998). Body size, sex ratio and diet composition of *Arctogadus glacialis* (Peters, 1874) (Pisces: Gadidae) in the Northeast Water Polynya (Greenland). Polar Biology, 20(5), 357–363.

Walters, V. (1961). Winter Abundance of *Arctogadus glacialis* in the Polar Basin. Copeia.

Wassmann, P., Duarte, C. M., Agustí, S., & Sejr, M. K. (2011). Footprints of climate change in the Arctic marine ecosystem. Global Change Biology, 17(2), 1235–1249.

Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of F_{st}. American Naturalist, 186, 24–36.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … & Yutani, H. (2019). Welcome to the Tidyverse. Journal of Open Source Software, 4(43), 1686.

Wright, S. (1951). The genetical structure of populations. Annals of Eugenics, 15(4), 323–354.

Zoller, K., Laberg, J. S., Rydningen, T. A., Husum, K., & Forwick, M. (2023). A high Arctic inner shelf-fjord system from the Last Glacial Maximum to the present: Bessel Fjord and Southwest Dove Bugt, northeastern Greenland. Climate of the Past, 19(7), 1321–1343.

Appendices

*Appendix 1. Purification of Total DNA from Animal Tissues (Spin-Column Protocol)***.** *This protocol was provided by Qiagen from their DNeasy Blood and Tissue Handbook.*

1. Cut up to 25 mg tissue (up to 10 mg spleen) into small pieces, and place in a 1.5 ml microcentrifuge tube. For rodent tails, place one (rat) or two (mouse) 0.4–0.6 cm lengths of tail into a 1.5 ml microcentrifuge tube. Add 180 µl Buffer ATL. Earmark the animal appropriately.

Ensure that the correct amount of starting material is used (see "Starting amounts of samples", page 15). For tissues such as spleen with a very high number of cells for a given mass of tissue, no more than 10 mg starting material should be used.

We strongly recommend to cut the tissue into small pieces to enable more efficient lysis. If desired, lysis time can be reduced by grinding the sample in liquid nitrogen* before addition of Buffer ATL and proteinase K. Alternatively, tissue samples can be effectively disrupted before proteinase K digestion using a rotor-stator homogenizer, such as the QIAGEN TissueRuptor, or a bead mill, such as the QIAGEN TissueLyser. A supplementary protocol for simultaneous disruption of up to 48 tissue samples using the TissueLyser can be obtained by contacting QIAGEN Technical Services.

For rodent tails, a maximum of 1.2 cm (mouse) or 0.6 cm (rat) tail should be used. When purifying DNA from the tail of an adult mouse or rat, it is recommended to use only 0.4–0.6 cm.

2. Add 20 µl proteinase K. Mix thoroughly by vortexing and incubate at 56°C until the tissue is completely lysed. Vortex occasionally during incubation to disperse the sample, or place in a thermomixer, shaking water bath, or on a rocking platform.

Lysis time varies depending on the type of tissue processed. Lysis is usually complete in 1–3 h or, for rodent tails, 6–8 h. If it is more convenient, samples can be lysed overnight; this will not affect them adversely.

After incubation the lysate may appear viscous but should not be gelatinous as it may clog the DNeasy Mini spin column.

Optional: If RNA-free genomic DNA is required, add 4 µl RNase A (100 mg/ml), mix by vortexing, and incubate for 2 min at room temperature before continuing with step 3.

Transcriptionally active tissues such as liver and kidney contain high levels of RNA, which will copurify with genomic DNA. For tissues that contain low levels of RNA, such as rodent tails, or if residual RNA is not a concern, RNase A digestion is not necessary.

3. Vortex for 15 s. Add 200 µl Buffer AL to the sample, and mix thoroughly by vortexing. Then add 200 µl ethanol (96–100%) and mix again thoroughly by vortexing.

It is essential that the sample, Buffer AL, and ethanol are mixed immediately and thoroughly by vortexing or pipetting to yield a homogeneous solution. Buffer AL and ethanol can be premixed and added together in one step to save time when processing multiple samples.

A white precipitate may form on addition of Buffer AL and ethanol. This precipitate does not interfere with the DNeasy procedure. Some tissue types (e.g., spleen, lung) may form a gelatinous lysate after addition of Buffer AL and ethanol. In this case, vigorously shaking or vortexing the preparation is recommended.

4. Pipet the mixture from step 3 (including any precipitate) into the DNeasy Mini spin column placed in a 2 ml collection tube (provided). Centrifuge at 6000 x g (8000 rpm) for 1 min. Discard flow-through and collection tube.*
5. Place the DNeasy Mini spin column in a new 2 ml collection tube (provided), add 500 µl Buffer AW1, and centrifuge for 1 min at 6000 x g (8000 rpm). Discard flow-through and collection tube.*
6. Place the DNeasy Mini spin column in a new 2 ml collection tube (provided), add 500 µl Buffer AW2, and centrifuge for 3 min at 20,000 x g (14,000 rpm) to dry the DNeasy membrane. Discard flow-through and collection tube.

It is important to dry the membrane of the DNeasy Mini spin column, since residual ethanol may interfere with subsequent reactions. This centrifugation step ensures that no residual ethanol will be carried over during the following elution.

Following the centrifugation step, remove the DNeasy Mini spin column carefully so that the column does not come into contact with the flow-through, since this will result in carryover of ethanol. If carryover of ethanol occurs, empty the collection tube, then reuse it in another centrifugation for 1 min at 20,000 x g (14,000 rpm).

7. Place the DNeasy Mini spin column in a clean 1.5 ml or 2 ml microcentrifuge tube (not provided), and pipet 200 µl Buffer AE directly onto the DNeasy membrane. Incubate at room temperature for 1 min, and then centrifuge for 1 min at 6000 x g (8000 rpm) to elute.

Elution with 100 µl (instead of 200 µl) increases the final DNA concentration in the eluate, but also decreases the overall DNA yield.

8. Recommended: For maximum DNA yield, repeat elution once as described in step 7.

This step leads to increased overall DNA yield.

A new microcentrifuge tube can be used for the second elution step to prevent dilution of the first eluate. Alternatively, to combine the eluates, the microcentrifuge tube from step 7 can be reused for the second elution step. Note: Do not elute more than 200 µl into a 1.5 ml micro-centrifuge tube because the DNeasy Mini spin column will come into contact with the eluate.

Appendix 2. Purification of Total DNA from Animal Tissues (DNeasy 96 Protocol). This protocol was provided by Qiagen from their DNeasy Blood and Tissue Handbook.

1. Cut up to 20 mg tissue (up to 10 mg spleen) into small pieces. For rodent tails, place one (rat) or two (mouse) 0.4–0.6 cm lengths of tail into a collection microtube. Earmark the animal appropriately. Use a 96-Well-Plate Register (provided) to identify the position of each sample.

Ensure that the correct amount of starting material is used (see "Starting amounts of samples", page 15). For tissues such as spleen with a very high number of cells for a given mass of tissue, no more than 10 mg starting material should be used.

We strongly recommend to cut the tissue into small pieces to enable more efficient lysis. If desired, lysis time can be reduced by disrupting the sample using a bead mill, such as the QIAGEN TissueLyser (see page 56 for ordering information), before addition of Buffer ATL and proteinase K. A supplementary protocol for simultaneous disruption of up to 48 tissue samples using the TissueLyser can be obtained by contacting QIAGEN Technical Services.

For rodent tails, a maximum of 1.2 cm (mouse) or 0.6 cm (rat) tail should be used. When purifying DNA from the tail of an adult mouse or rat, it is recommended to use only 0.4–0.6 cm.

Store the samples at -20°C until a suitable number has been collected (up to 192 samples). Samples can be stored at -20°C for several weeks to months without any reduction in DNA yield. DNA yields will be approximately 10–30 µg, depending on the type, length, age, and species of sample used (see "Expected yields", page 22).

Keep the clear covers from the collection microtube racks for use in step 3.

2. Prepare a proteinase K-Buffer ATL working solution containing 20 µl proteinase K stock solution and 180 µl Buffer ATL per sample, and mix by vortexing. For one set of 96 samples, use 2 ml proteinase K stock solution and 18 ml Buffer ATL. Immediately pipet 200 µl working solution into each collection microtube containing the tail sections or tissue samples. Seal the microtubes properly using the caps provided.

Note: Check Buffer ATL for precipitate. If necessary, dissolve the precipitate by incubation at 56°C for 5 min before preparing the working solution.

IMPORTANT: After preparation, the proteinase K-Buffer ATL working solution should be dispensed immediately into the collection microtubes containing the tail or tissue samples.

Incubation of the working solution in the absence of substrate for over 30 min reduces lysis efficiency and DNA purity.

3. Ensure that the microtubes are properly sealed to avoid leakage during shaking. Place a clear cover (saved from step 1) over each rack of collection microtubes, and mix by inverting the rack of collection microtubes. To collect any solution from the caps, centrifuge the collection microtubes. Allow the centrifuge to reach 3000 rpm, and then stop the centrifuge. It is essential that the samples are completely sub- merged in the proteinase K-Buffer ATL working solution after centrifugation.

If the proteinase K-Buffer ATL working solution does not completely cover the sample, increase the volume of the solution to 300 µl per sample (additional reagents are available separately; see page 56 for ordering information). Do not increase volumes above 300 µl as this will exceed the capacity of the collection microtubes in subsequent steps.

Keep the clear covers from the collection microtube racks for use in step 5.

4. Incubate at 56°C overnight or until the samples are completely lysed. Place a weight on top of the caps during the incubation. Mix occasionally during incubation to disperse the sample, or place on a rocking platform.

Lysis time varies depending on the type, age, and amount of tail or tissue being processed. Lysis is usually complete in 1–3 h or, for rodent tails, 6–8 h, but optimal results will be achieved after overnight lysis.

After incubation the lysate may appear viscous, but should not be gelatinous as it may clog the DNeasy 96 membrane. If the lysate appears very gelatinous, see the “Troubleshooting Guide” for recommendations.

Note: Do not use a rotary- or vertical-type shaker as continuous rotation may release the caps. If incubation is performed in a water bath make sure that the collection microtubes are not fully submerged and that any remaining water is removed prior to centrifugation in step 5.

5. Ensure that the microtubes are properly sealed to avoid leakage during shaking. Place a clear cover over each rack of collection microtubes and shake the racks vigorously up and down for 15 s. To collect any solution from the caps, centrifuge the collection microtubes. Allow the centrifuge to reach 3000 rpm, and then stop the centrifuge.

IMPORTANT: The rack of collection microtubes must be vigorously shaken up and down with both hands to obtain a homogeneous lysate. Inverting the rack of collection microtubes is

not sufficient for mixing. The genomic DNA will not be sheared by vigorous shaking. Keep the clear covers from the collection microtube racks for use in step 7.

Ensure that lysis is complete before proceeding to step 6. The lysate should be homogeneous following the vigorous shaking. To check this, slowly invert the rack of collection microtubes (making sure that the caps are tightly closed) and look for a gelatinous mass. If a gelatinous mass is visible, lysis needs to be extended by adding another 100 µl Buffer ATL and 15 µl proteinase K, and incubating for a further 3 h. It is very important to ensure that samples are completely lysed to achieve optimal yields and to avoid clogging of individual wells of the DNeasy 96 plate.

Optional: If RNA-free genomic DNA is required, add 4 µl RNase A (100 mg/ml). Close the collection microtubes with fresh caps, mix by shaking vigorously, and incubate for 5 min at room temperature. To collect any solution from the caps, centrifuge the collection microtubes. Allow the centrifuge to reach 3000 rpm, and then stop the centrifuge. Remove the caps, and continue with step 6.

Transcriptionally active tissues such as liver and kidney contain high levels of RNA, which will copurify with genomic DNA. For tissues that contain low levels of RNA, such as rodent tails, or if residual RNA is not a concern, RNase A digestion is usually not necessary.

6. Carefully remove the caps. Add 410 µl premixed Buffer AL-ethanol to each sample.

Note: Ensure that ethanol has been added to Buffer AL prior to use.

Note: A white precipitate may form upon addition of Buffer AL-ethanol to the lysate. It is important to apply all of the lysate, including the precipitate, to the DNeasy 96 plate in step 9. This precipitate does not interfere with the DNeasy procedure or with any subsequent application.

If the volumes of Buffer ATL and proteinase K were increased in steps 3 or 5, increase the volume of Buffer AL and ethanol accordingly. For example, 300 µl proteinase K-Buffer ATL working solution will require 615 µl Buffer AL-ethanol.

7. Ensure that the microtubes are properly sealed to avoid leakage during shaking. Place a clear cover over each rack of collection microtubes and shake the racks vigorously up and down for 15 s. To collect any solution from the caps, centrifuge the collection microtubes. Allow the centrifuge to reach 3000 rpm, and then stop the centrifuge.

Do not prolong this step.

IMPORTANT: The rack of collection microtubes must be vigorously shaken up and down with both hands to obtain a homogeneous lysate. Inverting the rack of collection microtubes is not sufficient for mixing. The genomic DNA will not be sheared by vigorous shaking. The lysate and Buffer AL-ethanol should be mixed immediately and thoroughly to yield a homogeneous solution.

8. Place two DNeasy 96 plates on top of S-Blocks (provided). Mark the DNeasy 96 plates for later sample identification.
9. Remove and discard the caps from the collection microtubes. Carefully transfer the lysate (maximum 900 µl) of each sample from step 7 to each well of the DNeasy 96 plates.

   Take care not to wet the rims of the wells to avoid aerosols during centrifugation. Do not transfer more than 900 µl per well.

Note: Lowering pipet tips to the bottoms of the wells may cause sample overflow and cross-contamination. Therefore, remove one set of caps at a time, and begin drawing up the samples as soon as the pipet tips contact the liquid. Repeat until all the samples have been transferred to the DNeasy 96 plates.

Note: If the volume of proteinase K-Buffer ATL working solution was increased in steps 3 or 5, transfer no more than 900 µl of the supernatant from step 7 to the DNeasy 96 plate. Larger amounts will exceed the volume capacity of the individual wells. Discard any remaining supernatant from step 7 as this will not contribute significantly to the total DNA yield.

10. Seal each DNeasy 96 plate with an AirPore Tape Sheet (provided). Centrifuge for 10 min at 6000 rpm.

AirPore Tape prevents cross-contamination between samples during centrifugation.

After centrifugation, check that all of the lysate has passed through the membrane in each well of the DNeasy 96 plates. If lysate remains in any of the wells, centrifuge for a further 10 min.

11. Remove the tape. Carefully add 500 µl Buffer AW1 to each sample. Note: Ensure that ethanol has been added to Buffer AW1 prior to use.

It is not necessary to increase the volume of Buffer AW1 if the volume of proteinase K-Buffer ATL working solution was increased in steps 3 or 5.

12. Seal each DNeasy 96 plate with a new AirPore Tape Sheet (provided). Centrifuge for 5 min at 6000 rpm.
13. Remove the tape. Carefully add 500 µl Buffer AW2 to each sample. Note: Ensure that ethanol has been added to Buffer AW2 prior to use.

It is not necessary to increase the volume of Buffer AW2 if the volume of proteinase K-Buffer ATL working solution was increased in steps 3 or 5.

14. Centrifuge for 15 min at 6000 rpm.

Do not seal the plate with AirPore Tape. The heat generated during centrifugation ensures evaporation of residual ethanol in the sample (from Buffer AW2) that might otherwise inhibit downstream reactions.

15. Place each DNeasy 96 plate in the correct orientation on a new rack of Elution Microtubes RS (provided).
16. To elute the DNA, add 200 µl Buffer AE to each sample, and seal the DNeasy 96 plates with new AirPore Tape Sheets (provided). Incubate for 1 min at room tem- perature (15–25°C). Centrifuge for 2 min at 6000 rpm.

200 µl Buffer AE is sufficient to elute up to 75% of the DNA from each well of the DNeasy 96 plate.

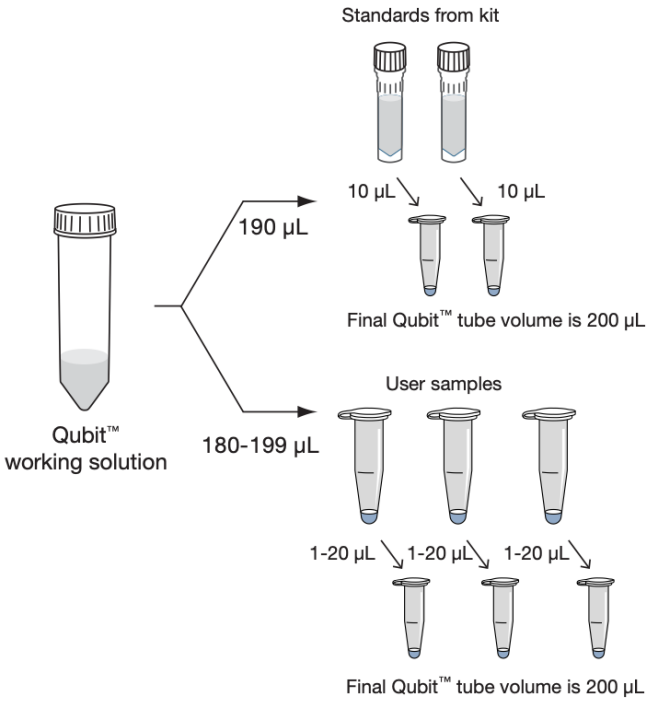
Elution with volumes less than 200 µl significantly increases the final DNA concentration of the eluate but may reduce overall DNA yield. For samples containing less than 1 µg DNA, elution in 50 µl Buffer AE is recommended.

17. Recommended: For maximum DNA yield, repeat step 16 with another 200 µl Buffer AE.

A second elution with 200 µl Buffer AE will increase the total DNA yield by up to 25%. However, due to the increased volume, the DNA concentration is reduced. If a higher DNA concentration is desired, the second elution step can be performed using the 200 µl eluate from the first elution. This will increase the yield by up to 15%.

Use new caps (provided) to seal the Elution Microtubes RS for storage.

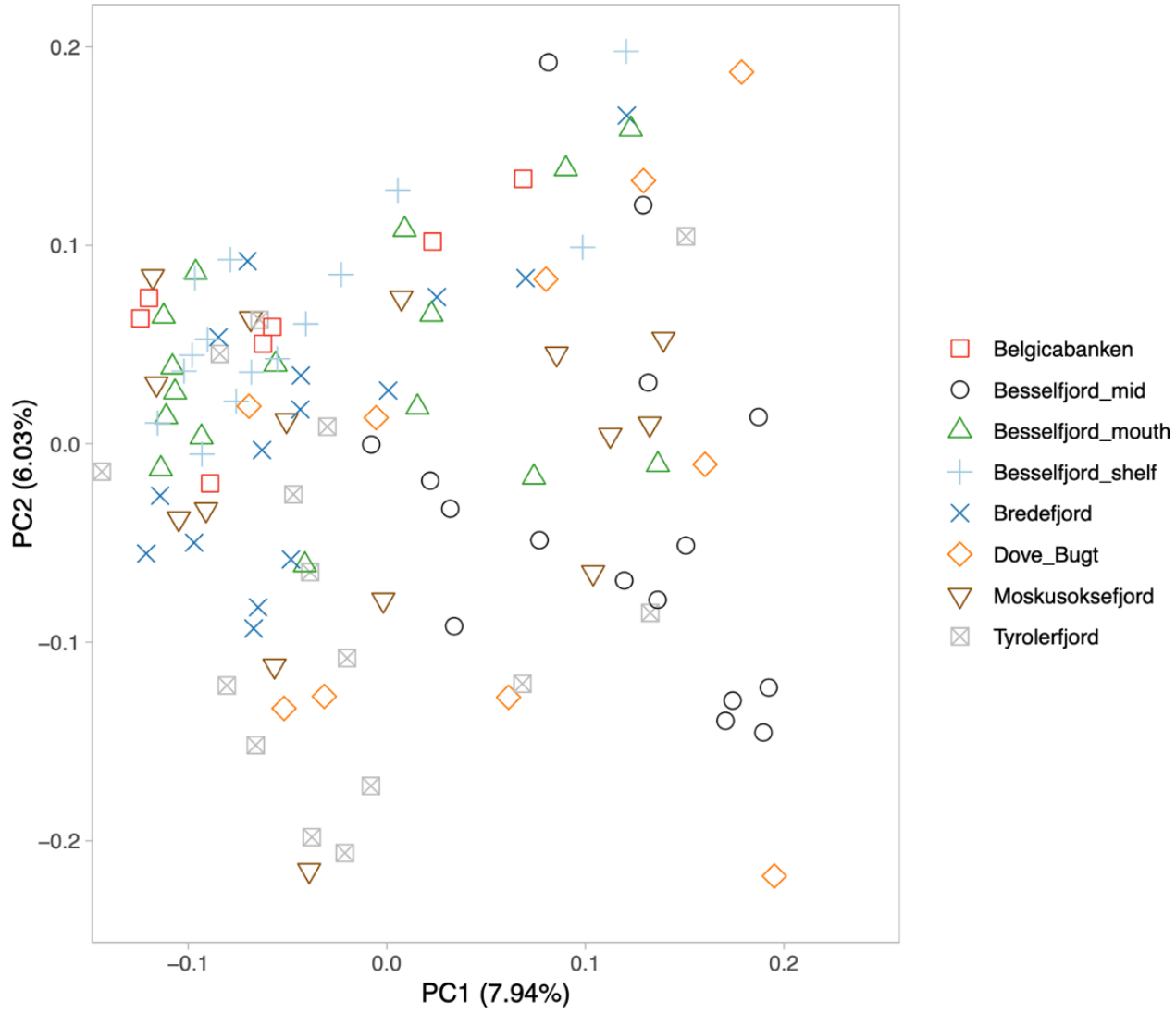
Appendix 3. Manual for Qubit 1X ds DNA BR Assay with a Qubit 4 Fluorometer. The manual was provided by Thermo Fisher Scientific.



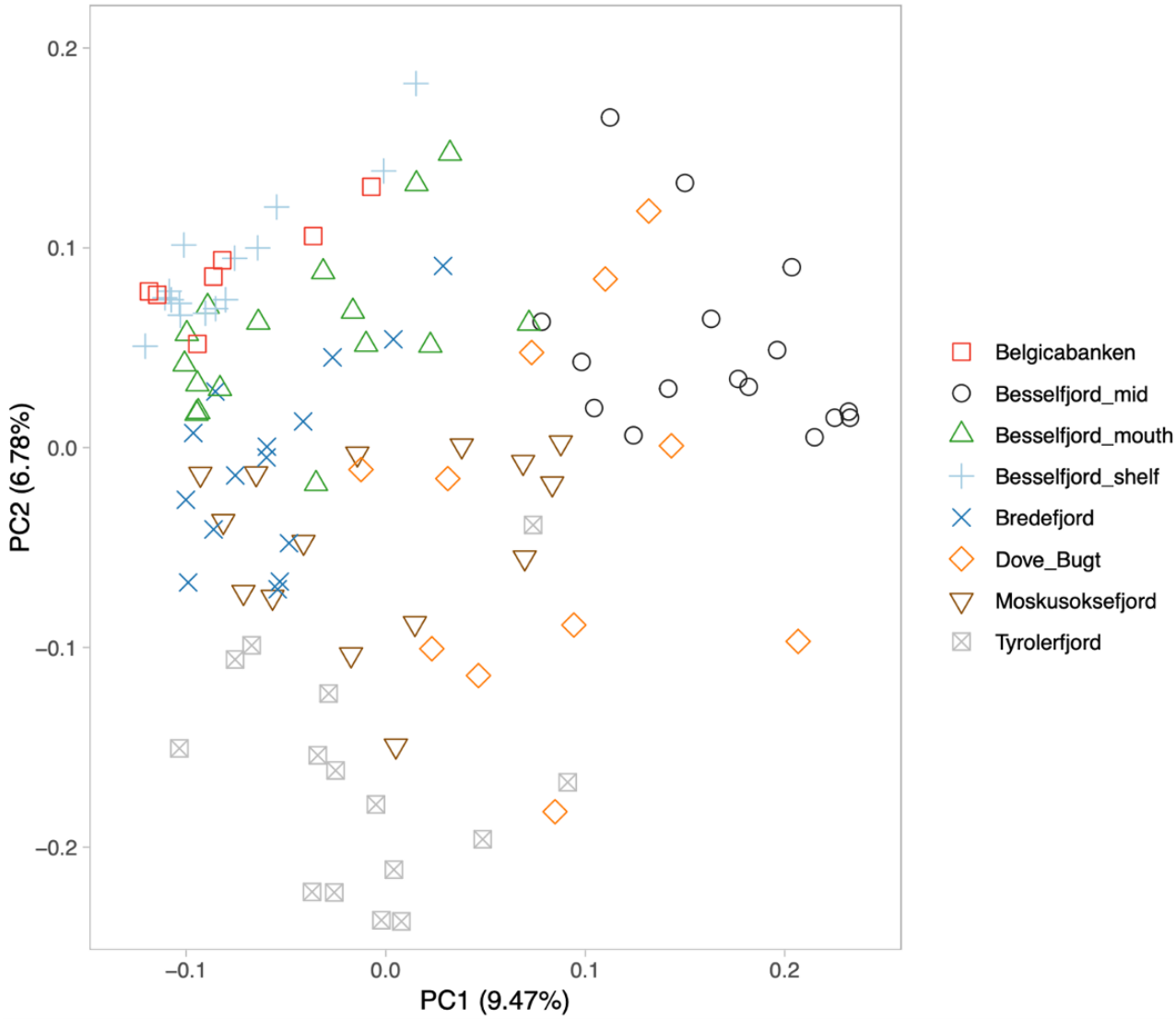
Appendix 4. Sequencing requirements. Genomic DNA (upper row) requirements were used as guideline. The requirements were provided by NOVOGENE, Co, Ltd.

Library Type	Sample Type	Amount (Qubit®)	Volume	Concentration	Purity (NanoDrop™/Agarose Gel)
Plant and animal whole genome library (⩽ 500 bp)	Genomic DNA	⩾ 200 ng	⩾ 20 µL	⩾ 10 ng/µL	OD260/280 = 1.8~2.0; no degradation, no contamination
	Genomic DNA (PCR-free non-350bp)	⩾ 5 µg	⩾ 20 µL	⩾ 30 ng/µL	
	Genomic DNA (PCR-free 350bp)	⩾ 1.2 µg	⩾ 20 µL	⩾ 20 ng/µL	

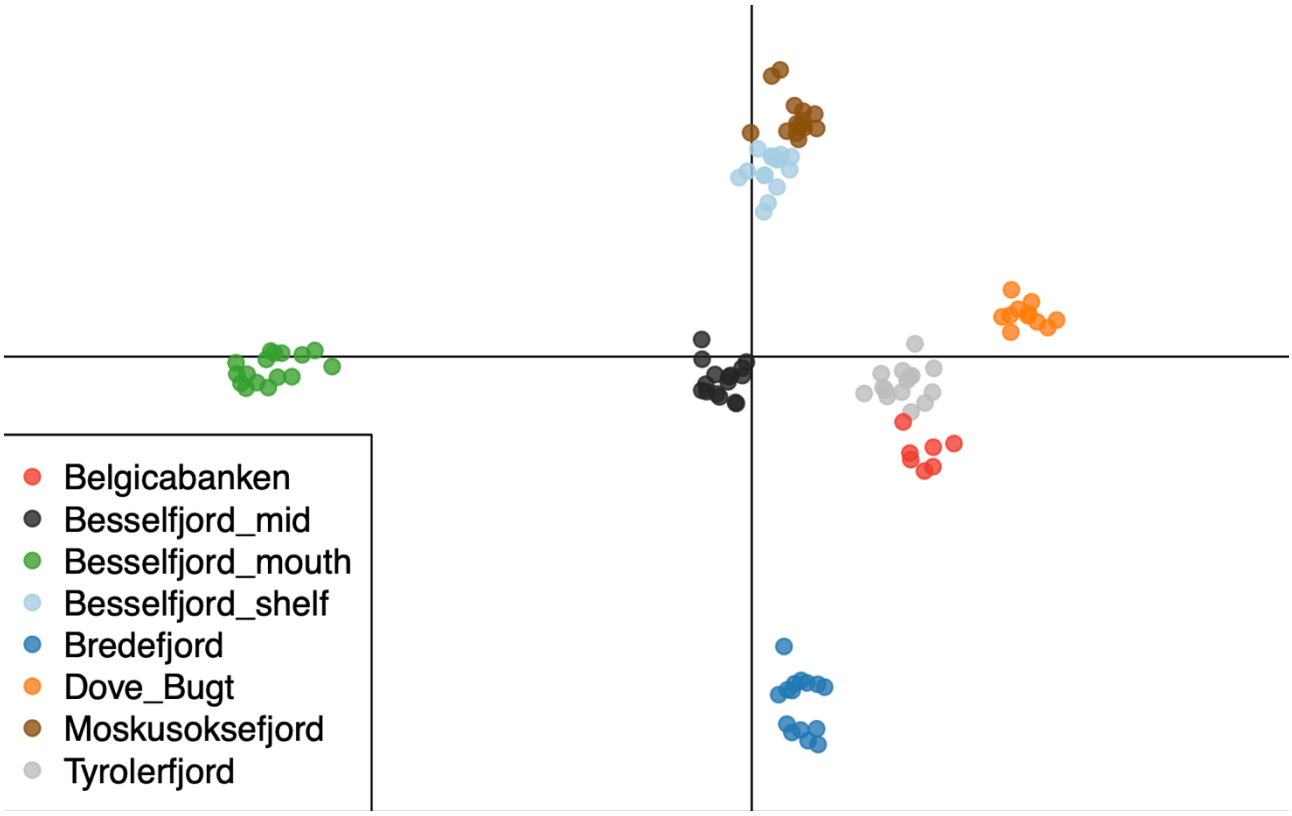
Appendix 5. Additional PCA plot 1. The plot is based on 542k SNPs. PC1 explains 7.94% of the allele frequency variation, whereas PC2 explains 6.03%.



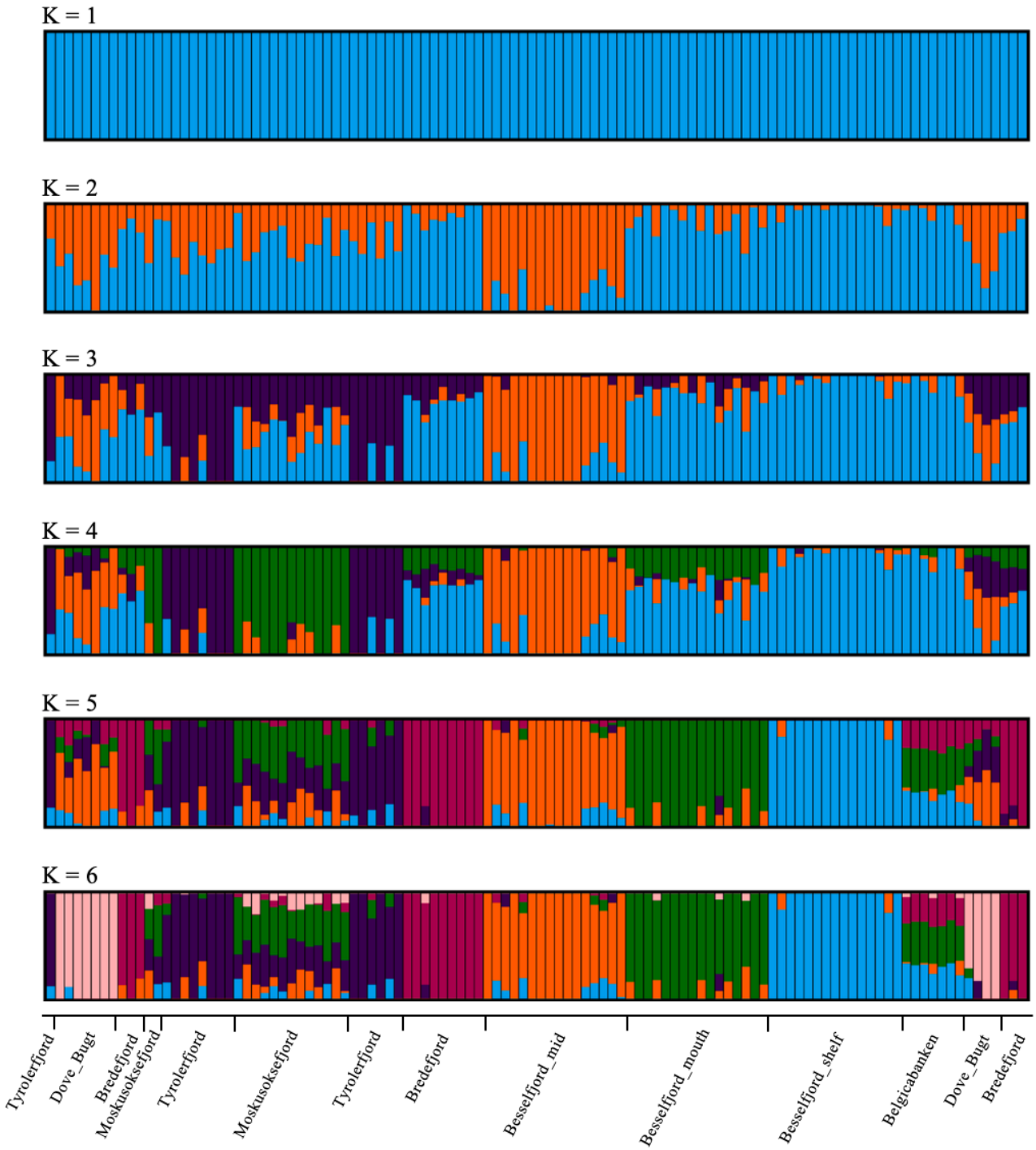
Appendix 6. Additional PCA plot 2. The plot is based on F_{st} values above 0. PC1 explains 9.47% of the variation in allele frequencies, whereas PC2 explains 6.78%.

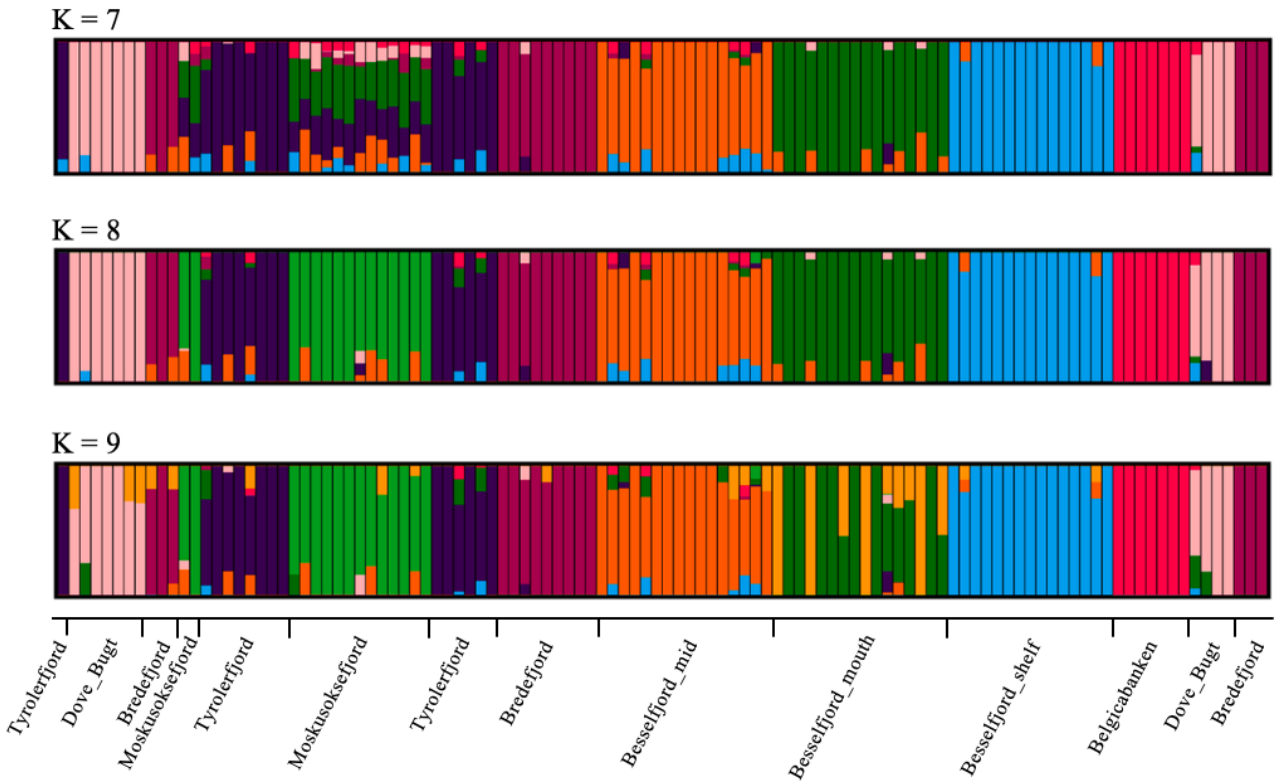


Appendix 7. DAPC scatter plot. The analysis is based on 10 PCs, which explains 64.69% of the total allele frequency variation in the data set.



Appendix 8. Admixture of all K values + 1 (K = 1–9). Be aware that the populations are not sorted and that the colors are different from Fig. 9. The figure continues on the next page.




K = 7
K = 8
K = 9
Tyrolerfjord
Dove_Bugt
Bredefjord
Moskusoksefjord
Tyrolerfjord
Moskusoksefjord
Tyrolerfjord
Bredefjord
Besselfjord_mid
Besselfjord_mouth
Besselfjord_shelf
Belgicabanken
Dove_Bugt
Bredefjord

Web links for workflow

Web link 1. Manual for Qubit 1X ds DNA BR Assay with a Qubit 4 Fluorometer (page 4 in the manual). First opened Nov 30, 2024: https://www.thermofisher.com/document-connect/document-connect.html?url=https://assets.thermofisher.com/TFS-Assets%2FLSG%2Fmanuals%2FMAN0019617_Qubit_1X_dsDNA_BR_Assay_UG.pdf

Web link 2. Guideline for Assessment of Nucleic Acid Purity by applying NanoDrop spectrophotometer. The guideline was provided by Thermo Fisher Scientific. First opened Nov 30, 2024: https://assets.thermofisher.com/TFS-Assets/CAD/Product-Bulletins/TN52646-E-0215M-NucleicAcid.pdf

Web link 3. R Core Team (2024). A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. First opened Apr 2, 2024: https://www.R-project.org/

Web link 4. Figtree. First opened Apr 19, 2024: http://tree.bio.ed.ac.uk/software/figtree/

Web link 5. Cartographic map of Bredefjord from Sep 4, 2022. First opened Apr 17, 2024: https://www.c-map.com/