UiT The Arctic University of Norway

Faculty of Biosciences, Fisheries and Economics, Department of Arctic and Marine Biology

Transcriptomic basis for differentiation of fjord and offshore *Boreogadus saida* (Polar cod) populations

Griffin Hill

BIO-3950 Master's Thesis in Biology, May 2021

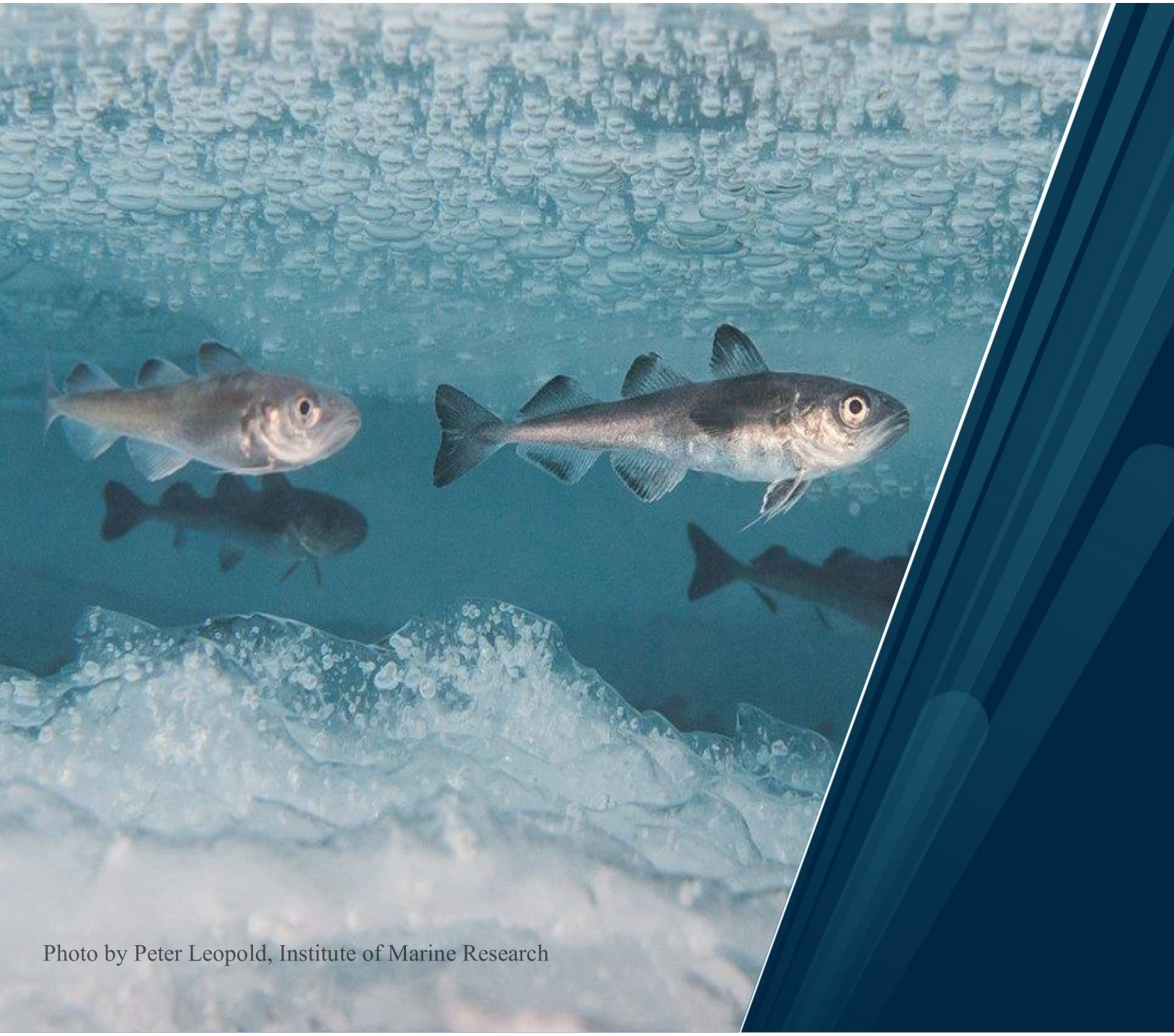


Photo by Peter Leopold, Institute of Marine Research

Abstract

Boreogadus saida (Polar cod) is a critical ecological linkage between planktonic primary production and megafaunal top consumers in the Arctic. Despite its critical ecology, the population structure, especially the relationship between fjord and oceanic groups inhabiting environments with different physical characteristics, is poorly described. Though the subject of several genetic studies, an underlying biological basis for this difference is yet to be identified. Total RNA was extracted from liver tissue of 144 Polar cod representing 3 pairs of neighboring fjord and offshore populations, comprising a North-South gradient of Eastern Greenland and Svalbard. 3' sequencing of the corresponding cDNA on two Illumina HiSeq 4000 lanes yielded approximately 14 million reads per sample upon which differential gene expression analysis was performed. After alignment and read quantification using the Atlantic Cod (*Gadus morhua*) genome, approximately 2,000 significantly differentially expressed (up or down regulated) genes per fjord-offshore pairing were identified. Gene ontology profiling of the differentially expressed genes by biological process revealed that distance between fjord and offshore populations is a strong predictor of gene expression between these populations. The differentially expressed biological pathways identified provide insight on the origin and structure of fjord and oceanic populations of Polar cod including evidence of gradual differentiation within the Greenland Sea based primarily on distance from the coast. This examination of a key Arctic species through the lens of a novel transcriptomic approach provides new context to previous genetic investigations as well as a unique foundational dataset for subsequent analyses of these populations.

Table of Contents

List of Figures

List of Tables

Foreword

This thesis is the culmination of the two year Master of Biology program from the department of Arctic and Marine Biology at UiT The Arctic University of Norway. This research was carried out with the support of the Research Group for Genetics (RGG) at the Norwegian College of Fishery Science from May 2020 to May 2021. For this project I contributed to the development of the project's narrative and overall design, carried out the extraction of total RNA from preserved liver tissue, cDNA library preparation for sequencing, post-processing of sequencing results, statistical analyses, and writing of this thesis. My advisors and other RGG group members provided support, both technical and emotional, throughout this undertaking. Among their direct contributions:

Kim Praebel, supervisor, conceived of the project, collected samples as part of the TUNU program, provided feedback on the project narrative and final manuscript.

Filipe Figueiredo, supervisor, provided instructions for RNA extraction and cDNA preparation, shared previous bioinformatic approaches to similar data.

Shripathi Bhat, bioinformatics extraordinaire, collected samples as part of the TUNU program, provided feedback on experimental design and assisted with bioinformatic processing of sequencing data.

Julie Bitz-Thorsen, lab manager, supervised lab activities and provided valuable input on the optimization of the cDNA library preparation method.

INTRODUCTION

Arctic marine ecosystems are changing at up to twice the speed of their lower latitude and terrestrial counterparts (Moritz et al., 2002, Figure 1). Anthropogenic climate change and its associated rapid alteration of Arctic marine habitat could lead to irreversible change in an ecosystem that remains poorly understood (Hoegh-Guldberg & Bruno, 2010). In such a dynamic system, it is critical to understand the changing relationship between key species and their environment, including economically important fishes. *Boreogadus saida* (Polar cod/ Polartorsk Lepechin 1774, hereafter referred to as *B. saida*) is one such species critically impacted by warming in the Arctic and of great consequence to the stability of Arctic marine ecosystems and the fisheries they support. While only subject to a modest direct fishery, the trophic niche inhabited by *B. saida* provides a critical linkage between primary producers and some of the largest biomass consumers in the Arctic (Hop & Gjøsæter, 2013). Despite its important role, the structure of the largest *B. saida* populations in the Greenland Sea is relatively unknown and has only recently begun to be revealed via genetic (Madsen et al., 2016) and now transcriptomic studies in this thesis. There is a pressing need to bring together growing knowledge of *B. saida* population structure and trophic interactions in considering its management in the context of a rapidly changing Arctic climate.

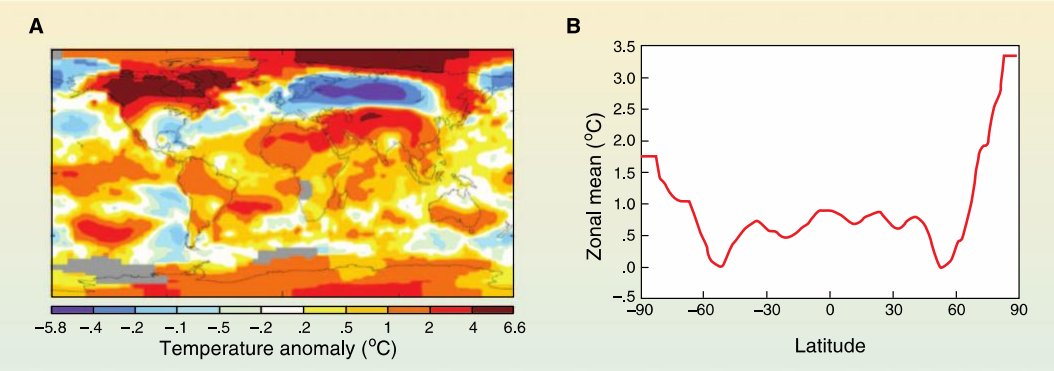


Figure 1: Recent changes in ocean temperature, acidity, and carbonate ion concentration. (A) Surface temperature anomaly for January 2010 relative to the mean for 1951–1980. (B) The same data presented in (A) as a function of latitude (Hoegh-Guldberg et al. 2010).

Boreogadus saida is a cryptic, poorly understood pan-Arctic forage fish of the *Gadidae* family (Hop & Gjøsæter, 2013). Despite its circumpolar distribution, the monospecific genus *Boreogadus* has been studied most intensively in the Russian/Norwegian waters of the Barents Sea and the North American Bering Strait and Chukchi Seas, leading to regional findings being widely applied to distinct reproductively isolated populations in diverse Arctic habitats (Hedges

et al., 2017). As an important trophic linkage, *B. saida* is a high energy, secondary consumer that is preferentially preyed upon by many marine mammals and seabirds, as well as by Atlantic cod (*Gadus morhua*, referred to by the abbreviated Latin name *G. morhua* from now on) (Bradstreet & Cross, 1982). This species produces antifreeze glycoproteins and relies on sea ice cover as a nursery environment for its buoyant eggs and sympagic larvae who preferentially feed on ice associated pelagic calanoid zooplankton (Graham & Hop, 1995; Mueter et al., 2016). Though thriving in temperatures from -2˚C to +5˚C, *B. saida* is unable to undergo normal embryonic development above this range (Drost et al., 2016). *Boreogadus saida* has also been observed schooling in large numbers in ice covered areas, but its range is not shared by any pelagic forage fish in comparable quantities (Melnikov & Chernova, 2013).

Figure 2: Map shows the maximum distribution of B. saida as observed from point data and includes both common and rare locations. Distribution extent is based upon literature review, museum specimens, and research sampling carried out by the Conservation of Arctic Flora and Fauna (CAFF) organization as part of the Circumpolar Biodiversity Monitoring Program (CBMP) (Hedges et al., 2017).

A unique trophic linkage

B. saida play a critical role in Arctic nutrient cycling, providing a critical trophic linkage between planktonic producers, some of the Arctic's top predators, and commercially important species. In the extremely seasonal production landscape of the light-limited Arctic, energy storage in the form of lipids is a key element in the survival strategy of most Arctic species (Falk-Petersen et al., 1990). On the ice edge, primary production is driven by under-ice diatom algae, which is subsequently consumed and converted to high energy lipid stores by amphipods, and Calanoid copepods (Bradstreet & Cross, 1982). Lipids in this form are largely inaccessible to marine mammals, birds, and large predatory fish that are not adapted to prey upon such small organisms in such an ice sheltered environment. However, *B. saida* prey on these invertebrate groups with great consistency across age class and population (inshore vs. offshore) (Bradstreet & Cross, 1982). In turn, *B. saida* provide a key energy rich prey item for a variety of seabirds and marine mammals such as ringed seals, narwhals, and beluga (Bluhm & Gradinger, 2008, Figure 3). These large predators rely on prey hotspots to meet their annual energy requirements in an environment with only seasonal productivity. In following with this hotspot principle, occurrences of *B. saida* schools have been identified as the driving force behind some marine mammal congregations and migrations such as the narwhal in Crewell Bay (Welch et al., 1993). The observation of *B. saida* swarming under the ice (Melnikov & Chernova, 2013) and at the

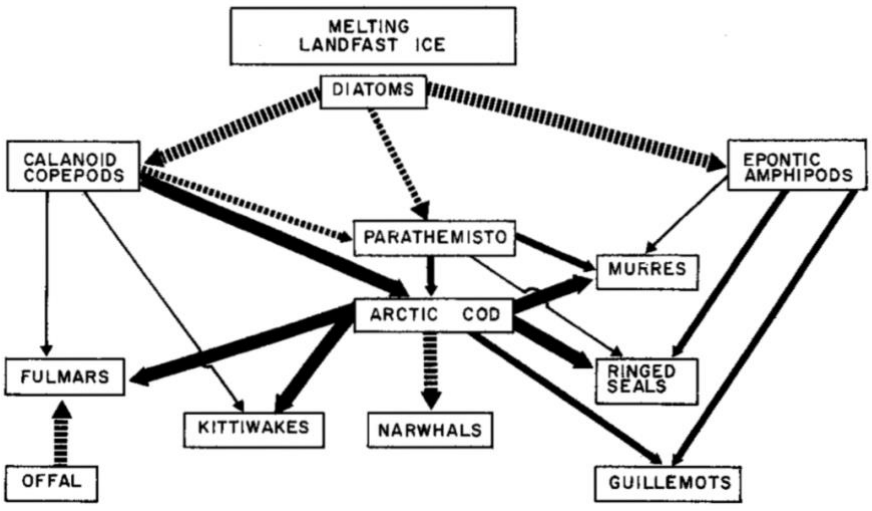


Figure 3: Trophic pathways along offshore ice edges in the Canadian High Arctic. Decreasing widths of lines reflect pathways of major, moderate, and minor importance, respectively. Solid lines are based on dry weight analysis methods. Hatched lines are based on wet weight (narwhal) or occurrence (all other) analyses (Bradstreet and Cross 1982). Subsequent studies have shown beluga and a greater variety of pinnipeds and seabirds also directly rely on B. saida (here referred to as Arctic cod) in this way.

ice edge further establishes this species as a prey item that, being energy rich and seasonally concentrated, meets the requirements to support large mammalian predators with high energy needs. This schooling migration, unique among true Arctic fishes, drives huge seasonal horizontal energy transfer between ecosystems and is the subject of regional specialized predation, likely irreplaceable by stationary boreal species. Maintaining the presence of *B. saida* as a mid-trophic level "wasp-waist" consumer confers a degree of ecological stability to systems that are, even at the best of times, delicate.

Stable isotope studies bring the threat of sea ice loss into focus, showing that ice-algae-derived carbon underpins nearly the entire *B. saida* diet and subsequently large swaths of Arctic marine food webs (Kohlbach et al., 2017). Furthermore, the *B. saida* lifecycle is contingent upon sea ice presence. Without ice, it is likely that their positively buoyant eggs will suffer significantly decreased performance as they are subject to UV damage, predation, and lack access to suitable under ice habitat upon hatching. Despite being of only minor commercial interest (currently only harvested by Russia), *B. saida* has an outsized ecosystem impact as the primary vertical transport of lipids between ice associated primary producers and some of the largest keystone predators in the Arctic. As commercially important species such as Atlantic cod and haddock expand poleward, increased predation upon, as well as bycatch of, *B. saida* is inevitable (Renaud et al., 2018). This further situates an already vulnerable fish at the center of an increasingly fragile food web.

Possible impacts of local extinction

It is hard to predict the impacts of localized *B. saida* extinction/replacement as the Arctic warms, but some outcomes are more likely than others. Natural predation upon capelin and herring would likely increase as these species are already present in the diet of many marine mammals who predate on *B. saida*. Coupled with fishing pressure and greater variation in sea surface temperature-driven seasonal spawning conditions, this could lead to greater volatility of the populations of both predator and prey as fewer alternative resources are available. The impact on commercial fisheries is also a source of speculation. With a higher lipid and overall energy content than Atlantic cod, *B. saida* may prove difficult to replace without significant impacts to other stocks (Lawson et al., 1998). Even if the current consumed biomass of *B. saida* was evenly distributed among the comparable commercial species as natural predation, it is likely that commercial harvest will decrease, or at least become much more variable as the number of stressors on traditional target species increases. However, the direct predation upon

B. saida by commercial target species or even as bycatch by the larger commercial fisheries is not well studied, even as both instances of mortality may be increasing.

It is also likely that regional productivity will be spatially restructured. Even as the increase of warm Atlantic inflow and loss of sea ice is forecast to increase primary productivity, this is anticipated to promote a shift to smaller, lipid-poor zooplankton. The loss of the energy-rich zooplankton primary consumer, paired with a lack of intermediate predators such as *B. saida* in ice adjacent areas, may limit the accessibility and ecological utility of this increased biomass (Bluhm & Gradinger, 2008). Declines in condition, growth or production of marine mammals have already been observed in recent decades, possibly reflecting decreased availability of *B. saida* in the Beaufort Sea (Mueter et al., 2016). Without *B. saida*, the ice associated lipid source may be lost even if the ice persists, as boreal replacements adopt a more general pelagic feeding strategy.

Population structure of *B. saida*

Distinct population structure is less common in marine populations where absolute barriers to gene flow are few (Palumbi, 1994). However, the distinct fjord and offshore habitats along north-east Greenland and Svalbard have given rise to distinct populations of *B. saida* with apparent genetic structure (Madsen et al., 2016). Historical structure and distribution of north-east Atlantic *B. saida* populations can be reconstructed from genetic and geological records. At the last glacial maximum approximately 20,000 years ago, the fjords now inhabited by *B. saida* in north-east Greenland would have been ice covered and inaccessible. However, there is mitochondrial DNA evidence that the *B. saida* were already present in the adjacent north-east Atlantic (Pálsson et al., 2009). The results of microsatellite studies have brought this history and the overall population structure into further focus. Surprisingly, distance is not quite the predictor of connectivity that it was previously understood to be. Microsatellites have shown adjacent fjord and coastal populations are often more distinct than two fjord populations are, even when the fjords are not geographically close (Madsen et al., 2016). While there are several possible explanations for this phenomenon, a founder effect caused by limited fjord recolonization post glaciation and subsequent reproductive isolation by changing land/sea levels, fits the combination of glacial, fossil, and genetic evidence. However, other studies have found gene flow appears to occur over long distances, with no isolation by distance occurring at scales up to 2000 kilometers (twice the breadth of the Greenland Sea study area) (Maes et al., 2021).

The structure and ongoing divergence of *B. saida* populations have substantial implications for our understanding of climate change impacts and management requirements. Furthermore, genetic analysis of *G. morhua*, a close relative of *B. saida*, has revealed divergence in the *Pan I* gene governing transmembrane protein activity. Stationary fjord and migratory coastal *G. morhua* populations display differential *Pan I* activity as a response to different environmental regimes of light, temperature, salinity, and depth (Andersen, Johnsen, et al., 2015). While historically resilient in the face of natural climatic fluctuation, the magnitude and rate of anthropogenic climate change threatens both the migratory and stationary *B. saida* populations. However, these populations are subject to very different challenges and may be sufficiently divergent to have different capacities for adaptation. Changes in seasonal fjord ice cover threaten the reproductive life history of stationary populations of *B. saida*. The rapid increase of Arctic temperatures endangering critical marine ecosystems, is thought to be largely driven by changes in sea ice (Landrum & Holland, Marika, 2020). A physical study of a typical *B. saida* containing fjord in NE Greenland found that ice cover "can be predicted to decline approximately from 1.4–1.9 (depending on snow thickness) to 0.8–1.3 m, with the expected increase in the ice-free season from 2.5 months today to 4.7 months at the end of the century" (Christensen et al., 2007; Rysgaard et al., 2003). This would vastly alter the quality and quantity of habitat for the sympagic juvenile stage *B. saida* that are categorically defined by their development in association with ice cover. These findings are supported at a larger scale by runs of the Community Earth System Model Multi Model Large Ensemble (CESM-MMLE) with different model groups showing 1) a decline in mean sea ice extent 2) an increase in year to year sea ice variability and 3) the possibility of an ice-free Arctic ($<1 \times 10^6$ km^2 ice area) as early as 2023 under the RCP 8.5 "business as usual" climate scenario (Landrum & Holland, Marika, 2020). This leads to the creation of a habitat more suited for boreal adapted species such as capelin and southern gadids, increasing potential competition for resources and reducing the effective niche of the Arctic specialized *B. saida* (Mueter et al., 2016). While also subject to increased competition from boreal invaders, migratory populations of *B. saida* face a greater threat from variable sea ice cover and the commercial fishing fleets tracking boreal stocks into the Arctic. Already experiencing substantial stock fluctuations, *B. saida* is forecast to undergo increased recruitment variability as a result of changing sea ice extent and sea surface temperatures (van der Meeren et al., 2019). The recruitment of two spawning populations originating from Svalbard and the Pechora sea respectively, was modeled using Lagrangian particle physics as well as ice extent, sea surface temperature, and the following year's survey catch (Huserbråten et al., 2019). The outcome is an apparent negative relationship

between sea surface temperature and recruitment, as well as a positive relationship between sea ice extent and recruitment (Figure 4).

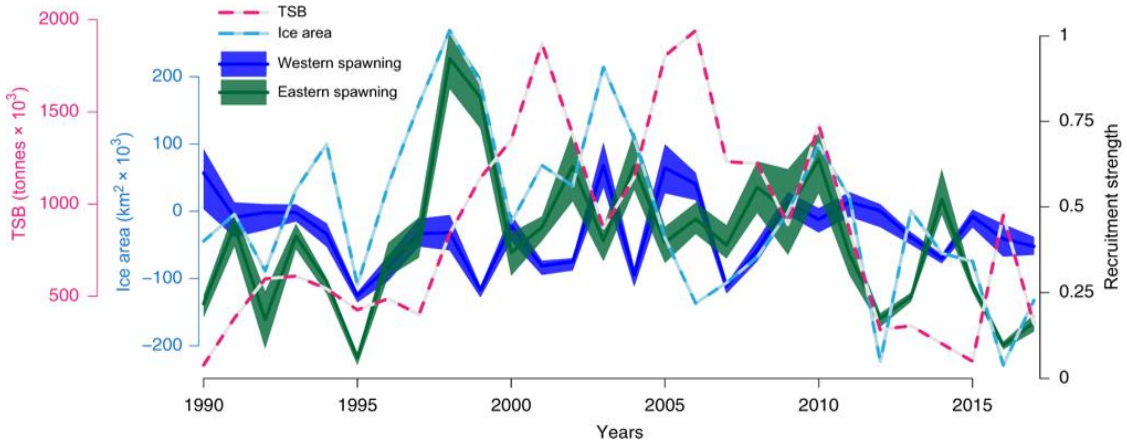


Figure 4: Total stock biomass (TSB), maximum ice cover, and the back-calculated recruitment strength of B. saida from spawning assemblages east of Svalbard and Pechora Sea. Here estimated TSB is measured in thousand metric tonnes and represented by a pink line, and the light blue line represents the annual maximum marginal ice cover (i.e., area with ice concentration above 15% per grid cell) (Huserbråten et al., 2019).

Though likely the aggregation of many factors, the ice facilitated co-occurrence of juvenile *B. saida* and the energy rich naupliar stages of *Calanus* copepods is a major determinant of recruitment success, as it is for many species in the Barents Sea (Castonguay et al., 2008; Kvile et al., 2016). In this way, the Southeastern Barents Sea may undergo a shift from immensely productive *B. saida* nursery to larval death trap. Even if the Svalbard spawning ground remains relatively unaltered, the increasing scarcity of ice-associated prey and predation by boreal arrivals, coupled with summer temperatures of trapped water masses exceeding hospitable levels, threatens critical *B. saida* life stages (Huserbråten et al., 2019). While the replacement of Arctic sympagics like *B. saida* by boreal species may seem to soften the blow of biodiversity loss, the trophic ecology of *B. saida* is likely irreplaceable in the system as a whole.

Gene expression as a tool to explore biological relationships

Traditionally, the investigation of population structure and the relationship between groups was done at the genetic level through the identification of mutations in a single gene encoding DNA region. Additionally, this type of genetic study was initially only carried out in so called “model organisms” whose biology was already well documented and exhaustively studied via traditional non-molecular methods (Müller & Grossniklaus, 2010). As technology

has progressed, studies have expanded from targeting the genotype of single genes, to characterizing genotype and differential gene expression (DGE) in many genes at once through what is known as a genome wide study using high throughput tools such as microarrays (Nuwaysir et al., 1999). Now, there is no need to identify a selection of genes to be profiled, as full genome and transcriptome sequencing allows for an organism's entire library of genetic material, including unknown regions, to be assessed at once (*de novo*) (Rao et al., 2019). Full transcriptome RNA sequencing captures an expression profile of an organism at the moment of collection, providing a basis of comparison between groups in different environments, even if they share the same genotype (Yang et al., 2015). Furthermore, this approach is now possible in non-model organisms. Recent studies characterizing divergent ecotypes (Ishikawa et al., 2017), gene expression in migrants between environments (Lohman et al., 2017), and the evolution of gene expression profiles (Z. Chen et al., 2008), all applying an RNA sequencing approach to non-model organisms (three spine stickleback (*Gasterosteaus aculeatus*) and an Antarctic notothenioid (*Pagothenia borchgrevinki*)) show the sheer potential of such methods to address research questions related to expression-level adaptation in new environments. This newfound accessibility of exhaustive gene expression information in non-model organisms brings to bear a powerful investigative technique in understudied, yet ecologically significant, species. The *QuantSeq 3' mRNA library preparation* (Lexogen) approach employed here consistently produces high read coverage suitable for downstream DGE and biological pathway enrichment analysis, all while requiring only a small amount of input total RNA and less per-sample investment (Corley et al., 2019). As a non-model organism with unknown population structure encompassing a wide range of habitats, *B. saida* is a prime candidate for novel study via a newly possible full transcriptome sequencing approach.

Despite persistent questions related to the basis of observed structuring in the Greenland Sea, the biological drivers of *B. saida* population structure remain unexplored. Investigative genetic studies have led to a widespread suggestion that these populations are diverging in the same manner as coastal and migratory populations of Atlantic cod (Karlsen et al., 2013), or at least exhibiting differential phenotypic plasticity/gene expression in response to different home environments (Andersen, Johnsen, et al., 2015; Fevolden & Pogson, 1997; Madsen et al., 2016). However, there have not been transcription based expression studies to verify this fact. This approach is better suited to non-model organisms than full genome sequencing and is easier to assemble due to the expectation of uneven coverage (Gibbons et al., 2009). As a result of the limitations of a purely genetic approach, the biological basis of the observed genetic differences between fjord and oceanic groups remains uncharacterized. This makes it impossible to

adequately describe the subtle differences in ecology between these groups. Furthermore, the development of a high-quality *B. saida* genome resource lags significantly behind that of the closely related *G. morhua* (Atlantic cod). Identification of specific biological processes exhibiting divergent regulation in response to the differences between fjord and coastal environments could provide important quantitative insight into the basis for future acclimation potential.

Study objectives

This study seeks primarily to determine whether fjord and offshore populations of *B. saida* are exhibiting significantly different gene expression, in order to identify patterns of ongoing divergence, if any is occurring. Emerging full transcriptome sequencing approaches have the potential to map sequencing reads to gene ontologies, effectively highlighting specific biological pathways being differentially regulated between populations (Eldem et al., 2017). Additionally, a full transcriptome approach allows for the identification of underlying biological processes contributing to any significant differential expression, providing a quantitative measure of environmental influence as a possible driver of divergence. In this way, this project also aims to provide a novel characterization of the transcriptome of *B. saida*, a species of importance in Arctic food webs and at significant risk of decline due to climate change.

Hypotheses

1. Is there significant relative differential gene expression (DGE) between neighboring Greenland fjord and shelf populations of *B. saida*?
    a. **H0:** There is no difference in relative gene expression between adjacent fjord and shelf populations of *B. saida*.
    b. **H1:** There is significant differential gene expression between adjacent fjord and shelf populations of *B. saida*.
2. Is there a connection between the physical characteristics of Greenland fjord/shelf environments and the pathways being differentially expressed in the *B. saida* populations they host?
    a. **H0:** There is no connection between the physical characteristics and pathways – differentially expressed genes are distributed randomly/evenly throughout transcriptome.
    b. **H1:** Differences in expression are clustered around pathways associated with metabolism, homeostasis, and osmoregulation, reflecting the different thermal and chemical environments associated with fjord vs. coastal habitats.

METHODS

Tissue collection

Individuals were collected via trawl on the TUNU cruise track in sampling years 2013, 2015, and 2017 (Figure 5). The first 20-30 dissections from each trawl were preserved in RNAlater, separated by tissue, and incubated at 4˚C for 12-18 hours prior to storage at -80 ˚C.

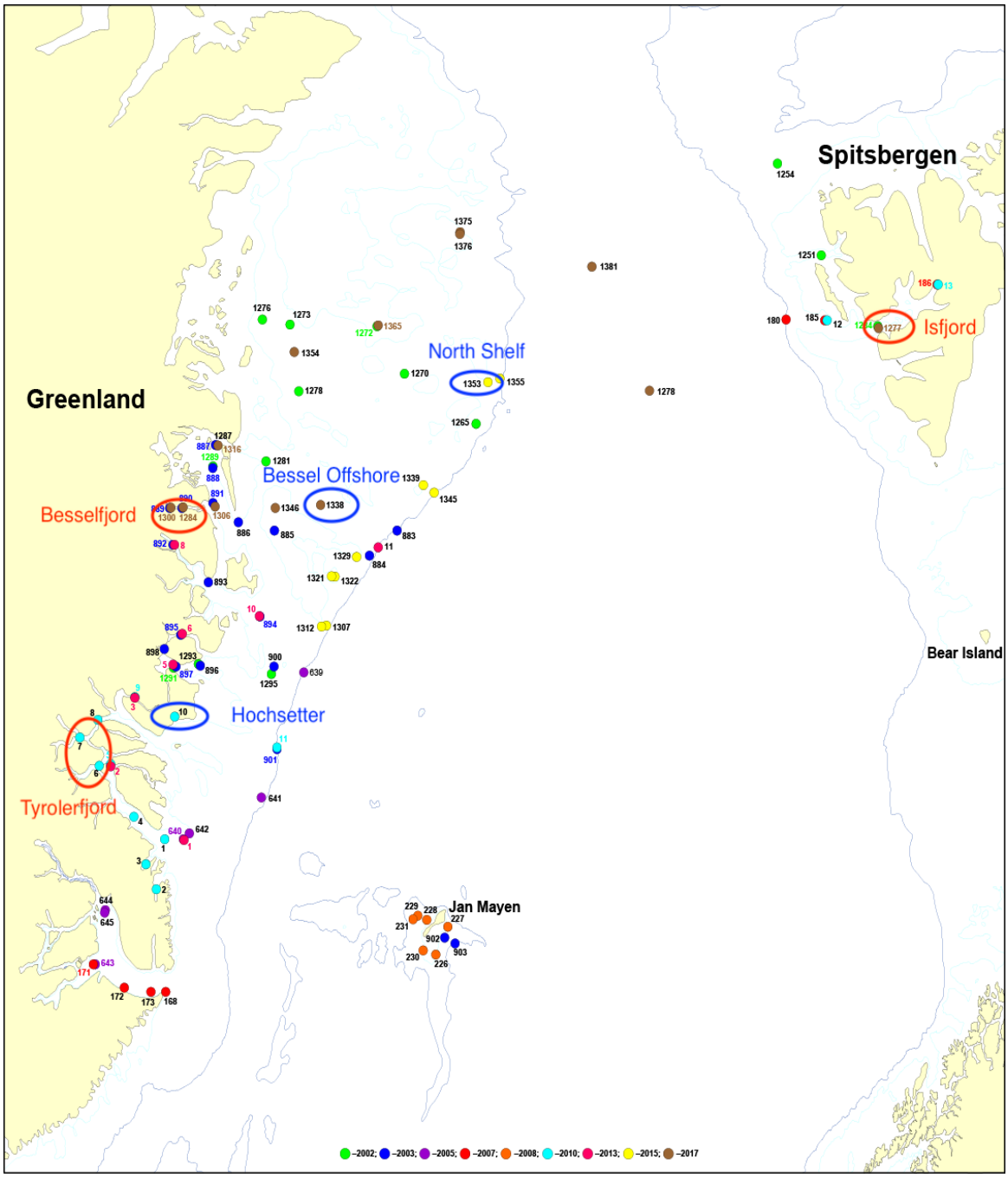


Figure 5: Stations used to represent fjord environments are circled in red, stations used as offshore environments are circled in blue, and the fill color of the individual points corresponds to the year of collection.

Stations were paired by year, except for Isfjord (station 1277, collected 2017) and Shelf (station 1353, collected 2015), which were paired due to the lack of samples from the same year (see Table 1 for details).

Table 1: Collection site metadata, color coded to reflect neighboring fjord-offshore pairings.

Location	Environment	TUNU Station(s)	Year	Count
Besselfjord	Fjord	1284/1300	2017	27
Bessel Offshore	Offshore	1338	2017	26
Isfjord	Fjord	1277	2017	16
North Shelf	Offshore	1353	2015	25
Tyrolerfjord	Fjord	6/7	2013	25
Hochsetter	Offshore	10	2013	25

Extraction

Extractions were made from the RNAlater preserved tissues using the Protocol for the Purification of Total RNA from Animal Tissues in the *Qiagen RNeasy Mini Kit* (Qiagen, USA). The extraction method was as specified by the kit including the optional protocols of adding DTT to the RLT buffer, and homogenizing using a 2 min. run in the *Qiagen TissueLyser II* (Qiagen, USA), and a drying centrifugation step prior to elution. The full kit protocol used can be found in Appendix I on page 55.

Quality control of eluted RNA

Post-extraction RNA concentration and quality were verified using *NanoDrop* (Thermo Fisher Scientific, USA) and the *2100 Bioanalyzer* (Agilent Technologies, USA). The *Qubit Broad Range RNA* (Thermo Fisher Scientific, USA) protocol was used prior to cDNA library preparation to obtain more accurate RNA concentration estimates.

cDNA library preparation

cDNA libraries were prepared using the *QuantSeq 3‘ mRNA-Seq Library Prep Kit* (Lexogen, Austria) and stored at -20˚C prior to pooling and sequencing on the *Illumina HiSeq 4000* platform. Dual indexing was applied using the *i5* and *i7* indices provided (full protocol can be found in Appendix I on page 55). The only modifications to the QuantSeq protocol concerned the number of PCR cycles used (13) and the drying time during bead purification (shortened from 5-10 min. to 3-4 min.). Samples were divided into two groups representing the two final sequencing pools and randomly assigned to library prep batches.

Two duplicate samples were added to each lane from the other (for a total of 4 duplicate samples per lane) to provide a basis for assessing lane effects. *Spike-in RNA Variant* (SIRV) (Lexogen, Austria) controls were added in the volume of 1 µL to every sample at the beginning of the cDNA synthesis protocol. Pools were constructed to a target concentration of greater than 2 nM with 10 fmol of each sample being added for a total volume of approximately 261 µL (74 samples) per pool. Pools were then shipped frozen to Novogene (United Kingdom) for sequencing on the Illumina HiSeq 4000 platform (1 pool per lane).

Statistical treatment and bioinformatic analysis

Prior to DGE analysis, morphological characteristics including length, weight, and condition, as assessed using Fulton's K, were assessed to verify the inter-comparability of samples both within and across sampling sites. After testing for normality and homogeneity of variances, *Analysis of variance* (ANOVA) and post-hoc *Tukey Honest Significant Difference* (HSD) tests were utilized in this assessment. Based on established weight-length relationships in fish, the weight of an individual is expected to increase at a cubic factor relative to the length, giving a slope of 3 when plotted as a regression line in the logarithmic form (Froese, 2006). This cubic relationship justifies a weight-length linear regression model based on log transformed data. Fulton's K condition factor was used to roughly approximate condition based on the relationship between length and weight, as physiological condition could affect gene expression (Striberny et al., 2019). ANOVA was used to verify the inter-comparability of sampled fish based on length-weight relationship and Fulton's K condition factor. Its assumptions are met based on the normal distribution of the Fulton's K and homogenous variance across locations. Normal probability plots verifying the normal distribution and homogeneity of variance of Fulton's K as well as measured length and weight can be found in Suppl. Figure 1.

Following sequencing, bioinformatic analysis was performed on the Research Genetics Group's *Spygene* server, starting with demultiplexing via the *demuxFQ* package (https://github.com/gdbzork/demuxFQ). The quality of each step's output was verified using the *FastQC* (Andrews, 2010) and *MultiQC* (Ewels et al., 2016) tools. After data demultiplexing, adapter and junk sequences were removed using *Bbduk*, (https://github.com/BioInfoTools/BBMap/blob/master/sh/bbduk.sh) as recommended by Lexogen. All reads shorter than 20bp after trimming were discarded. For sequence alignment, the *SIRVome* was appended to the *G. morhua* 3.0 genome, which was then indexed using *STAR*

(https://github.com/alexdobin/STAR). The *B. saida* genome was initially selected for alignment, but was replaced by that of *G. morhua* due to the higher quality assembly and close evolutionary relationship between the two species. Due to 30% of the reads failing to align due to length, STAR's parameters were optimized for short reads *(--outFilterScoreMinOverLread 0 --outFilterMatchNminOverLread 0 --outFilterMatchNmin 0*). Once aligned, the reads were distributed into gene feature counts with *htseq-count* (Anders et al., 2014), using *gene-id* as the main identifier for downstream expression analysis. Gene counts were collected by lane and provided as input for the *DESeq2* R package (Love et al., 2014). DESeq2 was used to generate distance estimates, *log2fold* change expression differences, and tables of significance. Based primarily on sample-to-sample distance, 11 outliers were removed from the DESeq2 tables used for differential gene expression (Suppl. Table 1). Genes identified by the DESeq2 results function as significant at the adjusted p-value <0.01, were fed into the NCBI *eFetch* utility (Eric Sayers, 2018) and converted to standardized unique identifiers (UIDs). These UIDs were then used for annotation via the *FishEnrichr* tool from the Maayan research group (E. Y. Chen et al., 2013; Kuleshov et al., 2016). The *FishEnrichr* Gene Ontology (GO) ID outputs were then clustered and visualized using the *simplifyEnrichment* package for R. A detailed step by step description of the bioinformatic analysis can be found in Appendix II on page 61.

RESULTS

Inter-comparability of sampled *Boreogadus saida*

Prior to DGE analysis, the inter-comparability of the individual samples and sample groupings was verified at multiple levels. First, basic morphological characteristics of sampled *Boreogadus saida* were compared to ensure that morphology was not a primary contributor to the observed gene expression. Histograms of the length distribution by location showed roughly normal distributions covering different ranges of lengths, with Isfjord's distribution reflecting a smaller mean (Figure 6).

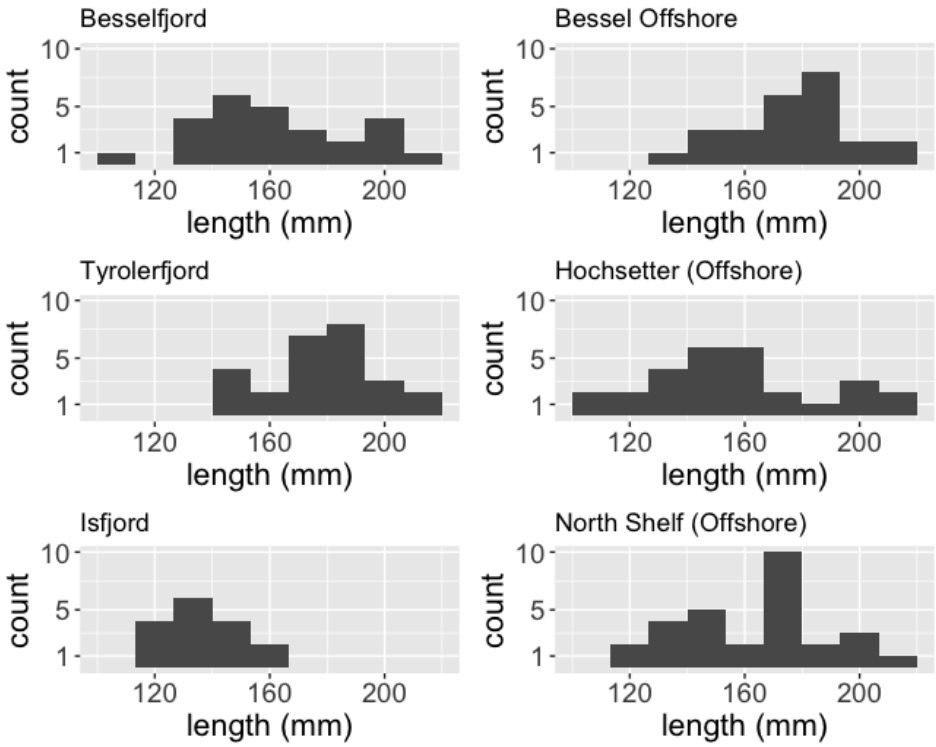


Figure 6: Histograms of length in mm by station location. Fjord locations are in the left column and the corresponding offshore location is in the right column.

Linear models relating weight to length were fit to each site, with shape of the data and residuals (homogeneity of variance improved, no pattern in residual plot after log transformation) suggesting a log-linear fit between length and weight. Highly significant log10-linear models were fit to each collection site (Figure 7).

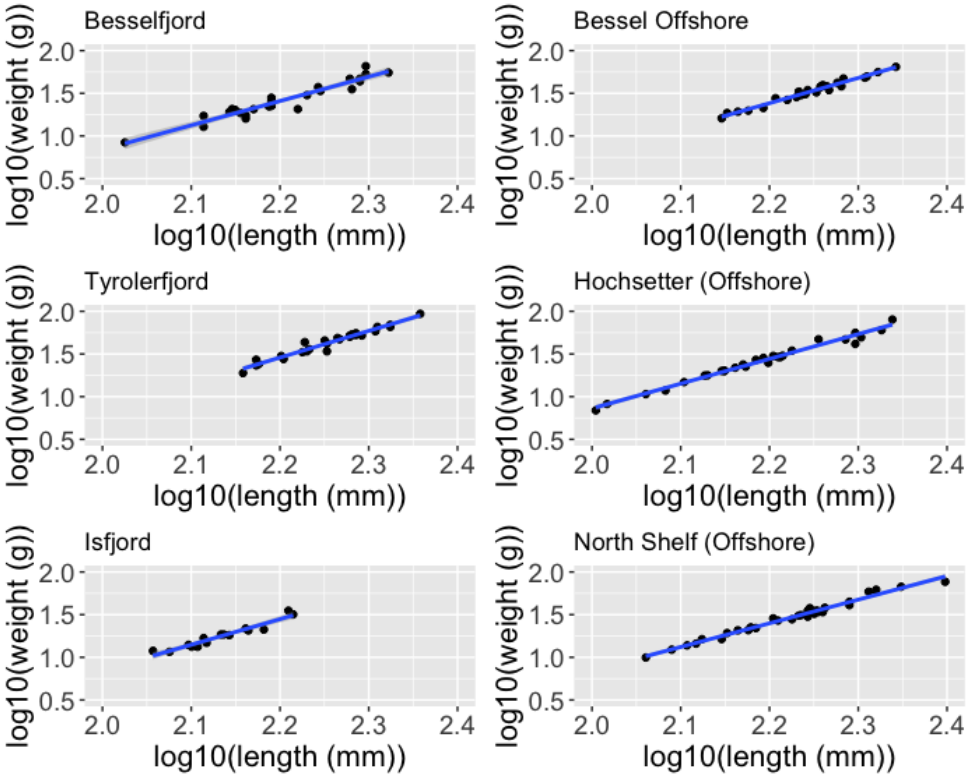


Figure 7: Linear regressions of log10 transformed length (mm) and weight (g) data by capture location. Fjord locations are in the left column and corresponding offshore sites are in the right column.

Based on the equivalent slopes as verified via ANOVA on the linear regressions, there is no significant difference between the log(weight-length) relationships at the study locations (Table 2).

Table 2: ANOVA output table based on a test for homogeneity of the slopes of the linear models describing the relationship between log transformed length and weight measurements across all sampling locations. The lack of significance in the relationship between location and log(length) (highlighted row) suggests slope homogeneity.

ANOVA Table (type II tests)					
Test for homogeneity of linear regression slopes log10(weight-length)					
Effect	DFn	DFd	F	p	ges
location	5	145	20.305	<<0.01	0.412
log10(length)	1	145	4625.523	<<0.01	0.97
location:log10(length)	5	145	1.487	0.198	0.049

To incorporate both length and weight measurements, as well as estimate a condition factor of the study individuals, Fulton's K (Figure 8) was used for an ANOVA comparison within fjord-offshore pairings.

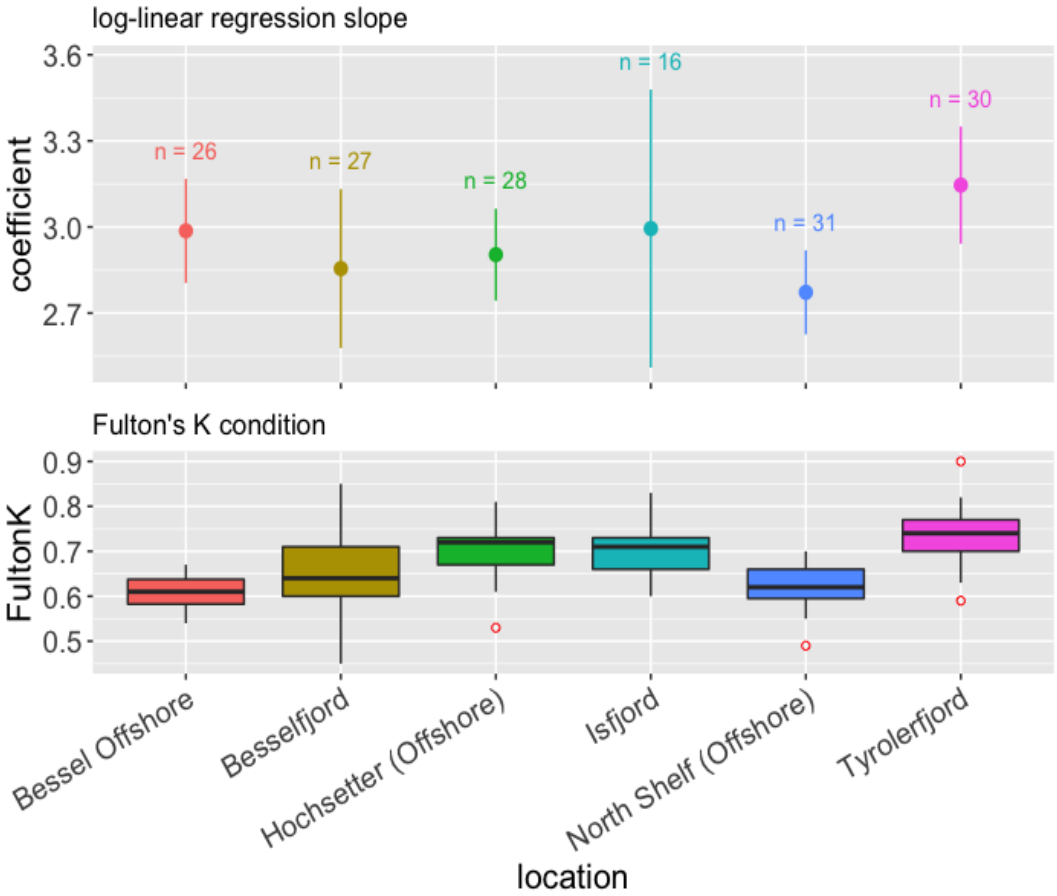


Figure 8: Slopes of log-linear fitted regression lines (top) by location with 95% confidence intervals. Boxplots of Fulton's K condition index by location with outliers as unfilled points (bottom).

Significant differences in condition as measured by Fulton's K are identified via a one way ANOVA (Table 3). This analysis included all possible location pairings, including those within environment types (fjord-fjord, offshore-offshore).

Table 3: ANOVA output for a model explaining differences in Fulton's K based on location. A p-value much smaller than 0.001 suggests there are highly significantly differences between the Fulton's K values of certain location pairings, but this analysis does not identify which pairings are driving the significant result.

Fulton's K ANOVA Table					
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
location	5	0.362	0.0723	20.06	p<.001
Residuals	151	0.544	0.0036		

A post-hoc Tukey HSD test was employed to identify the study-relevant pairings with significant differences in Fulton's K. The ANOVA and Tukey HSD tests revealed that there was no significant difference in Fulton's K condition factor between the habitat pairings of Besselfjord and Bessel Offshore as well as Tyrolerfjord and Hochsetter, however, there was a significant difference (North Shelf-Isfjord difference -0.087±0.057, p-value <0.01) between the Isfjord and Northern Shelf stations (Figure 9). There was no significant difference between any other habitat pairing relevant to the study.

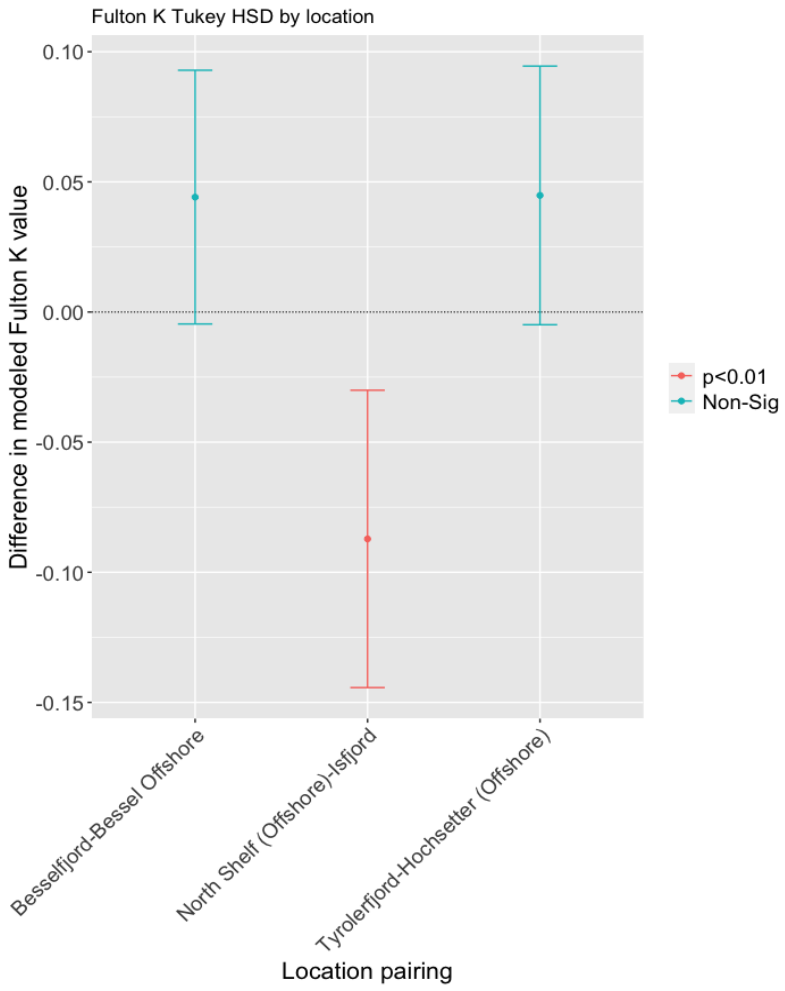


Figure 9: Results of the post-hoc Tukey Honest Significant Differences test by location pairing. A confidence interval overlapping zero indicates no significant difference.

Differential gene expression

Environmental groupings

When grouped by environment type (fjord vs. offshore) alone, Tyroler/Hochsetter and Isfjord/Shelf North tend to cluster most distinctly based on principal components of variance, with Besselfjord/Offshore overlapping with one or both other groupings (Figure 10). All Principal Componenet Analyses (PCAs) presented include ellipses drawn at 95% confidence level based on a multivariate t distribution. This differentiation is strongest between the Tyrolerfjord and Isfjord stations. Outliers from a number of the sites were removed prior to this analysis due to anomalous sample to sample distance or low read count. As a result, sample sizes varied from group to group as well (Fjord:Bessel n=26, Fjord:IsFj_Shelf n=15, Fjord:Tyroler n=24, Offshore:Bessel n=26, Offshore:IsFj_Shelf n=23, Offshore:Tyroler n=25). Additional PCAs including less explanatory components are presented in Suppl. Figure 2.

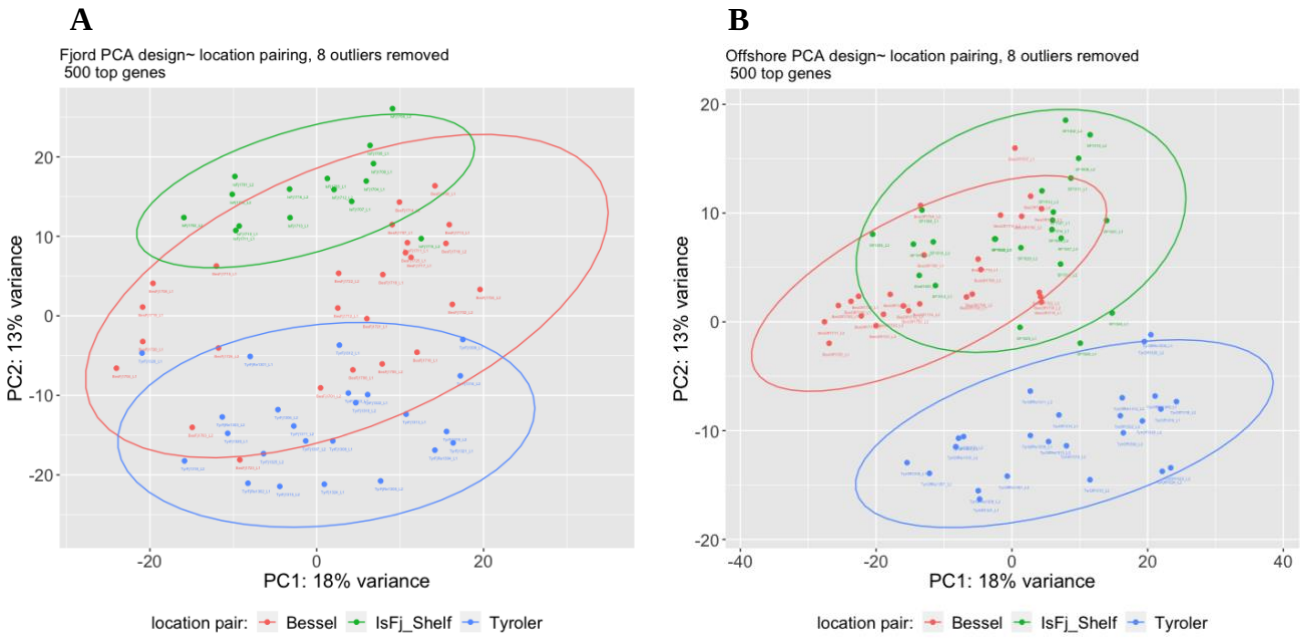


*Figure 10: Principal Component Analysis (PCA) of the environment groups (all fjords in panel **A**, all offshore locations in panel **B**) based on 500 most differentially expressed genes. PCA ellipses drawn at 95% confidence level based on a multivariate t distribution.*

Isfjord and the North Shelf

Station 1277 (Isfjord, TUNU 2017) on western Svalbard and Station 1355 (North Shelf, TUNU 2015) are the most geographically and temporally distant pairing. Isfjord had the lowest sample count after sequencing and outlier removal with 15 samples compared to 23 for the offshore pairing (3 outliers based on extreme sample to sample distance removed). A two

component PCA of the fjord-offshore pairing did not yield significantly different clusters (Figure 11).

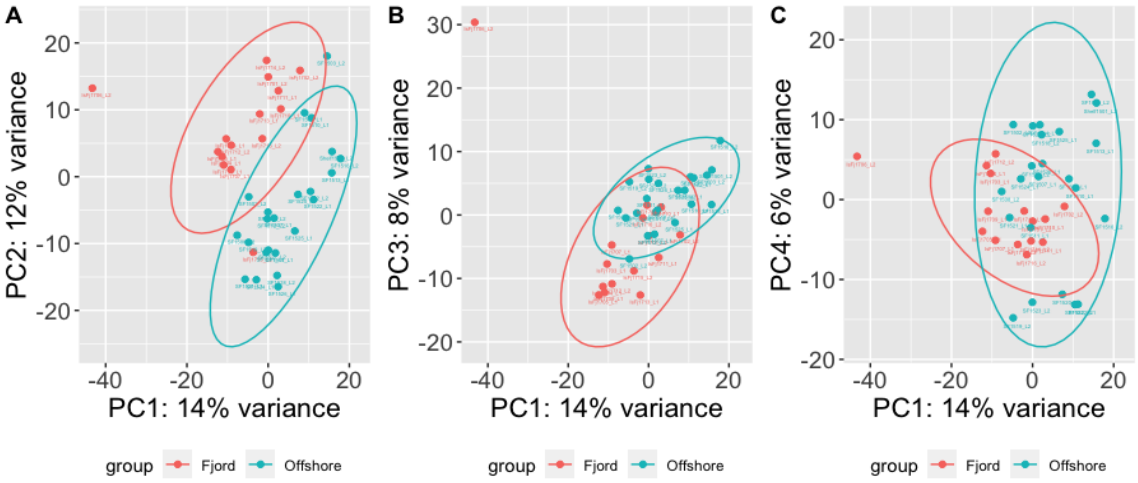


Figure 11: Principal component analysis (PCA) of Isford and North Shelf Offshore, with 5 outliers removed, for the top 4 identified explanatory components from most explanatory (A) to least (C) based on the top 500 most differentially expressed genes. PCA ellipses drawn at 95% confidence level based on a multivariate t distribution.

During differential gene expression using DESeq2, 19,919 genes with greater than 5 reads were assessed. Based on log2 fold change (LFC) at a p-value < 0.1, 882, 4.4% of these had LFC > 0 (upregulated) while 957, 4.8% had LFC < 0 (downregulated). Additionally, 4660 (23%) had low counts, and 0 were identified as outliers. From this set, 618 genes emerged as significant with a p-value adjusted for multiple testing of less than 0.01. These genes were annotated with Gene Ontology (GO) terms by biological process using the FishEnrichr enrichment analysis tool (Figure 12).

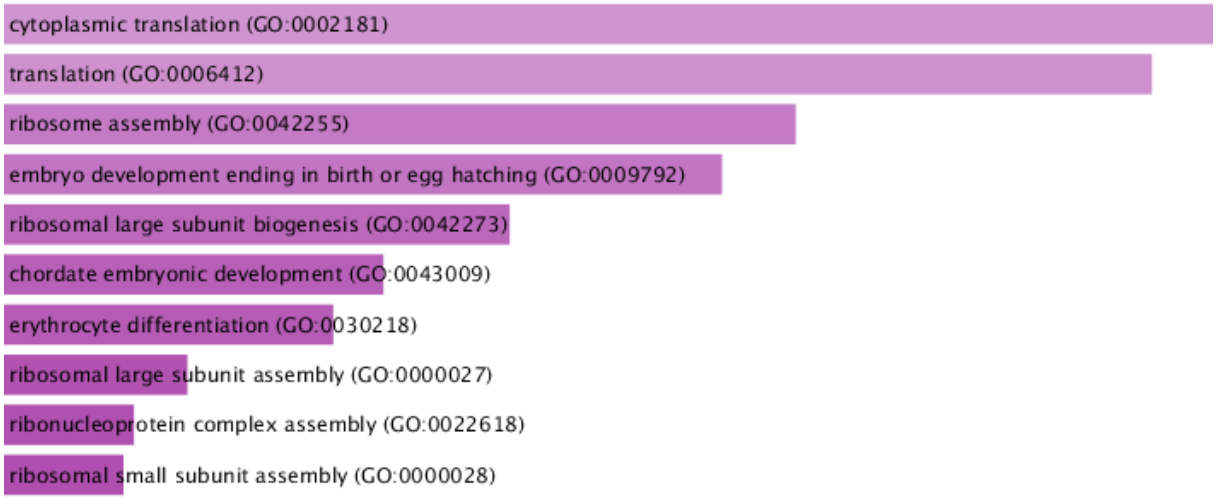


Figure 12: Significant (p-value <0.01) gene ontology clustering by biological process ranked by adjusted p-value. Biological processes are determined and annotated based on differentially expressed genes between the environmental groups, in this case Isfjord and the Northern Shelf.

GO terms with an adjusted p-value of less than 0.05 were then clustered and visualized, giving rise to several distinct groups (n=43, Figure 13). These groups are plotted according to GO term similarity and annotated with word clouds based on GO cluster keywords.

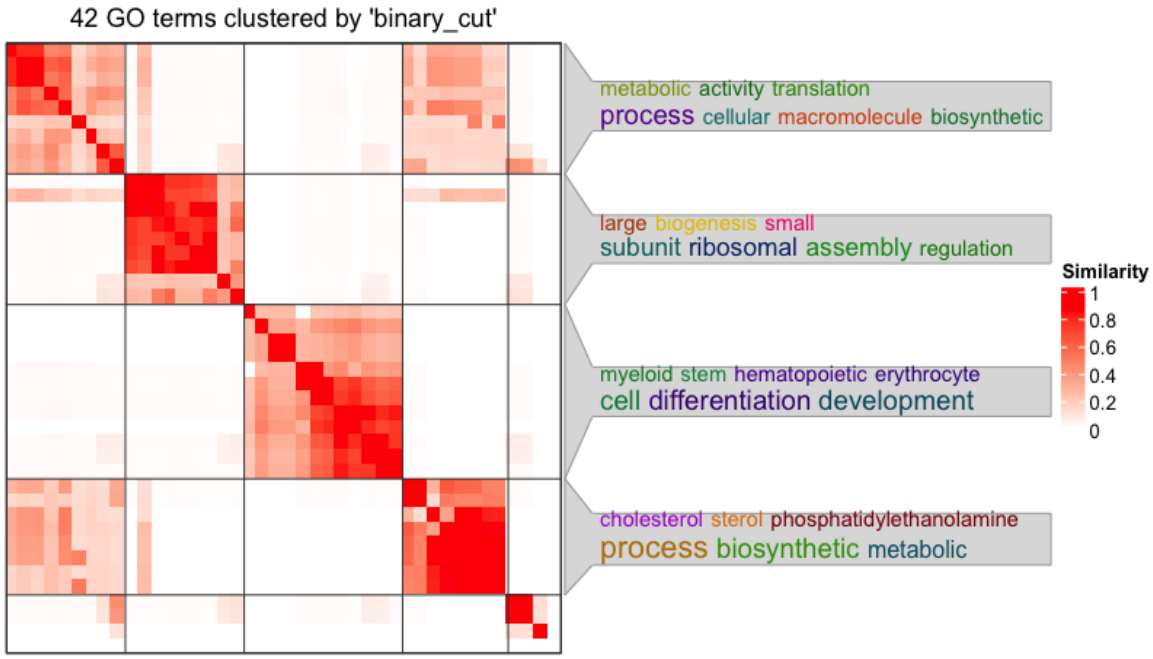


Figure 13: Heatmap of similarity between gene ontology clusters of significantly differentially expressed genes between Isfjord and Shelf North.

Besselfjord and Bessel Offshore

Station 1300 (Besselfjord, TUNU 2017) and Station 1338 (Bessel Offshore, TUNU 2017) are a geographically intermediate fjord-offshore pairing. Station 1300 is located at the back of the W-E facing multi-basin Besselfjord (Zoller 2020). Station 1338 is located approximately midway between the coastline and shelf break, providing a longitudinal midpoint between Hochsetter and the Northern Shelf stations. A two component PCA did not show separate clustering among the top four generated components (Figure 14).

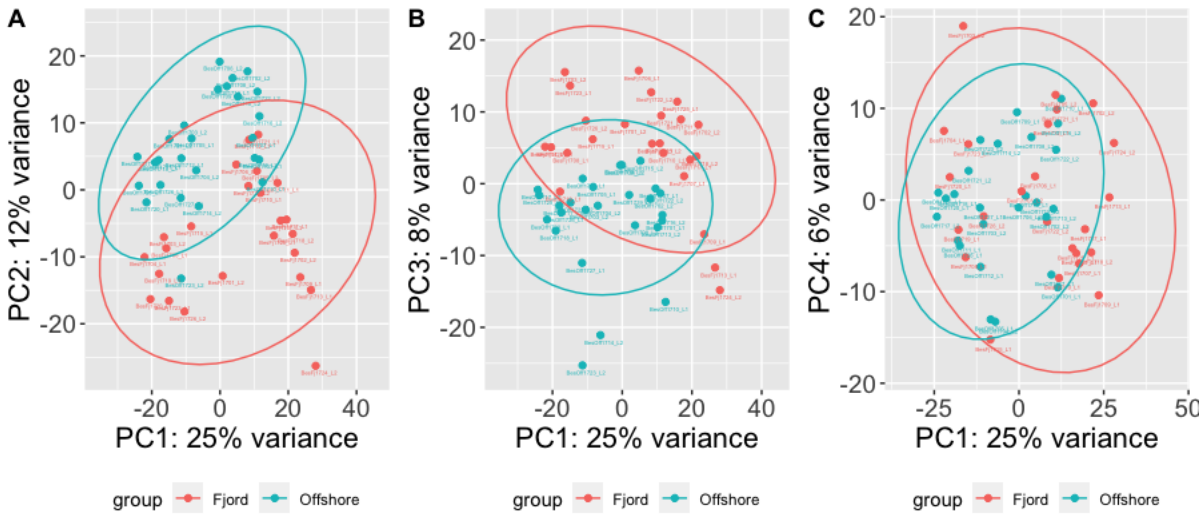


Figure 14: Principal component analysis (PCA) of Besselfjord and Bessel Offshore for the top 4 identified explanatory components, with 3 outliers removed, from most explanatory (A) to least (C) based on the top 500 most differentially expressed genes. PCA ellipses drawn at 95% confidence level based on a multivariate t distribution.

Through differential gene expression analysis, 20,514 genes were identified with a read count greater than 5. Of those, based on log2 fold change, 1565, 7.6%, had LFC > 0 (upregulated) and 1142, 5.6%, had LFC < 0 (downregulated). Additionally, 5175, 25% had low counts and 0 were identified as outliers. From this set, 976 genes were selected for GO annotation with a p-value adjusted for multiple sampling < 0.01. These genes were annotated with Gene Ontology (GO) terms by biological process using the FishEnrichr enrichment analysis tool (Figure 15).

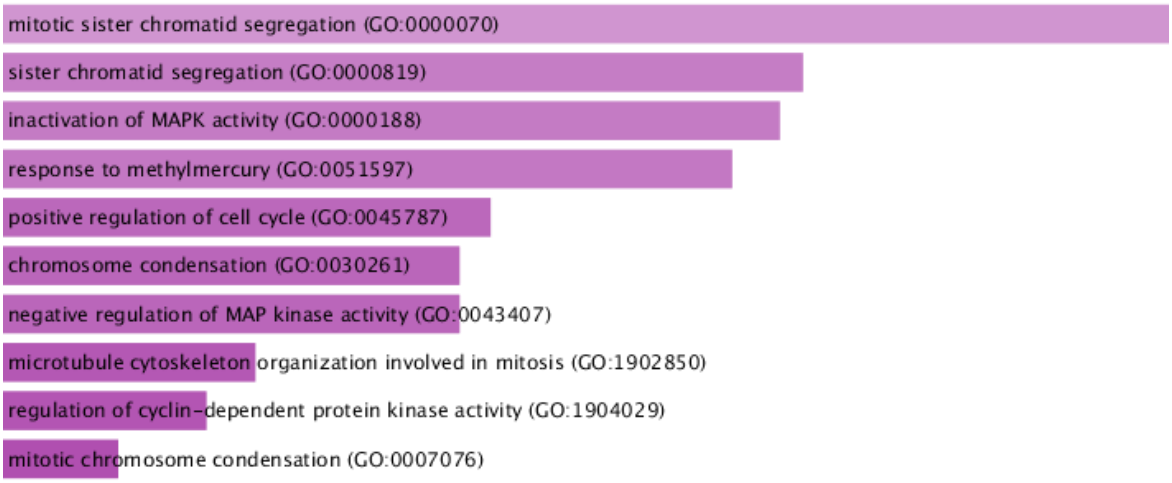


Figure 15: Significant (p-value <0.01) gene ontology clustering by biological process ranked by adjusted p-value. Biological processes are determined and annotated based on differentially expressed genes between the environmental groups, in this case Besselfjord and Bessel Offshore.

Only 7 GO terms emerged as significant at the p-value < 0.1 level, so this level was used in an attempt to have some terms to characterize the site pairing. These terms were assembled into 3 cluster groups and visualized according to GO term similarity (Figure 16). The clusters tracked along 3 biological processes: MAPK inactivation and deregulation, chromosome condensation and sister chromatid segregation, and response to methylmercury.

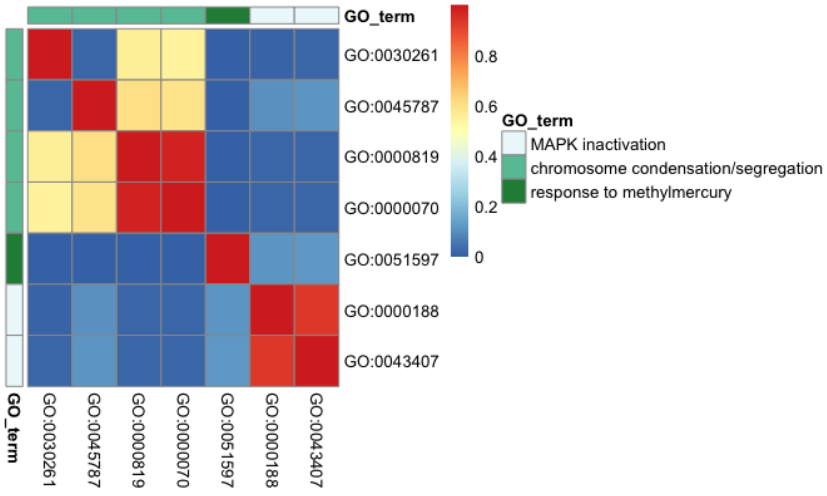


Figure 16: Heatmap of GO ID similarity with warmer colors indicating more similar and cooler colors indicating less. Consolidated GO term groupings for each GO ID are indicated by the green color scale

Tyrolerfjord and Hochsetter

The southernmost pairing consists of Station 6/7 (Tyrolerfjord, TUNU 2013) and Station 10 (Hochsetter, TUNU 2013). This pairing is the most coastal, with even the "offshore" Hochsetter site being much nearer to land on the shelf than Bessel Offshore or the Northern Shelf. Tyrolerfjord is a long, silled fjord lying W-E at 74N further separated from the shelf waters by Young Sound and a second outer sill. The outer sill rises to a depth of 45m and the inner to 56m, approximately 70km from the outer sill, with the back of Tyrolerfjord and the surrounding land terminating glaciers of the Greenland ice sheet sitting 90km from the entrance to the outer fjord (Boone et al. 2017). A 2 factor PCA failed to show clustering by environment group (fjord vs. offshore) based on the top 4 components ranked by explained proportion of the observed variance (Figure 17). Through differential gene expression analysis in the DESeq2

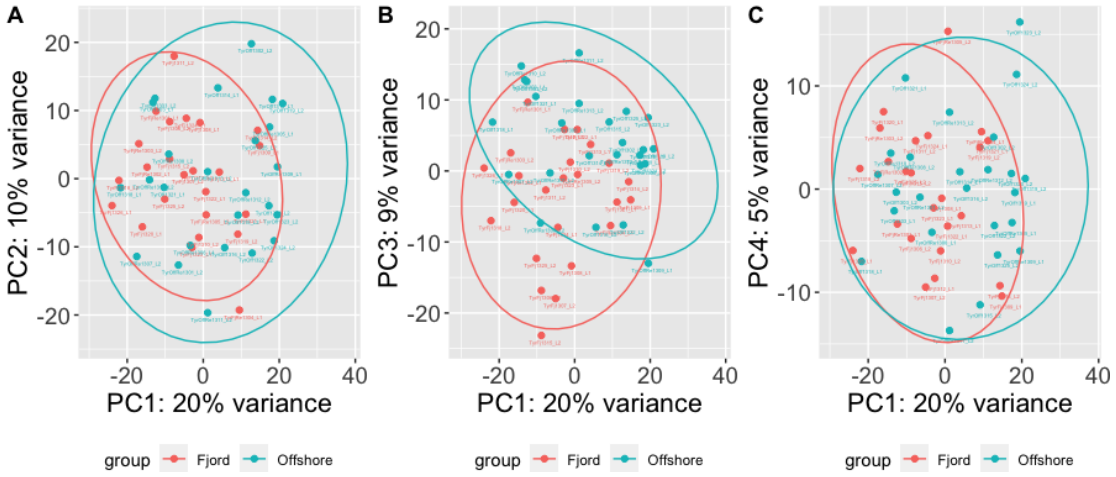


Figure 17: Principal component analysis (PCA) of Tyrolerfjord and Hochsetter for the top 4 identified explanatory components, with 3 outliers removed, from most explanatory (A) to least (C) based on the top 500 most differentially expressed genes. PCA ellipses drawn at 95% confidence level based on a multivariate t distribution.

package, 23,614 genes were identified with at least 5 reads. Of the identified genes with adjusted p-value < 0.1 1,015 (4.3%) had a log2 fold change greater than zero (upregulated) versus 1,007 (4.3%) with a log2 fold change less than zero (downregulated). Additionally, 8,602 (36%) genes had low counts (mean count < 1) and none were identified as outliers in this analysis. From this set, 602 genes were selected for GO annotation with a p-value adjusted for multiple sampling < 0.01. These genes were annotated with Gene Ontology (GO) terms by biological process using the FishEnrichr enrichment analysis tool (Figure 18).

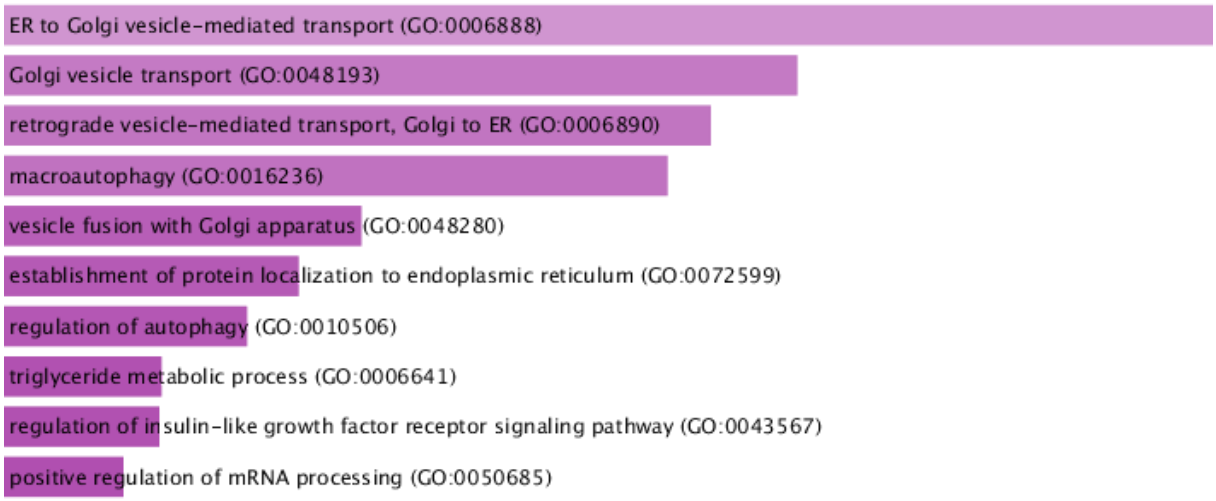


Figure 18: Significant (p-value <0.01) gene ontology clustering by biological process ranked by adjusted p-value. Biological processes are determined and annotated based on differentially expressed genes between the environmental groups, in this case Tyrolerfjord and Hochsetter.

No GO terms were significant at the adjusted p-value < 0.1 level for this location. The highest scoring terms (combined score = log(p-value from Fisher exact test)*Z score of the deviation from the expected rank, highest scoring set as combined score > 10) covered 10 GO terms divided into 5 clusters (consolidated to 4 clusters for visualization, Figure 19). These consolidated clusters can be summarized as: signaling in response to stress and growth factors, endoplasmic reticulum transport and localization, regulation of oligodendrocyte progenitor proliferation, positive regulation of mRNA splicing.

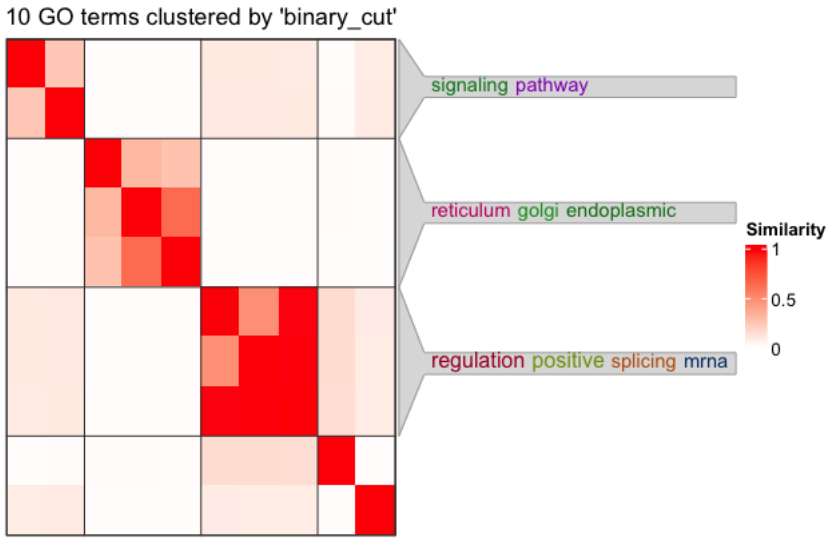


Figure 19: Heatmap of similarity between GO term clusters of high scoring terms from FishEnrichr analysis, combined score > 10. Labeled with term keywords by cluster. N terms = 10, n clusters = 4.

DISCUSSION

This study chronicles the relationship between populations inhabiting different environments. Through the application of an mRNA-seq transcriptome sequencing approach, this study identified genes being differentially expressed between fjord and offshore populations of *B. saida.* These genes map to specific biological pathways regulating stress responses (MAPK inactivation, methylmercury response, stimuli signal transduction) as well as more general pathways associated with homeostasis and cellular maintenance (chromosome segregation, cytoplasmic translation, vesicle mediated transport). The relationship between fjord and offshore populations is revealed in findings at three levels: differentially expressed genes (DEGs) based on normalized reads from full transcriptome sequencing, principal component analysis (PCA) based on the top 500 DEGs, and GO clustering based on all DEGs. These findings are discussed in the context of differences between fjord and offshore populations as well as in relation to environmental drivers. Future opportunities for additional research based on these results are also presented.

Relative differential gene expression (DGE) between neighboring Greenland fjord and offshore populations of *B. saida*

Significant up and down regulation of thousands of genes in each environment pair points to differences in gene expression between fjord and offshore populations. This finding allows us to reject the null hypothesis; that there is no difference in relative gene expression between adjacent fjord and shelf populations of *B. saida*. This differentiation appears to be correlated with spatial distance as more separated locations exhibit more distant clustering based on the significantly differentially expressed genes (DEGs). As geographical separation increases from south to north, the groupings of DEGs based on gene ontology also becomes more disparate. Groupings within fjord and offshore environments (across all locations) also showed a similar signal driven by distance. This finding demonstrates that relationships between fjord and offshore populations are characterized differently by local adaptation in the form of gene expression, rather than by purely genetic differences. For example, Madsen et al. (2016) found via microsatellites that fjord populations, even at great distance, exhibited a closer genetic relationship than adjacent fjord and offshore pairings, while the transcriptomic findings presented here show geographical distance to be a strong predictor of differential gene expression, regardless of environment type. The population structure identified via neutral genetic markers unsurprisingly differs from that identified here via full transcriptome

sequencing. This alternative characterization includes the Tyrolerfjord and Hochsetter pairing, the most spatially proximate in both studies. This pairing grouped the most tightly based on transcriptomic data (not significantly different in PCA or GO analysis), but was among the most distant based on microsatellites. The connection between genotype, expression profile at a certain timepoint, and environmental characteristics is too broad and nuanced for a single study to reveal. Mechanisms of adaptation fall on a spectrum including genomic sequences (such as those included in microsatellite assays and SNP analysis), physical (chromosome rearrangement), epigenetic modifications (nucleotide methylation), and actual gene expression (as represented by the transcriptome) (Fargeot et al., 2021; Grenier et al., 2016; Schneider & Grosschedl, 2007). Compared to other vertebrates, fish show very plastic response to the environmental stimuli, often causing individual studies targeting a single element of adaptation to produce apparently conflicting results. Population characterizations based on microsatellites (Madsen et al., 2016; Maes et al., 2021), genomic architecture (Barth et al., 2017), or transcriptome-derived gene expression (this study) may appear contradicting, but by selecting elements from each as new information becomes available, a summative representation of reality can be achieved. Only by combining multiple approaches can we "unite disparate information sources together to identify patterns that would be invisible when investigated through the lens of a single method" (Jones et al., 2013). By elucidating specific biological pathways driving DGE between populations, the transcriptomic dataset presented herein provides a valuable supplemental characterization and foundational context to other investigations, both past and future.

Studies covering large geographic regions have shown that species with long range dispersal potential often exhibit genetic homogeneity and a lack of regional genetic structuring despite well described basin-scale ecotypes (Barth et al., 2017; Maes et al., 2021). The results of this study point to regulation of gene expression as a putative explanatory factor of observed adaptation and phenotypic plasticity. However, merely categorizing habitats as fjord versus offshore likely does not adequately characterize the environmental experience with respect to the transcriptomic response of a fish at these locations. The transcriptomic findings of this study support the idea that regional differentiation in *B. saida* is driven by expression, though additional analyses of this dataset including SNP profiling would be required to eliminate (or confirm) genetic population structure as a contributing factor.

Connecting physical characteristics of Greenland fjord/shelf environments to differentially expressed biological pathways

A major weakness of prior genetic studies addressing the population structure of *B. saida*, and many other non-model organisms, is the lack of connection between proposed population structure and biology. While useful in identifying population level traits such as reproductive isolation and areas of high or low gene flow, basic genetic tools are restricted in their ability to characterize and compare adaptation under specific circumstances. Most genetic modifications take place over many generations and are not responsive to environmental conditions over an organism's lifetime. Even single nucleotide polymorphism (SNP) based approaches struggle to link individual gene mutations to biological functions controlled by many genes (Berg & Coop, 2014). These biological mechanisms linking key traits to divergence is often revealed through DGE via full transcriptome sequencing (Oomen & Hutchings, 2017). The transcriptomic approach was selected in order to describe the biological processes underlying any observed differences between populations at the gene family level. These new comparisons highlight active avenues of adaptation unrevealed by traditional genetic approaches.

The differences in expression in the liver tissue of *B. saida* characterized in this study provides evidence that the environment fundamentally alters basic biological processes such as homeostasis, stress response, and protein synthesis. The linkages discussed here provide evidence for rejecting the second null hypothesis; there is no connection between the physical environment and pathways being differentially expressed. The mapping of reads to DEGs associated with innate immunity, protein synthesis, and cellular transport, matches the findings of other studies of cold adapted Antarctic fish (Bilyk & Cheng, 2013; Z. Chen et al., 2008). Stress response and habitat specific stressors are of particular interest in understanding resilience in the face of future climate change and human impacts. Previous studies have shown that the liver acts as a stress response initiator in *B. saida* when exposed to toxins, as well as more chronic stressors such as heat (Andersen, Frantzen, et al., 2015). Legacy pollutants, such as organochlorine pesticides (OCPs) including DDT, have been identified in the muscle and liver tissue of *B. saida*, with greater concentrations observed in fjord specimens (Spataro et al., 2021). This is a specific example of a chemical stressor being present at different levels in fjord versus offshore environments, an occurrence that could contribute to the differential gene expression observed in this study. Additionally, the isolation of these stress-associated terms in a single fjord-offshore pairing suggests diversity within fjord and offshore environments.

Beyond specific stressors, abiotic characteristics of different habitats could drive differential regulation of even basal biological functions. Fjord and offshore shelf environments are distinguishable by a number of physical characteristics such as bathymetry, temperature, sedimentation rates, and chemistry (Fredriksen et al., 2020). Within the study area, encompassing the Greenland Sea and fjords of both eastern Greenland and Svalbard, there is a great deal of physical diversity within these habitat categories. The fjords included in this study represent a gradient of sub-Arctic and Arctic classified fjords, with varying degrees of glacial influence, seasonal Atlantic seawater exchange, and ice cover. The fjords sampled for this study, exhibit temperature differences of up to 1.5°C from surface to bottom and likely even larger differences at the sill-Atlantic exchange, especially on Svalbard where the mouth of Isfjord is influenced by the warmer West Spitsbergen Current (Boone et al., 2017; Nilsen et al., 2008; Zoller, 2020). In this way, sites having more disparate gene ontology profiles at longer distance from one another can be seen as a function of environmental differences along a gradient rather than reproductive isolation. Conductivity, temperature, and depth (CTD) readings from the sites themselves confirm that environmental diversity exists within the grouped fjord and offshore sites. The capture depth of specimens within habitat types varies by more than 100m with the associated water temperature differing by 1°C or more. Even temperate fish species (not cold-adapted) have exhibited temperature sensitivity to changes less than 0.5°C and, in some cases, 0.05°C (Bardach & Bjorklund, 1957). It is unreasonable to expect uniform expression patterns in categorical groups as broad as "fjord" and "offshore" even if there was observed genetic structuring, as suggested by Madsen et al. (2016), who suggested fjord-offshore genetic structuring could be caused by post-glacial recolonization and subsequent adaptive divergence driven by the different environments.

Gene expression modification can play an important role as a driver of divergence in addition to, or in place of, actual mutation of the gene-encoding DNA sequence. By studying organisms that have undergone significant divergence, such as humans and mice, characteristics of a gene that restrict or promote change over time can be identified. This approach has revealed that genes with high levels of expression and low tissue specificity tend to have their sequences and expression level highly conserved (Liao & Zhang, 2006). Genes identified in *B. saida* governing processes such as cellular transport and chromosomal segregation are assumed to be highly conserved and widely expressed across tissues due to their fundamental nature. The direct relationship between sequence conservation and gene expression profile remains a subject of debate, with studies of orthologs and duplicated genes in a variety of organisms yielding conflicting results (Jordan et al., 2005; Liao & Zhang, 2006;

Wagner, 2000). A study of recently diverged bird species (<1 million years) also found that genes with highly variable expression patterns tended to correspond to more specialized extracellular components or were unannotated (Uebbing et al., 2016). These observations from other species provide context to the observed differential expression of genes associated with both fundamental biological processes and more specific stress responses in *B. saida.* Differential expression in *B. saida* of highly conserved genes suggests that regulation of even the genes governing basic, highly conserved biological processes may be playing a role in adaptation. It is logical that adjusting gene expression in response to a new environment would be a "safer" path toward an adaptive advantage, and less subject to purifying selection (the removal of deleterious variants), than gene sequence modification (DNA mutation). Changes to protein coding DNA sequences are irreversible over the lifetime of an organism and are more often deleterious than advantageous (there are more ways to "break" a protein than improve it). In this way, most natural selection acts to remove harmful mutations (negative selection) rather than maintain positive changes (positive or Darwinian selection) (Jordan et al., 2005). By differentially regulating core biological processes, *B. saida* exhibits adaptation at the gene expression profile level to a new environmental regime.

One of the major threats of climate change in marine systems is the rate of environmental change exceeding the rate of natural selection for well-adapted traits. Phenotypic plasticity and adaptation over an organism's lifetime via up or down-regulating specific genes may provide additional resilience in a rapidly changing environment. Understanding the expression-derived adaptive response of different groups of *B. saida* could provide an indicator of potential resilience in the face of emerging climate stresses. The inclusion of GO terms corresponding to a response to methylmercury and regulation of the Mitogen-Activated Protein Kinase (MAPK) immune pathway suggest that fjord and offshore groups are encountering different pathological and chemical environmental stressors. The MAPK cascade is especially interesting, as it is preserved across vertebrates and integral in translating stimuli into biological responses, especially with regards to immune system function, but remains understudied in fish (Wei et al., 2020). This novel identification of specific stress pathways in wild *B. saida* provides a specific avenue for future comparisons of latent resilience between populations with yet to be identified genetic structuring.

Future opportunities

The transcriptomic approach employed in this study provides a new perspective on the population structure and environmental relationship of an ecologically significant cryopelagic Arctic gadid. However, there are many gaps remaining in the foundational knowledge required to manage this species, and others like it, in a changing climate. There are several specific improvements upon the approach presented here that could make subsequent studies more productive. These optimizations can be broken into three main categories: method diversity, sample diversity and improved reference materials.

Gene expression, genotype, phenotype, and epigenetic/post-transcriptional modifications such as DNA methylation and miRNA activity are inextricably connected to adaptation in an organism. While the characterization of gene expression based on environment type is enormously useful in pointing to the aspects of an organism that are being differentially influenced by environmental factors, it does not explain how this influence is occurring. The characterization of DNA methylation state could provide the same vital mechanistic context to a transcription study that gene expression provides to a genetic study. This three factor inter-relation between genotype (SNP data), gene expression (RNA-seq), and epigenetic modification (DNA methylation) drives how an organism interacts with and adapts to its surroundings over its lifetime (Jones et al., 2013). In this way, an epigenetic study on the same fjord and offshore ecotypes of *B. saida* could reveal the final aspects of the relationship between these groups currently untouched upon by genetic and transcriptomic research approaches.

Transcription is, biologically, a snapshot of molecular activity representative only of the time, tissue, condition, location, and individual from which it is sampled. By combining and normalizing over dozens of individual "snapshots", this project sought to create a larger image of the relationship between different habitats. In order to improve our understanding at a system level across the full spectrum of habitats and timepoints, many more snapshots and composite images are required. Applying this approach to other tissues and locations is a logical next step in shading in some of the details of the gradient of expression that has already begun to emerge from the three pairs of fjord and offshore stations in this study. Collecting samples for RNA extraction has become a regular part of the TUNU cruise protocol, providing a time series and deep backlog of samples with rich metadata and information on physical environmental characteristics. The wide availability of safe preservatives like *RNAlater* and growing interest in RNA sequencing will hopefully lead to samples of this nature being regularly collected on research cruises throughout the Arctic. Though any tissue would provide useful supplementary

comparisons for this study, those related to immune function such as head kidney and spleen may provide unique insight into the differential stressors faced by fish from different habitats. The TUNU program has also been collecting blood from these fish and preserving it in *RNAlater*, though at the time of writing no reliable method exists for the extraction of high-quality RNA from these samples. Despite not being the lowest hanging fruit, characterizing blood expression profiles would be undoubtedly novel and fascinating research that would, after RNA extraction, be able to follow the same blueprint as this study.

The real work of a full transcriptome approach begins after the sequencing is completed. Selecting and aligning to a reference genome is the first major analytical decision in the bioinformatic process. The study of *B. saida* is substantially limited by the lack of a high-quality reference genome. Though the well-studied *G. morhua* provides a solid entry point and made this study possible, a more specifically annotated genome could enable the identification of more nuanced pathways and biological processes under differential regulation. Directly related is the limited number of suitable databases for non-model organisms. The process of converting reads to genes and genes to biological pathways requires multiple high-quality intermediary references, without which large numbers of genes and reads remain uncharacterized as was the case in this study. Improved genomic and database resources could dramatically improve the quality, quantity, and confidence of results from future studies following this transcriptomic model.

CONCLUSION

The findings of this study support the alternative hypotheses that there are significant differences in gene expression and these differences cluster around basal biological processes between fjord and offshore populations of *B. saida* in the Greenland Sea and adjacent fjords. Furthermore, this study validates full transcriptome sequencing as an approach to understanding local adaptation in populations of a non-model organism. Follow up investigation is warranted to broaden the characterization of these populations using additional tissues and locations. An in-depth characterization of other adaptation avenues such as epigenetic modification would also provide valuable context to the transcriptomic findings presented here. Future gene expression research can use the pipeline developed for this study with minimal adjustments, streamlining knowledge development around a key species and cutting-edge technique.

Acknowledgements

This project has pushed my abilities to the limit and would certainly not have been possible without the support of a network of amazing humans that stretches from Tromsøya all the way to California.

First and foremost, my parents, who have always encouraged me to choose fun and pursue the next adventure without hurry or worry, no matter how far from home it takes me. I owe so much of my love for science and nature to you both.

I am exceedingly grateful for the many hours my advisors Kim, Filipe, and Shripathi have put into helping me design and carry out a master's project that challenged and excited me. You've allowed me to create something I am truly proud of and imparted invaluable knowledge along the way.

I was lucky to have several professors that went above and beyond to help me develop ideas that would end up structuring major sections of my project. Einar Nilssen, Jørgen Christiansen, Roland Sauter, Laura Jaakola, and Mette Svenning, among others, shared a passion for their subject areas that inspired me to bring new energy to my own investigation. Thank you for being so selfless with your time and enthusiastic about the possibilities of scientific research.

The process of writing a master's thesis is not without its challenges, but having my RGG colleagues to share the weekly highs and lows with made it much more enjoyable. Julie, Kamilla, Ingrid, Almu, Elmedina, Rebecca, Meghan, Marta, Emma, Stine, Yngvild, Stian, Julien, Gled, Nathan, and many more: your advice, encouragement, diversion, and humor kept things in perspective even when the finish line seemed out of reach.

A final thank you to Corine, my endless supporter and statistically significant other! I could not have asked for a better co-pilot, be it rocket ship or struggle bus, through the wild journey that the last two years have been. You've made quarantine fun, every meal a feast, and skiing look way too easy. From lab to summit, your determination and love have inspired me to be my best, most authentic self, and for that, as well as a million other things, I thank you.

References

Anders, S., Pyl, P. T., & Huber, W. (2014). HTSeq – A Python framework to work with high-throughput sequencing data. *BioRxiv*, 002824. https://doi.org/10.1101/002824

Andersen, Ø., Frantzen, M., Rosland, M., Timmerhaus, G., Skugor, A., & Krasnov, A. (2015). Effects of crude oil exposure and elevated temperature on the liver transcriptome of polar cod (*Boreogadus saida*). *Aquatic Toxicology*, *165*, 9–18. https://doi.org/10.1016/j.aquatox.2015.04.023

Andersen, Ø., Johnsen, H., De Rosa, M. C., Præbel, K., Stjelja, S., Kirubakaran, T. G., Pirolli, D., Jentoft, S., & Fevolden, S.-E. (2015). Evolutionary history and adaptive significance of the polymorphic Pan I in migratory and stationary populations of Atlantic cod (*Gadus morhua*). *Marine Genomics*, *22*, 45–54. https://doi.org/10.1016/j.margen.2015.03.009

Andrews, S. (2010). *Babraham Bioinformatics—FastQC A Quality Control tool for High Throughput Sequence Data*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bardach, J. E., & Bjorklund, R. G. (1957). The Temperature Sensitivity of Some American Freshwater Fishes. *The American Naturalist*, *91*(859), 233–251.

Barth, J. M. I., Berg, P. R., Jonsson, P. R., Bonanomi, S., Corell, H., Hemmer-Hansen, J., Jakobsen, K. S., Johannesson, K., Jorde, P. E., Knutsen, H., Moksnes, P.-O., Star, B., Stenseth, N. C., Svedäng, H., Jentoft, S., & André, C. (2017). Genome architecture enables local adaptation of Atlantic cod despite high connectivity. *Molecular Ecology*, *26*(17), 4452–4466. https://doi.org/10.1111/mec.14207

Berg, J. J., & Coop, G. (2014). A Population Genetic Signal of Polygenic Adaptation. *PLOS Genetics*, *10*(8), e1004412. https://doi.org/10.1371/journal.pgen.1004412

Bilyk, K. T., & Cheng, C.-H. C. (2013). Model of gene expression in extreme cold—Reference transcriptome for the high-Antarctic cryopelagic notothenioid fish *Pagothenia borchgrevinki*. *BMC Genomics*, *14*(1), 634. https://doi.org/10.1186/1471-2164-14-634

Bluhm, B. A., & Gradinger, R. (2008). Regional Variability in Food Availability for Arctic Marine Mammals. *Ecological Applications*, *18*(sp2), S77–S96. https://doi.org/10.1890/06-0562.1

Boone, W., Rysgaard, S., Kirillov, S., Dmitrenko, I., Bendtsen, J., Mortensen, J., Meire, L., Petrusevich, V., & Barber, D. G. (2017). Circulation and fjord-shelf exchange during the ice-covered period in Young Sound-Tyrolerfjord, Northeast Greenland (74°N). *Estuarine, Coastal and Shelf Science*, *194*. https://doi.org/10.1016/j.ecss.2017.06.021

Bradstreet, M. S. W., & Cross, W. E. (1982). Trophic Relationships at High Arctic Ice Edges. *ARCTIC*, *35*(1), 1–12. https://doi.org/10.14430/arctic2303

Castonguay, M., Plourde, S., Robert, D., Runge, J. A., & Fortier, L. (2008). Copepod production drives recruitment in a marine fish. *Canadian Journal of Fisheries and Aquatic Sciences*, *65*(8), 1528–1531. https://doi.org/10.1139/F08-126

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*, *14*, 128. https://doi.org/10.1186/1471-2105-14-128

Chen, Z., Cheng, C.-H. C., Zhang, J., Cao, L., Chen, L., Zhou, L., Jin, Y., Ye, H., Deng, C., Dai, Z., Xu, Q., Hu, P., Sun, S., Shen, Y., & Chen, L. (2008). Transcriptomic and genomic evolution under constant cold in Antarctic notothenioid fish. *Proceedings of the National Academy of Sciences*, *105*(35), 12944–12949. https://doi.org/10.1073/pnas.0802432105

Christensen, O. B., Drews, M., Christensen, J. H., Dethloff, K., Ketelsen, K., Hebestadt, I., & Rinke, A. (2007). *The HIRHAM Regional Climate Model. Version 5 (beta)*. https://orbit.dtu.dk/en/publications/the-hirham-regional-climate-model-version-5-beta

Corley, S. M., Troy, N. M., Bosco, A., & Wilkins, M. R. (2019). QuantSeq. 3′ Sequencing combined with Salmon provides a fast, reliable approach for high throughput RNA expression analysis. *Scientific Reports*, *9*(1), 18895. https://doi.org/10.1038/s41598-019-55434-x

Drost, H. E., Lo, M., Carmack, E. C., & Farrell, A. P. (2016). Acclimation potential of Arctic cod (*Boreogadus saida*) from the rapidly warming Arctic Ocean. *The Journal of Experimental Biology*, *219*(19), 3114–3125. https://doi.org/10.1242/jeb.140194

Eldem, V., Zararsiz, G., Taşçi, T., Duru, I. P., Bakir, Y., & Erkan, M. (2017). Transcriptome Analysis for Non-Model Organism: Current Status and Best-Practices. In *Applications of RNA-Seq and Omics Strategies—From Microorganisms to Human Health*. IntechOpen. https://doi.org/10.5772/intechopen.68983

Eric Sayers. (2018). The E-utilities In-Depth: Parameters, Syntax and More. In *Entrez Programming Utilities Help [Internet]*. National Center for Biotechnology Information (US). https://www.ncbi.nlm.nih.gov/books/NBK25499/

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Falk-Petersen, S., Hopkins, C., & Sargent, J. R. (1990). Trophic relationships in the pelagic, Arctic food web. *Trophic Relationships in the Marine Environment*, 315–333.

Fargeot, L., Loot, G., Prunier, J. G., Rey, O., Veyssière, C., & Blanchet, S. (2021). Patterns of Epigenetic Diversity in Two Sympatric Fish Species: Genetic vs. Environmental Determinants. *Genes*, *12*(1), 107. https://doi.org/10.3390/genes12010107

Fevolden, S. E., & Pogson, G. H. (1997). Genetic divergence at the synaptophysin (Syp I) locus among Norwegian coastal and north-east Arctic populations of Atlantic cod. *Journal of Fish Biology*, *51*(5), 895–908. https://doi.org/10.1111/j.1095-8649.1997.tb01529.x

Fredriksen, R., Christiansen, J. S., Bonsdorff, E., Larsen, L.-H., Nordström, M. C., Zhulay, I., & Bluhm, B. A. (2020). Epibenthic megafauna communities in Northeast Greenland vary across coastal, continental shelf and slope habitats. *Polar Biology*, *43*(10), 1623–1642. https://doi.org/10.1007/s00300-020-02733-z

Froese, R. (2006). Cube law, condition factor and weight-length relationships: History, meta-analysis and recommendations. *Journal of Applied Ichthyology*, *22*(4), 241–253. https://doi.org/10.1111/j.1439-0426.2006.00805.x

Gibbons, J. G., Janson, E. M., Hittinger, C. T., Johnston, M., Abbot, P., & Rokas, A. (2009). Benchmarking Next-Generation Transcriptome Sequencing for Functional and Evolutionary Genomics. *Molecular Biology and Evolution*, *26*(12), 2731–2744. https://doi.org/10.1093/molbev/msp188

Graham, M., & Hop, H. (1995). Aspects of Reproduction and Larval Biology of Arctic Cod (*Boreogadus saida*). *ARCTIC*, *48*(2), 130–135. https://doi.org/10.14430/arctic1234

Grenier, S., Barre, P., & Litrico, I. (2016). Phenotypic Plasticity and Selection: Nonexclusive Mechanisms of Adaptation. *Scientifica*, *2016*, e7021701. https://doi.org/10.1155/2016/7021701

Hedges, K., Macphee, S., Valtysson, H., Johannesen, E., & Mecklenburg, C. (2017). *State of the Arctic Marine Biodiversity Report: Chapter 3.4: Marine fishes* (pp. 109–127).

Hoegh-Guldberg, O., & Bruno, J. F. (2010). The Impact of Climate Change on the World's Marine Ecosystems. *Science*, *328*(5985), 1523–1528. https://doi.org/10.1126/science.1189930

Hop, H., & Gjøsæter, H. (2013). Polar cod (*Boreogadus saida*) and capelin (*Mallotus villosus*) as key species in marine food webs of the Arctic and the Barents Sea. *Marine Biology Research*, *9*(9), 878–894. https://doi.org/10.1080/17451000.2013.775458

Huserbråten, M. B. O., Eriksen, E., Gjøsæter, H., & Vikebø, F. (2019). Polar cod in jeopardy under the retreating Arctic sea ice. *Communications Biology*, *2*(1), 1–8. https://doi.org/10.1038/s42003-019-0649-2

Ishikawa, A., Kusakabe, M., Yoshida, K., Ravinet, M., Makino, T., Toyoda, A., Fujiyama, A., & Kitano, J. (2017). Different contributions of local- and distant-regulatory changes to transcriptome divergence between stickleback ecotypes. *Evolution*, *71*(3), 565–581. https://doi.org/10.1111/evo.13175

Jones, M. J., Fejes, A. P., & Kobor, M. S. (2013). DNA methylation, genotype and gene expression: Who is driving and who is along for the ride? *Genome Biology*, *14*(7), 126. https://doi.org/10.1186/gb-2013-14-7-126

Jordan, I. K., Mariño-Ramírez, L., & Koonin, E. V. (2005). Evolutionary significance of gene expression divergence. *Gene*, *345*(1), 119–126. https://doi.org/10.1016/j.gene.2004.11.034

Karlsen, B. O., Klingan, K., Emblem, Å., Jørgensen, T. E., Jueterbock, A., Furmanek, T., Hoarau, G., Johansen, S. D., Nordeide, J. T., & Moum, T. (2013). Genomic divergence between the migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, *22*(20), 5098–5111. https://doi.org/10.1111/mec.12454

Kohlbach, D., Schaafsma, F. L., Graeve, M., Lebreton, B., Lange, B. A., David, C., Vortkamp, M., & Flores, H. (2017). Strong linkage of polar cod (*Boreogadus saida*) to sea ice algae-produced carbon: Evidence from stomach content, fatty acid and stable isotope analyses. *Progress in Oceanography*, *152*, 62–74. https://doi.org/10.1016/j.pocean.2017.02.003

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., Lachmann, A., McDermott, M. G., Monteiro, C. D., Gundersen, G. W., & Ma'ayan, A. (2016). Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, *44*(W1), W90-97. https://doi.org/10.1093/nar/gkw377

Kvile, K. Ø., Langangen, Ø., Prokopchuk, I., Stenseth, N. C., & Stige, L. C. (2016). Disentangling the mechanisms behind climate effects on zooplankton. *Proceedings of the National Academy of Sciences*, *113*(7), 1841–1846. https://doi.org/10.1073/pnas.1525130113

Landrum, L. & Holland, Marika. (2020). Extremes become routine in an emerging new Arctic. *Nature Climate Change*, 17.

Lawson, J., Magalhães, A., & Miller, E. (1998). Important prey species of marine vertebrate predators in the northwest Atlantic: Proximate composition and energy density. *Marine Ecology Progress Series*, *164*, 13–20. https://doi.org/10.3354/meps164013

Liao, B.-Y., & Zhang, J. (2006). Low Rates of Expression Profile Divergence in Highly Expressed Genes and Tissue-Specific Genes During Mammalian Evolution. *Molecular Biology and Evolution*, *23*(6), 1119–1128. https://doi.org/10.1093/molbev/msj119

Lohman, B. K., Stutz, W. E., & Bolnick, D. I. (2017). Gene expression stasis and plasticity following migration into a foreign environment. *Molecular Ecology*, *26*(18), 4657–4670. https://doi.org/10.1111/mec.14234

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

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. https://doi.org/10.1007/s00300-015-1812-y

Maes, S. M., Christiansen, H., Mark, F. C., Lucassen, M., Putte, A. V. de, Volckaert, F. A. M., & Flores, H. (2021). High gene flow in polar cod (*Boreogadus saida*) from West-Svalbard and the Eurasian Basin. *Journal of Fish Biology*, *n/a*(n/a). https://doi.org/10.1111/jfb.14697

Melnikov, I. A., & Chernova, N. V. (2013). Characteristics of under-ice swarming of polar cod *Boreogadus saida* (*Gadidae*) in the Central Arctic Ocean. *Journal of Ichthyology*, *53*(1), 7–15. https://doi.org/10.1134/S0032945213010086

Moritz, R. E., Bitz, C. M., & Steig, E. J. (2002). Dynamics of Recent Climate Change in the Arctic. *Science*, *297*(5586), 1497–1502. https://doi.org/10.1126/science.1076522

Mueter, F. J., Nahrgang, J., John Nelson, R., & Berge, J. (2016). The ecology of gadid fishes in the circumpolar Arctic with a special emphasis on the polar cod (*Boreogadus saida*). *Polar Biology*, *39*(6), 961–967. https://doi.org/10.1007/s00300-016-1965-3

Müller, B., & Grossniklaus, U. (2010). Model organisms—A historical perspective. *Journal of Proteomics*, *73*(11), 2054–2063. https://doi.org/10.1016/j.jprot.2010.08.002

Nilsen, F., Cottier, F., Skogseth, R., & Mattsson, S. (2008). Fjord–shelf exchanges controlled by ice and brine production: The interannual variation of Atlantic Water in Isfjorden, Svalbard. *Continental Shelf Research*, *28*(14), 1838–1853. https://doi.org/10.1016/j.csr.2008.04.015

Nuwaysir, E. F., Bittner, M., Trent, J., Barrett, J. C., & Afshari, C. A. (1999). Microarrays and toxicology: The advent of toxicogenomics. *Molecular Carcinogenesis*, *24*(3), 153–159. https://doi.org/10.1002/(SICI)1098-2744(199903)24:3<153::AID-MC1>3.0.CO;2-P

Oomen, R. A., & Hutchings, J. A. (2017). Transcriptomic responses to environmental change in fishes: Insights from RNA sequencing. *FACETS*, *2*(2), 610–641. https://doi.org/10.1139/facets-2017-0015

Pálsson, S., Källman, T., Paulsen, J., & Árnason, E. (2009). An assessment of mitochondrial variation in Arctic gadoids. *Polar Biology*, *32*(3), 471–479. https://doi.org/10.1007/s00300-008-0542-9

Palumbi, S. R. (1994). Genetic Divergence, Reproductive Isolation, and Marine Speciation. *Annual Review of Ecology and Systematics*, *25*, 547–572.

Rao, M. S., Van Vleet, T. R., Ciurlionis, R., Buck, W. R., Mittelstadt, S. W., Blomme, E. A. G., & Liguori, M. J. (2019). Comparison of RNA-Seq and Microarray Gene Expression Platforms for the Toxicogenomic Evaluation of Liver From Short-Term Rat Toxicity Studies. *Frontiers in Genetics*, *9*. https://doi.org/10.3389/fgene.2018.00636

Renaud, P. E., Daase, M., Banas, N. S., Gabrielsen, T. M., Søreide, J. E., Varpe, Ø., Cottier, F., Falk-Petersen, S., Halsband, C., Vogedes, D., Heggland, K., & Berge, J. (2018). Pelagic food-webs in a changing Arctic: A trait-based perspective suggests a mode of resilience. *ICES Journal of Marine Science*, *75*(6), 1871–1881. https://doi.org/10.1093/icesjms/fsy063

Rysgaard, S., Vang, T., Stjernholm, M., Rasmussen, B., Windelin, A., & Kiilsholm, S. (2003). Physical Conditions, Carbon Transport, and Climate Change Impacts in a Northeast Greenland Fjord. *Arctic, Antarctic, and Alpine Research*, *35*(3), 301–312. https://doi.org/10.1657/1523-0430(2003)035[0301:PCCTAC]2.0.CO;2

Schneider, R., & Grosschedl, R. (2007). Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes & Development*, *21*(23), 3027–3043. https://doi.org/10.1101/gad.1604607

Spataro, F., Patrolecco, L., Ademollo, N., Præbel, K., Rauseo, J., Pescatore, T., & Corsolini, S. (2021). Multiple exposure of the *Boreogadus saida* from bessel fjord (NE Greenland) to legacy and emerging pollutants. *Chemosphere*, *279*, 130477. https://doi.org/10.1016/j.chemosphere.2021.130477

Striberny, A., Jørgensen, E. H., Klopp, C., & Magnanou, E. (2019). Arctic charr brain transcriptome strongly affected by summer seasonal growth but only subtly by feed deprivation. *BMC Genomics*, *20*(1), 529. https://doi.org/10.1186/s12864-019-5874-z

Uebbing, S., Künstner, A., Mäkinen, H., Backström, N., Bolivar, P., Burri, R., Dutoit, L., Mugal, C. F., Nater, A., Aken, B., Flicek, P., Martin, F. J., Searle, S. M. J., & Ellegren, H. (2016). Divergence in gene expression within and between two closely related flycatcher species. *Molecular Ecology*, *25*(9), 2015–2028. https://doi.org/10.1111/mec.13596

van der Meeren, G., Prozorkevich, D., & Trofimov, A. (2019). *(PDF) Survey report from the joint Norwegian/Russian ecosystem survey in the Barents Sea and adjacent waters, August-October 2018* (IMR/PINRO Joint Report Series, 2-2019). Institute of Marine Research in Norway.

Wagner, A. (2000). Decoupled evolution of coding region and mRNA expression patterns after gene duplication: Implications for the neutralist-selectionist debate. *Proceedings of the National Academy of Sciences*, *97*(12), 6579–6584. https://doi.org/10.1073/pnas.110147097

Wei, X., Zhang, Y., Li, C., Ai, K., Li, K., Li, H., & Yang, J. (2020). The evolutionarily conserved MAPK/Erk signaling promotes ancestral T-cell immunity in fish via c-Myc–mediated glycolysis. *Journal of Biological Chemistry*, *295*(10), 3000–3016. https://doi.org/10.1074/jbc.RA119.012231

Welch, H. E., Crawford, R. E., & Hop, H. (1993). Occurrence of Arctic Cod (*Boreogadus Saida*) Schools and Their Vulnerability to Predation in the Canadian High Arctic. *ARCTIC*, *46*(4), 331–339. https://doi.org/10.14430/arctic1361

Yang, W. E., Woods, C. W., & Tsalik, E. L. (2015). Chapter 13—Host-Based Diagnostics for Detection and Prognosis of Infectious Diseases. In A. Sails & Y.-W. Tang (Eds.), *Methods in Microbiology* (Vol. 42, pp. 465–500). Academic Press. https://doi.org/10.1016/bs.mim.2015.06.001

Zoller, K. M. (2020). Last Glacial Maximum – Holocene palaeoenvironment in Bessel Fjord and southwestern Dove Bugt, Northeast Greenland. https://munin.uit.no/handle/10037/18467

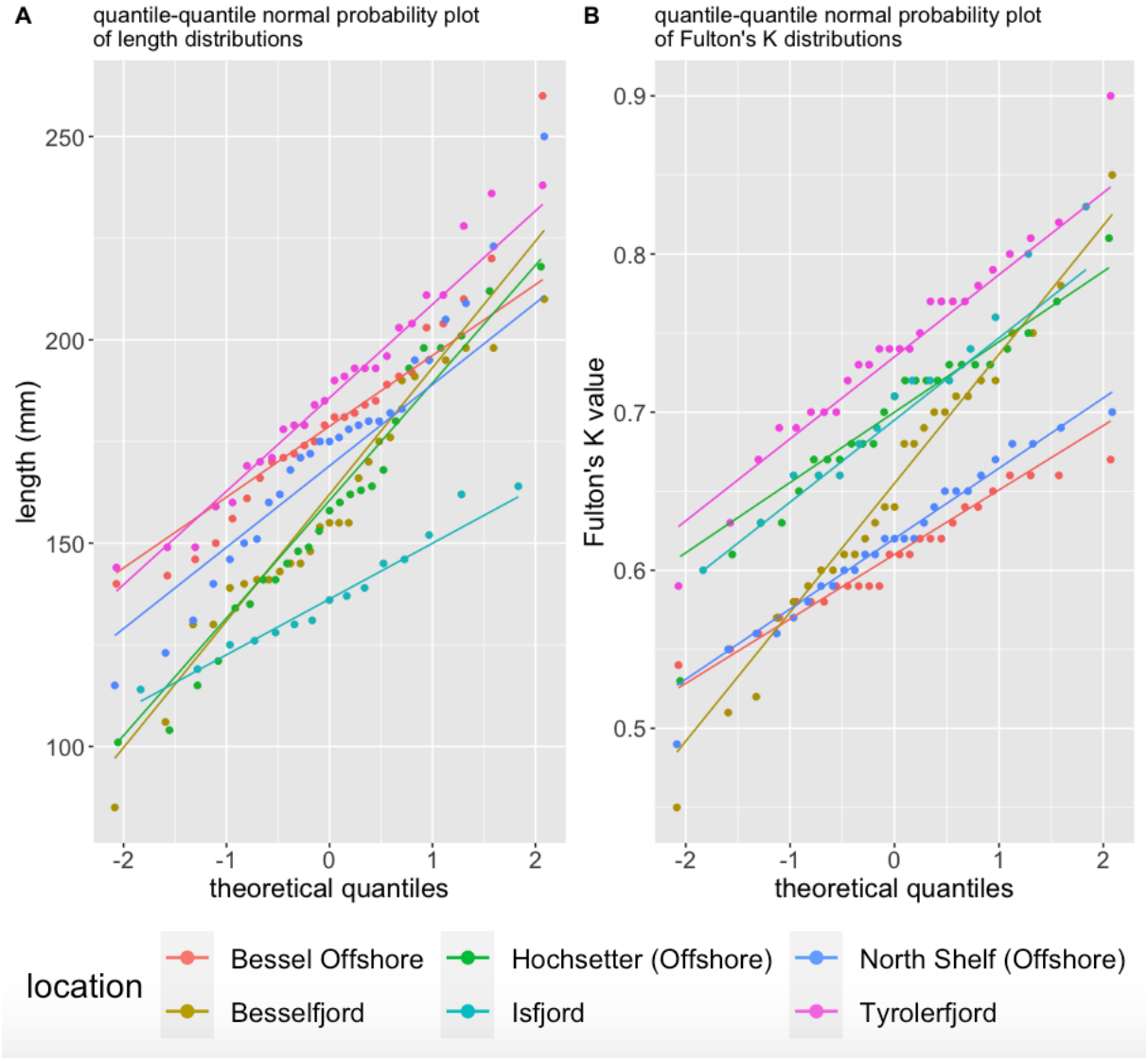
Supplementary plots and materials

Suppl. Table 1: Outlier removal details

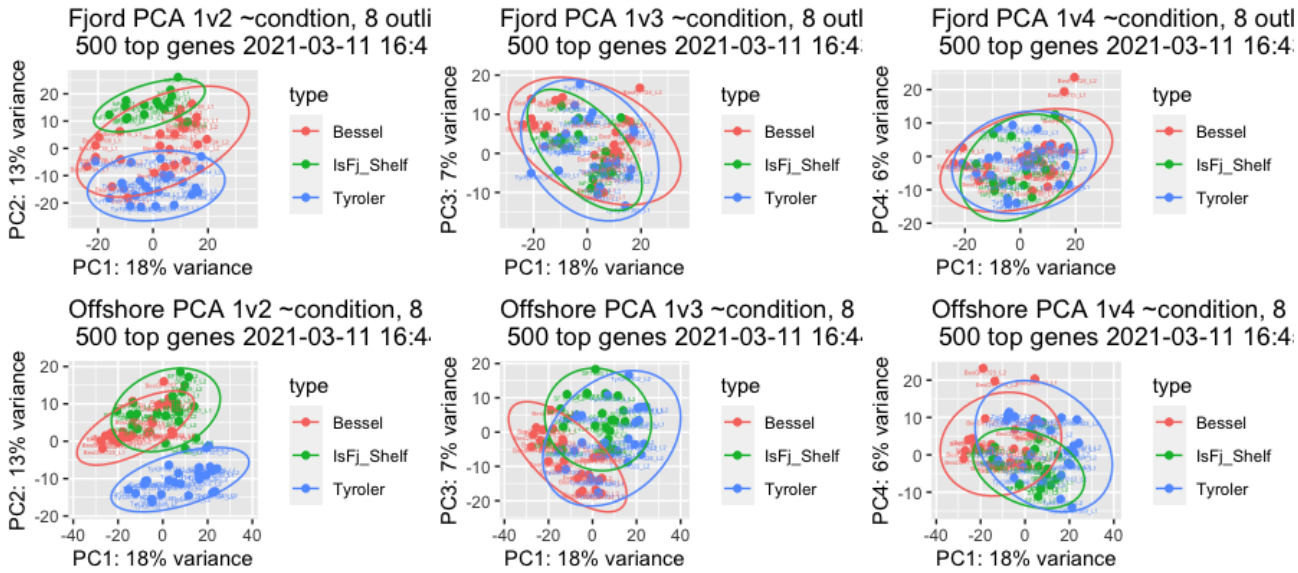
Fish_ID	GH_ID	location	length	weight	FultonK	type	notes
Bs13241	TyrFj1317	Tyroler	191	53.6	0.77	Tyr	Sample to sample dist.
Bs13403	TyrOff1317	Hoch	135	17.8	0.72	Tyr	Low input reads
Bs13406	TyrOffRe1304	Hoch	201	49.5	0.61	Tyr	Sample to sample dist.
Bsa17056	BesFj1714	Besselfjord	148	20.6	0.64	Bess	Sample to sample dist.
Bsa17046	BesFj1716	Besselfjord	130	12.8	0.58	Bess	Sample to sample dist.
Bsa17008	IsFj1708_L2	Svalbard, Isfjord	125	14	0.72	Isfj	Low input reads
Bsa17011	IsFj1711_L2	Svalbard, Isfjord	131	14.8	0.66	Isfj	Low input reads
Bs15242	SF1509	Shelf-N	175	29.9	0.56	Isfj	Sample to sample dist.
Bs15248	SF1515	Shelf-N	151	22.45	0.65	Isfj	Sample to sample dist.
Bs15250	SF1517	Shelf-N	176	37.75	0.69	Isfj	Sample to sample dist.
Bsa17069	BesFj1727	Besselfjord	154	22.2	0.61	Bess	Demux failed, bad index

Suppl. Table 2: Collected linear regression coefficients and p-values of the log10 transformed length and weight model fittings.

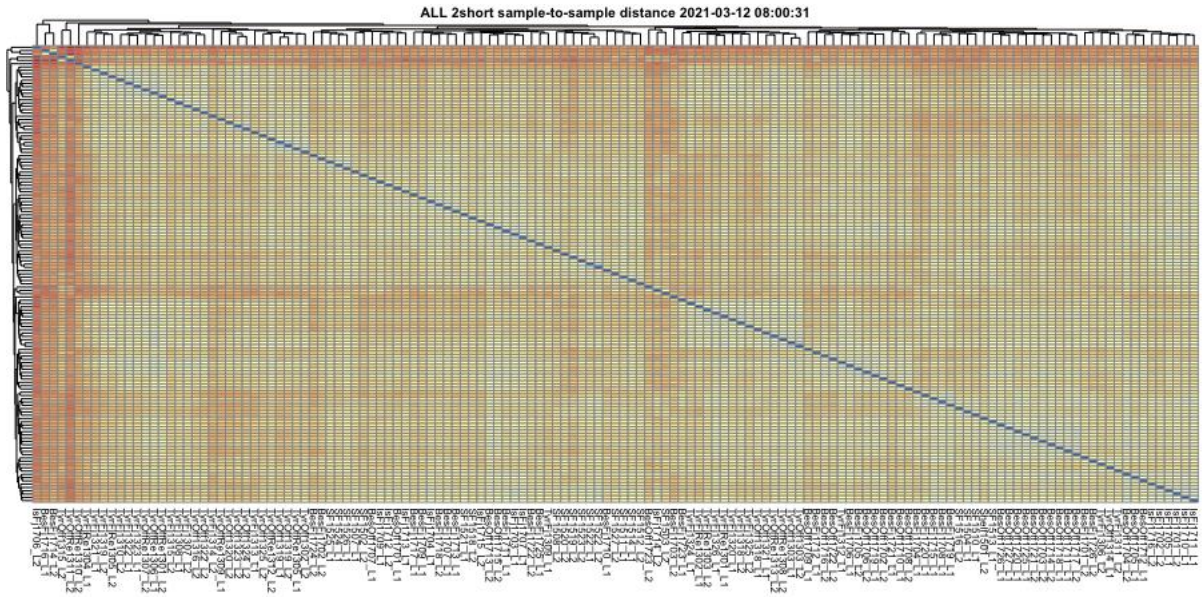
location	n	log_intercept	slope	std_error	r_squared	f_value	Pr_fvalue
Besselfjord	27	2.855	2.855	0.134	0.945	451.34	< .001
Bessel Offshore	26	2.986	2.986	0.088	0.979	1155.523	< .001
Tyrolerfjord	30	3.146	3.146	0.1	0.972	996.968	< .001
Hochsetter (Offshore)	28	2.904	2.904	0.078	0.981	1390.466	< .001
Isfjord	16	2.995	2.995	0.224	0.927	178.276	< .001
North Shelf (Offshore)	31	2.772	2.772	0.072	0.98	1499.421	< .001



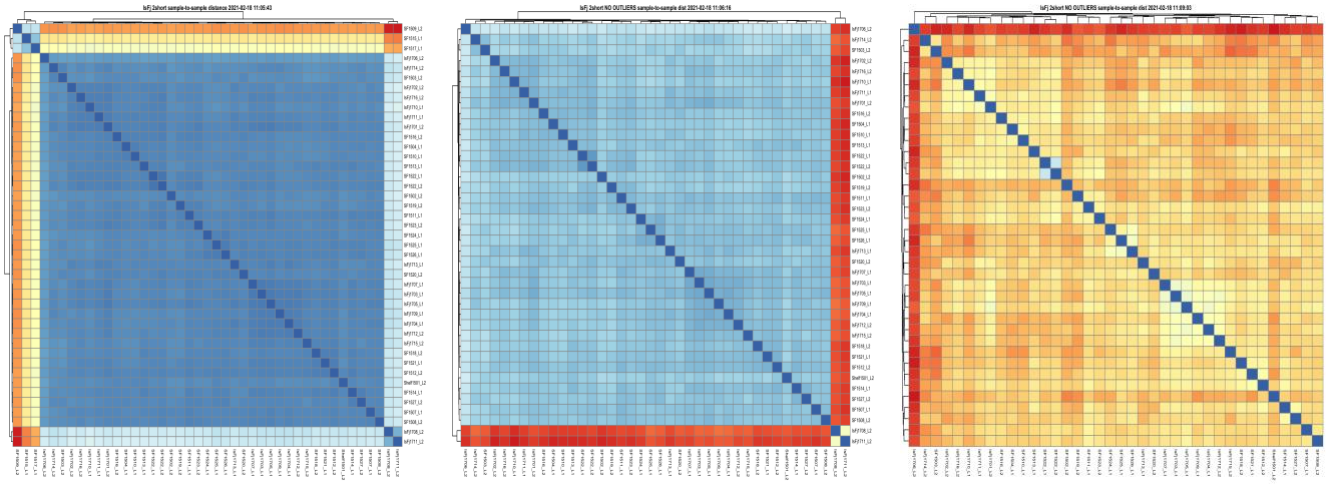
*Suppl. Figure 1: Quantile-quantile normal probability plots for length (**A**) and Fulton's K condition factor (**B**) based on location. Lines of the same or similar slope follow normal distributions of a similar shape. Point distribution represents the shape of the residuals, with a linear trend implying the variances are relatively homogenous.*



Suppl. Figure 2: Additional PCAs of the full environment groups incorporating less explanatory PC3 and PC4.



Suppl. Figure 3: Sample to sample distance for all groups combined, 8 outliers removed.



Suppl. Figure 4: Progressive outlier removal by sample to sample distance of Isfjord and Northern Shelf.

Appendix I: RNA extraction and cDNA library preparation kit protocols

Qiagen Mini Kit RNA extraction protocol:

Protocol: Purification of Total RNA from Animal Tissues

This protocol requires the RNeasy Mini Kit or RNeasy Protect Mini Kit.

Determining the correct amount of starting material

It is essential to use the correct amount of starting material in order to obtain optimal RNA yield and purity. A maximum amount of 30 mg fresh or frozen tissue or 15–20 mg RNAprotect stabilized tissue (which is partially dehydrated) can generally be processed. For most tissues, the RNA binding capacity of the RNeasy spin column and the lysing capacity of Buffer RLT will not be exceeded by these amounts. Average RNA yields from various tissues are given in Table 2 (page 21).

Some tissues, such as spleen, parts of brain, lung, and thymus are more difficult to lyse or tend to form precipitates during RNA purification. The volume of Buffer RLT may need to be increased to facilitate complete homogenization and to avoid significantly reduced RNA yields, DNA contamination, or clogging of the RNeasy spin column. See the procedure below for details.

RNA yields from fibrous tissues, such as skeletal muscle, heart, and skin, may be low due to the abundance of contractile proteins, connective tissue, and collagen. For maximum RNA yields from these tissues, we recommend using the RNeasy Fibrous Tissue Mini Kit instead. See page 91 for ordering information.

Greater RNA yields from fatty tissues, such as brain and adipose tissue, can be achieved using the RNeasy Lipid Tissue Mini Kit, which uses QIAzol Lysis Reagent for optimal tissue lysis. See page 91 for ordering information.

If there is no information about the nature of your starting material, we recommend starting with no more than 10 mg tissue. Depending on RNA yield and purity, it may be possible to use up to 30 mg tissue in subsequent preparations.

Do not overload the RNeasy spin column, as this will significantly reduce RNA yield and quality.

Weighing tissue is the most accurate way to quantitate the amount of starting material. As a guide, a 3 mm cube (27 mm3) of most animal tissues weighs 30–35 mg.

Important points before starting

- If using the RNeasy Kit for the first time, read "Important Notes" (page 19). If working with RNA for the first time, read Appendix A (page 73).
- For optimal results, stabilize harvested tissues immediately in RNAprotect Tissue Reagent (see protocol on page 41). Tissues can be stored in the reagent for up to 1 day at 37°C, 7 days at 15–25°C, or 4 weeks at 2–8°C, or archived at −30 to −15°C or −90 to −65°C.
- Fresh, frozen, or RNAprotect stabilized tissues can be used. Tissues can be stored at −90 to −65°C for several months. Flash-freeze tissues in liquid nitrogen, and immediately transfer to −90 to −65°C. Do not allow tissues to thaw during weighing or handling prior to disruption in Buffer RLT. Homogenized tissue lysates from step 4 can also be stored at −90 to −65°C for several months.

Incubate frozen lysates at 37°C in a water bath until completely thawed and salts are dissolved before continuing with step 5. Avoid prolonged incubation, which may compromise RNA integrity.

- λ If desired, more than 30 mg tissue can be disrupted and homogenized at the start of the procedure (increase the volume of Buffer RLT proportionately). Use a portion of the homogenate corresponding to no more than 30 mg tissue for RNA purification, and store the rest at −90 to −65°C.
- λ Buffer RLT may form a precipitate upon storage. If necessary, redissolve by warming, and then place at room temperature.

- λ Buffer RLT and Buffer RW1 contain a guanidine salt and are therefore not compatible with disinfecting reagents containing bleach. See page 7 for safety information.
- λ Perform all steps of the procedure at room temperature. During the procedure, work quickly.
- λ Perform all centrifugation steps at 20–25°C in a standard microcentrifuge. Ensure that the centrifuge does not cool below 20°C.

Things to do before starting

- λ β-Mercaptoethanol (β-ME) must be added to Buffer RLT before use. Add 10 µl β-ME per 1 ml Buffer RLT. Dispense in a fume hood and wear appropriate protective clothing. Buffer RLT containing β-ME can be stored at room temperature for up to 1 month.
- λ Alternatively, add 20 µl of 2 M dithiothreitol (DTT) per 1 ml Buffer RLT. The stock solution of 2 M DTT in water should be prepared fresh or frozen in single-use aliquots. Buffer RLT containing DTT can be stored at room temperature for up to 1 month.
- λ Buffer RPE is supplied as a concentrate. Before using for the first time, add 4 volumes of ethanol (96–100%) as indicated on the bottle to obtain a working solution.
- λ If performing optional on-column DNase digestion, prepare DNase I stock solution as described in Appendix D (page 82).

Procedure

1. Excise the tissue sample from the animal or remove it from storage. Remove RNAprotect stabilized tissues from the reagent using forceps. Determine the amount of tissue. Do not use more than 30 mg.

Weighing tissue is the most accurate way to determine the amount.

Note: If the tissues were stored in RNAprotect Reagent at −30 to −15°C, be sure to remove any crystals that may have formed.

2. Follow either step 2a or 2b.

1. 2a. For RNAprotect stabilized tissues:
   If using the entire tissue, place it directly into a suitably sized vessel for disruption and homogenization, and proceed to step 3.
   If using only a portion of the tissue, cut it on a clean surface. Weigh the piece to be used, and place it into a suitably sized vessel for disruption and homogenization. Proceed to step 3.
   RNA in RNAprotect stabilized tissues is protected during cutting and weighing of tissues at ambient temperature (15–25°C). It is not necessary to cut the tissues on ice or dry ice or in a refrigerated room. Remaining tissues can be stored in RNAprotect Tissue Reagent. Previously stabilized tissues can be stored at −90 to −65°C without the reagent.
2. 2b. For unstabilized fresh or frozen tissues:
   If using the entire tissue, place it directly into a suitably sized vessel for disruption and homogenization, and proceed immediately to step 3.
   If using only a portion of the tissue, weigh the piece to be used, and place it into a suitably sized vessel for disruption and homogenization. Proceed immediately to step 3.
   RNA in harvested tissues is not protected until the tissues are treated with RNAprotect Tissue Reagent, flash-frozen, or disrupted and homogenized in step 3. Frozen tissues should not be allowed to thaw during handling. The relevant procedures should be carried out as quickly as

possible.
Note: Remaining fresh tissues can be placed into RNAprotect Tissue Reagent to stabilize RNA (see protocol on page 36). However, previously frozen tissues thaw too slowly in the reagent, preventing the reagent from diffusing into the tissues quickly enough to prevent RNA degradation.

3. Disrupt the tissue and homogenize the lysate in Buffer RLT (do not use more than 30 mg tissue) according to step 3a, 3b, 3c, or 3d.

See "Disrupting and homogenizing starting material", page 22, for more details on disruption and homogenization.

Note: Ensure that β-ME is added to Buffer RLT before use (see "Things to do before starting").

After storage in RNAprotect Tissue Reagent, tissues may become slightly harder than fresh or thawed tissues. Disruption and homogenization using standard methods is usually not a problem. For easier disruption and homogenization, we recommend using 600 µl Buffer RLT.

Note: Incomplete homogenization leads to significantly reduced RNA yields and can cause clogging of the RNeasy spin column. Homogenization with the TissueLyser and rotor–stator homogenizers generally results in higher RNA yields than with other methods.

Table 8. Volumes of Buffer RLT for Tissue Disruption and Homogenization

* Use 600 µl Buffer RLT for tissues stabilized in RNAprotect Tissue Reagent or for difficult-to-lyse tissues.

Table 8. Volumes of Buffer RLT for Tissue Disruption and Homogenization

Amount of starting material (mg) | Volume of Buffer RLT (µl)

* Use 600 µl Buffer RLT for tissues stabilized in RNAprotect Tissue Reagent or for difficult-to-lyse tissues.

<20 | 350 or 600*

20–30 | 600

3a. Disruption and homogenization using a rotor–stator homogenizer:
Place the weighed (fresh, frozen, or RNAprotect stabilized) tissue in a suitably sized vessel. Add the appropriate volume of Buffer RLT (see Table 8). Immediately disrupt and homogenize the tissue using a conventional rotor–stator homogenizer until it is uniformly homogeneous (usually 20–40 s). Proceed to step 4.

3b. Disruption using a mortar and pestle followed by homogenization using a QIAshredder homogenizer:
Immediately place the weighed (fresh, frozen, or RNAprotect stabilized) tissue in liquid nitrogen, and grind thoroughly with a mortar and pestle. Decant tissue powder and liquid nitrogen into an RNAse-free, liquid-nitrogen–cooled, 2 ml microcentrifuge tube (not supplied). Allow the liquid nitrogen to evaporate, but do not allow the tissue to thaw.

Add the appropriate volume of Buffer RLT (see Table 8). Pipet the lysate directly into a QIAshredder spin column placed in a 2 ml collection tube, and centrifuge for 2 min at full speed. Proceed to step 4.

3c. Disruption using a mortar and pestle followed by homogenization using a needle and syringe:

Immediately place the weighed (fresh, frozen, or RNAprotect stabilized) tissue in liquid nitrogen, and grind thoroughly with a mortar and pestle. Decant tissue powder and liquid nitrogen into an RNAse-free, liquid-nitrogen–cooled, 2 ml microcentrifuge tube (not supplied). Allow the liquid nitrogen to evaporate, but do not allow the tissue to thaw.

Add the appropriate volume of Buffer RLT (see Table 8), and homogenize by passing the lysate at least 5 times through a blunt 20-gauge needle fitted to an RNAse-free syringe. Proceed to step 4.

3d. Disruption and homogenization using the TissueLyser: See the TissueLyser Handbook. Then proceed to step 4.

4. Centrifuge the lysate for 3 min at full speed. Carefully remove the supernatant by pipetting, and transfer it to a new microcentrifuge tube (not supplied). Use only this supernatant (lysate) in subsequent steps.

In some preparations, very small amounts of insoluble material will be present after the 3 min centrifugation, making the pellet invisible.

5. Add 1 volume of 70% ethanol* to the cleared lysate, and mix immediately by pipetting. Do not centrifuge. Proceed immediately to step 6.

   Note: The volume of lysate may be less than 350 µl or 600 µl due to loss during homogenization and centrifugation in steps 3 and 4.

   Note: Precipitates may be visible after addition of ethanol. This does not affect the procedure.

6. Transfer up to 700 µl of the sample, including any precipitate that may have formed, to an RNeasy spin column placed in a 2 ml collection tube (supplied). Close the lid gently, and centrifuge for 15 s at ≥8000 x g (≥10,000 rpm). Discard the flow-through.†

   Reuse the collection tube in step 7.

   If the sample volume exceeds 700 µl, centrifuge successive aliquots in the same RNeasy spin column. Discard the flow-through after each centrifugation.†

   Optional: If performing optional on-column DNase digestion (see "Eliminating genomic DNA contamination", page 26), follow Appendix D (page 82), steps 1–4, after performing this step.

7. Add 700 µl Buffer RW1 to the RNeasy spin column. Close the lid gently, and centrifuge for 15 s at ≥8000 x g (≥10,000 rpm) to wash the spin column membrane. Discard the flow-through.†

   Reuse the collection tube in step 8.

   Note: After centrifugation, carefully remove the RNeasy spin column from the collection tube so that the column does not contact the flow-through. Be sure to empty the collection tube completely.

   Skip this step if performing optional on-column DNase digestion (page 69).

* Using 50% ethanol (instead of 70% ethanol) may increase RNA yields from liver samples.
† Flow-through contains Buffer RLT or Buffer RW1 and is therefore not compatible with bleach. See page 6 for safety information.

8. Add 500 µl Buffer RPE to the RNeasy spin column. Close the lid gently, and centrifuge for 15 s at ≥8000 x g (≥10,000 rpm) to wash the spin column membrane. Discard the flow-through.

   Reuse the collection tube in step 9.

Note: Buffer RPE is supplied as a concentrate. Ensure that ethanol is added to Buffer RPE before use (see "Things to do before starting").

9. Add 500 µl Buffer RPE to the RNeasy spin column. Close the lid gently, and centrifuge for 2 min at ≥8000 x g (≥10,000 rpm) to wash the spin column membrane.

   The long centrifugation dries the spin column membrane, ensuring that no ethanol is carried over during RNA elution. Residual ethanol may interfere with downstream reactions.

   Note: After centrifugation, carefully remove the RNeasy spin column from the collection tube so that the column does not contact the flow-through. Otherwise, carryover of ethanol will occur.

10. **Optional**: Place the RNeasy spin column in a new 2 ml collection tube (supplied), and discard the old collection tube with the flow-through. Close the lid gently, and centrifuge at full speed for 1 min.

    Perform this step to eliminate any possible carryover of Buffer RPE, or if residual flow-through remains on the outside of the RNeasy spin column after step 9.

11. Place the RNeasy spin column in a new 1.5 ml collection tube (supplied). Add 30–50 µl RNAse-free water directly to the spin column membrane. Close the lid gently, and centrifuge for 1 min at ≥8000 x g (≥10,000 rpm) to elute the RNA.
12. If the expected RNA yield is >30 µg, repeat step 11 using another 30–50 µl RNAse-free water, or using the eluate from step 11 (if high RNA concentration is required). Reuse the collection tube from step 11.

    If using the eluate from step 11, the RNA yield will be 15–30% less than that obtained using a second volume of RNAse-free water, but the final RNA concentration will be higher.

Lexogen QuantSeq library preparation protocol

Detailed Protocol - Library Amplification

This PCR protocol replaces the single indexing PCR protocols in steps 25 - 28 of the QuantSeq Kits User Guides (015UG009, 015UG058, 015UG110), steps 35 - 38 of the SENSE mRNA-Seq V2 Kit User Guide (001UG004), or steps - of the SENSE Total RNA-Seq Kit User Guides (009UG013, 009UG102), respectively.

Preparation

PCR			Purification (Cat. No. 022)
From i5 Dual Indexing Add-on Kit: Dual PCR – thawed at RT **5001 - 5004 or 5001 - 5096 (i5 Index Plate) –** thawed at RT		**spin down before opening!**	**from standard library prep kits: PB** – stored at +4°C **PS** – stored at +4°C **80 % EtOH –** provided by user
From standard library prep kits: 7001 - 7096 (i7 Index Plate) – thawed at RT			**prepare fresh! EB** – stored at +4°C
Enzyme Mix – keep on ice or at -20 °C			
Thermocycler 98 °C, 30 sec			
98 °C, 10 sec 65 °C, 20 sec 72 °C, 30 sec		11- 27x see SENSE or QuantSeq User Guide recommendations or endpoint as determined by qPCR (Cat. No. 020.96)	
72 °C, 1 min 10 °C, ∞			

PCR

The library is amplified to add the complete adapter sequences required for cluster generation, to introduce i5 and i7 indices, and to generate sufficient material for quality control and sequencing.

We strongly recommend performing a qPCR assay to determine the optimal number of PCR cycles for the endpoint PCR. Please use the PCR Add-on Kit (Cat. No. 020.96) and see Instruc- tion Manual 020IM064 for assay details. The qPCR assay is equally efficient for single- and dual- indexed libraries.

ATTENTION: Important information for dual-indexed library amplification!

• •

• •

For dual indexing **REPLACE** the **PCR Mix** (**PCR or PCR**) from the standard library prep kit with the **Dual PCR Mix** (**Dual PCR**) supplied in the i5 Dual Indexing Add-on Kits!
The **Enzyme Mix** needed for the PCR is provided in the standard library prep kits: **E3** from QuantSeq Kits, **E2** from SENSE mRNA-Seq V2 Kit, **E2** from SENSE Total RNA-Seq Kit, or **E** from the PCR Add-on Kit. **Do not use E2 from QuantSeq Kits** for this PCR!

Each i5 Index Plate contains sufficient volume for one library prep per index and is intended for **single use only! Spin down the i5 and i7 Index Plates before opening!** Pierce or cut open the sealing foil of the wells containing only the desired barcodes. Reseal opened wells of the barcode plate after use to prevent cross contamination!

NOTE: At this point we recommend placing the purification components (**PB**, **PS**, **EB**, included in the QuantSeq / SENSE Kits) for step at room temperature, to give them enough time to equilibrate.

Prepare a mastermix containing 7 µl PCR Mix from the i5 Dual Indexing Add-on Kit (**Dual PCR**) and 1 µl **Enzyme Mix** from the standard library prep kits per reaction (see **ATTENTION** note p.4). **ATTENTION:** Do not use **E2** from QuantSeq for the PCR reaction! Do not use **PCR** or **PCR** from the standard kits if dual indexing is intended.

Add 8 µl of this **Dual PCR** / Enzyme mastermix to 17 µl of the eluted library.

Add 5 µl of the respective i5 Index Primer (**5001 - 5004**, in microtubes or **5001 - 5096**, in 96-well plate, from the i5 Dual Indexing Add-on Kits). **ATTENTION:** Spin down the i5 Index Plate before opening! Pierce or cut open the sealing foil of the wells con- taining the desired barcodes. Avoid cross contamination! Reseal opened wells of the barcode plate after use to prevent cross contamination!

Add 5 µl of the respective i7 Index Primer (**7001 - 7096**, in 96-well plate, supplied with the standard kits). **REMARK:** Ensure the total PCR volume is 35 µl, if necessary adjust the volume with Elution Buffer (**EB**) or molecular biology-grade water. Mix well by pi- petting. Seal the PCR plate and quickly spin down to make sure all liquid is collected at the bottom of the well. **ATTENTION:** Spin down the i7 Index Plate before opening! Pierce or cut open the sealing foil of the wells containing the desired barcodes. Avoid cross contamination! Reseal opened wells of the barcode plate after use to prevent cross contamination!

Conduct 11 - 27 cycles of PCR (see recommendations in SENSE and QuantSeq User Guides, or determine the cycle number to use by qPCR) with the follow- ing program: Initial denaturation at 98 °C for 30 seconds, 11 - 27 cycles of 98 °C for 10 seconds, 65 °C for 20 seconds and 72 °C for 30 seconds, and a final extension at 72 °C for 1 minute, hold at 10 °C. Safe stopping point. Libraries can be stored at -20 °C at this point.

Purification

The finished library is purified from PCR components that can interfere with quantification. The Purification Beads (**PB**) may have settled and must be properly resuspended before adding them to the reaction.

ATTENTION: Important information for purification of dual-indexed libraries!
• The following purification protocol replaces the Post PCR Purification described in steps - of the QuantSeq Kits, steps - of the SENSE mRNA-Seq V2 Kit, or steps

- of the SENSE Total RNA-Seq Kit.
• If PCR products were stored at -20 °C, ensure these are thawed and equilibrated to room

temperature before Purification Beads (**PB**) are added.

Add 35 µl of properly resuspended Purification Beads (**PB**) to each reaction, mix well, and incubate for 5 minutes at room temperature. **ATTENTION:** For SENSE Total RNA- Seq and for QuantSeq libraries generated from low RNA input or degraded RNA, add only 31.5 µl **PB**. For SENSE FFPE Total RNA-Seq library preps add only 29 µl **PB**.

Place the plate onto a magnetic plate and let the beads collect for 2 - 5 minutes or until the supernatant is completely clear.

Remove and discard the clear supernatant without removing the PCR plate from the magnetic plate. Make sure that accumulated beads are not disturbed.

Add 30 µl of Elution Buffer (**EB**), remove the plate from the magnet, and resuspend the beads properly in **EB**. Incubate for 2 minutes at room temperature.

Add 30 µl of Purification Solution (**PS**) to the beads / **EB** mix to re-precipitate the library. Mix thoroughly and incubate for 5 minutes at room temperature. **ATTENTION:** Add only 29 µl **PS** for SENSE FFPE Total RNA-Seq library preps.

Place the plate onto a magnetic plate and let the beads collect for 2 - 5 minutes or until the supernatant is completely clear.

Remove and discard the clear supernatant without removing the PCR plate from the magnetic plate. Make sure that accumulated beads are not disturbed.

Add 120 µl of 80 % EtOH and wash the beads for 30 seconds. Leave the plate in con- tact with the magnet as beads should not be resuspended during this washing step. Remove and discard the supernatant.

Repeat this washing step once for a total of two washes. Make sure to remove the supernatant completely.

Leave the plate in contact with the magnet and let the beads dry for 5 - 10 minutes or until all ethanol has evaporated. **ATTENTION:** Dry the beads only at room tempera- ture and do not let the beads dry too long (visible cracks appear). This will negatively influence the elution and hence the resulting library yield.

Add 20 µl of Elution Buffer (**EB**) per well, remove the plate from the magnet, and resus- pend the beads properly in **EB**. Incubate for 2 minutes at room temperature.

Place the plate onto a magnetic plate and let the beads collect for 2 - 5 minutes or until the supernatant is completely clear.

Transfer 15 - 17 µl of the supernatant into a fresh PCR plate. Make sure not to transfer any beads.

At this point, the libraries are finished and ready for quality control, pooling (for multi- plexing, see also Appendix A, p.7), and sequencing. For more details please refer to the respective SENSE and QuantSeq User Guides.

Appendix II: detailed bioinformatic analyses

Documentation of all bioinformatic analyses including gene expression analysis can be found in the project's GitHub repository: https://github.com/gghill/thesis-RNAseq.