Exploring population connectivity and adaptation in two deep sea fishes, *Molva molva* and *Molva dypterygia*

Lydia McGill

University of Salford

School of Science, Engineering and Environment

Submitted on Fulfilment of the Requirements for the Award of the

Degree of MRes Biological Sciences

March 2020

Table of Contents

List of Tables and Figures

Acknowledgements

More than anyone, I would like to acknowledge and thank my supervisors, Ilaria Coscia and Allan McDevitt, for taking me on for this project and being continuously supportive of me throughout my research. They have set a very high bar for supervision and support, which future supervisors will no doubt struggle to meet. I would also like to thank them for their role in helping me secure a PhD position, which without them I would not have had the knowledge or confidence to do so.

Next I would like to thank our research group, the MEG (Molecular Ecology Group), who were always there for discussing problems, practicing presentations, and celebrating results. Everyone involved in this group also had a great supportive role during my studies. Now that I have begun my PhD these people are all greatly missed.

Finally, I would like to thank the University of Salford, for allowing me to carry out my research; using their resources and setting me up with a desk for my work. I would also like to acknowledge the Santander grant I secured through the university, with which I was able to attend iMarco2019 (the international marine connectivity conference) and present my research.

Abstract

The deep sea is typically seen as a stable and constant environment. However, in recent years we are seeing increasing fragility caused by unprecedented human exploitation. Many deep-sea fish species are typically long lived and have specific traits which often limit their resilience to impact and change. Due to the expansion of fisheries, it is important to better understand population structure and connectivity of these deep-sea populations, in order to ensure their sustainable management.

In this study, samples of two deep sea species collected across their distribution range were screened: the common ling, *Molva molva*, and the deeper dwelling blue ling, *Molva dypterygia*. Using Genotype-By-Sequencing (GBS) and the *M. molva* reference genome, 6,566 and 3,073 neutral Single Nucleotide Polymorphism (SNPs) were identified, respectively. Results indicate how the two species exhibit a different structure pattern, with the deeper blue ling showing fine scale differentiation within population samples along the Norwegian coast, and the common ling being more homogeneous. By identifying 3 outlier loci in the common ling, and 5 outlier loci in the blue ling, adaptive divergence is explored. No candidate genes could be identified from the common ling data. With data for the blue ling it was possible to link outlier loci with multiple genes and speculate adaptive divergence from this. Genes linked with responses to environmental variables including light and temperature were among those found.

Overall, the findings presented in this study attempt at filling a knowledge gap about exploited deep sea fish species, and will hopefully aid the sustainable management of these species.

1. Introduction

1.1. Overview

The study of population connectivity is valuable for ecological, evolutionary, management and conservation applications (Etter and Bower, 2015). Connectivity studies in the marine environment have lagged behind that of terrestrial due to technical limitations, such as tracking individuals, especially those with a pelagic larval phase, over vast open spaces of water where there are few obvious barriers (Baxter, 2001; Sá-Pinto *et al.*, 2012). This lagging research has led to a belief of strong connectivity within the marine environment, a belief which is now being challenged in more recent studies (Cowen *et al.*, 2000). With genetic methods we can now explore structure within the marine environment, which is increasingly important for the future management of fisheries and conservation efforts (Hedgecock, Barber and Edmands, 2007).

Marine fisheries are a major source of food for people around the world and with an ever-growing population this pressure has increased over the years (Pauly and Zeller, 2016). It is now widely accepted that species across the world are becoming increasingly affected by human activity (Bernatchez, 2016). Due to this increased demand fisheries are now facing a decline in catches, which is possibly indicative of overfishing (Pauly and Zeller, 2016). Because of this there is now a substantial need to track the status of the oceans in terms of exploitation, to understand their current state and inform decisions of the future (Coll *et al.*, 2016).

In order to sustainably manage marine populations, a better understanding of the degree of connectivity between management units or populations is needed (Verhelst *et al.*, 2016; Coscia *et al.*, 2019). Here, the population structure of two commercially exploited species, *Molva molva* (the common ling) and *Molva dypterygia* (the blue ling), is explored using

genomic tools, in order to unravel the patterns of population structure. Using Genotype-By-Sequencing (GBS) data generated in the lab, bioinformatics is utilised to create a Single Nucleotide Polymorphism (SNP) array for each species using the reference genome recently published for the common ling (Malmstrøm *et al.*, 2017). With the help of these new genomic technologies, this study aims to create data which can inform management, to create better strategies to ensure the future sustainability of these fishes.

Combining any available data utilising different markers will allow for a broader understanding of interactions between populations (Santos *et al.*, 2012; Fernández-Pérez *et al.*, 2018). Here, the data and results generated for the common ling can be compared and combined with that which was found by Gonzalez *et al.* (2015) based on microsatellite markers, giving a stronger basis for management strategies in this species. Creating a SNP array for genomic analyses of the species will give a greater insight into the structural possibilities hinted at in Gonzalez's study, and may allow insight into finer scale differentiation and adaptation across the range. Broadening to include the blue ling in this study means this valuable information is gained for more than one commercially exploited species for which data is deficient (ICES, 2017) and separate management strategies can be inferred onto the two. As fisheries go deeper this is even more essential, and research should continue to delve into the deep to gain an understanding of the human-induced pressures experienced here.

After exploring population structure, adaptation within the species through the identification of outlier loci is then explored. This allows for an insight into any signatures of adaptive divergence occurring within populations, and thus what could be driving differentiation between them (Duforet-Frebourg, Bazin and Blum, 2014; de Villemereuil and Gaggiotti,

2015). This study aims to overcome the issue of an understudied area by adding to the current knowledge of deep-sea connectivity.

1.2. Marine genomics and fisheries

Due to the great connectivity and thus high gene flow of marine species, it has often proven difficult to detect structure within marine populations (Junge *et al.*, 2019). Due to this, more powerful tools were sought after (Bagley, Lindquist and Geller, 1999; Allendorf, 2017). Many studies have utilised microsatellites to begin exploring connectivity, often finding panmixia or low but significant genetic structure (Bagley, Lindquist and Geller, 1999; Allendorf, 2017; Saha *et al.*, 2017). Issues with microsatellite DNA mostly arise through technical difficulty of creating the markers, and genotyping errors due to low yield and poor quality DNA (De Barba *et al.*, 2016; Janjua *et al.*, 2020).

Further genetic research has allowed for the development of a greater biological understanding of the marine environment (Miller and Gunasekera, 2017). Next Generation Sequencing (NGS) has revolutionised research, where researchers have been able to expand on knowledge of population structure, traceability, phylogenetics, and so on (Kumar and Kocour, 2017). These new technologies are aiding exploration of structure in marine environments after decades of molecular studies in which limited data could be found (Gagnaire *et al.*, 2015). Although the likes of genetic tools such as microsatellites are still greatly popular in marine studies, the use of genomic Single Nucleotide Polymorphisms (SNPs) is on the rise, making it increasingly possible to detect structure (Selkoe *et al.*, 2016). SNPs occur upon the change of a single base in DNA sequence, and thus are the most common source of genetic variation in and among species (Ruperao and Edwards, 2015).

Application of SNP markers has already begun to unravel patterns of population structure for many marine species. This has been successful for Atlantic cod (Bradbury *et al.*, 2013), Atlantic mackerel (Rodríguez-Ezpeleta *et al.*, 2016), albacore and Atlantic blue fin tuna (Albaina *et al.*, 2013), European hake (Milano *et al.*, 2014), Atlantic herring (Lamichhaney *et al.*, 2017), and many other species. This focus on commercially valuable species has proven important for management.

To begin on the path to better fisheries management, there is need for an increase in the use of NGS technologies (Cuéllar-pinzón *et al.*, 2016). This means more genomic studies will be used to improve understanding of environmental variation (Bernatchez *et al.*, 2017). This is paramount for assessing fish stocks worldwide, to understand exploitation in order to maintain food security, to conserve, and to achieve sustainability (Rosenberg *et al.*, 2018). This has failed in the past for many reasons, including a lack of data on the issue (Rosenberg *et al.*, 2018).

The deep sea is the largest ecosystem in the world, with great biodiversity and importance (Clarke *et al.*, 2015). Due to the great depths and distance from land, it is a difficult habitat to explore and monitor (Miller and Gunasekera, 2017). The species found here have slower life histories than those in shallower waters and may also be more vulnerable to fishing through aggregation at sea mounts, ridges, or slopes; commonly targeted by fisheries (Victorero *et al.*, 2018). It is assumed that deep sea species have a higher degree of connectivity, with greater dispersal than those in shallower waters (Baco *et al.*, 2016). With higher connectivity there is an added assumption of greater resilience (Cowen *et al.*, 2007). The deep sea is poorly understood and there are multiple misconceptions leading to these suppositions (Baco *et al.*, 2016).

Combining the decline of stocks with advancement of fishing technology has led to an expansion of fisheries away from the continental shelf into deeper waters (Mangi *et al.*, 2016; Collie *et al.*, 2017; Victorero *et al.*, 2018). It is already known that many deep-sea fisheries are unsustainable, with an exhibited decline in abundance of commercial fish species since deep sea fishing began (Clarke *et al.*, 2015). Negative impacts of fishing increase from 600m and deeper, along with a decrease in catch value from these depths (Clarke *et al.*, 2015). This is salient, indicating that the deeper we go, the more impact we may be having on population levels and their recovery. It is also increased effort with a lesser payoff, and thus does not appear viable to continue in the long term.

Genetic data has now allowed a greater insight into this habitat, helping towards estimation of connectivity and predicting decreases in diversity in the deep sea (Miller and Gunasekera, 2017). Such studies are starting to indicate that the deep-sea is not as well connected as originally thought. It has been shown that deep basins may reduce gene flow, through limited adult migration and poor drift of pelagic larvae (Knutsen *et al.*, 2009). Such limitations are likely to lead to population differentiation due to poor gene flow (Knutsen *et al.*, 2009).

With genomic data this is now being expanded on. Research using SNP markers on the deep sea orange roughy found population structure, with evidence for local adaptation and isolation by distance (Gonçalves da Silva *et al.*, 2019). Genetic studies originally struggled to find any structure within this species and improved to find low levels of differentiation (Varela, Ritchie and Smith, 2013). This uncertainty can now be clarified through SNP analyses, and future studies may want to continue using these methods at differing geographic scales to further explore structure within this species.

Considering the current knowledge available for population genomics shows how this increasingly useful tool can be used to tackle the challenges which are associated with sustainable fisheries management, and more effort should be directed towards integrating such methods into these practices (Bernatchez *et al.*, 2017; Mullins *et al.*, 2018; Pecoraro *et al.*, 2018). We can now use genomics to gain a genome-wide and allele-specific view into the structure and diversity of any organism (Valenzuela-Quiñonez, 2016). It is crucial that we have been able to identify structure in highly mobile species of commercial importance (Junge *et al.*, 2019). Whether this structure be that of high connectivity or showing differentiation between populations, there is knowledge which can then be taken into consideration with fisheries management (Junge *et al.*, 2019). Therefore, the use of genomics to explore population structure is needed in marine species for conservation and stock management (Lal, Southgate, Jerry and Zenger, 2016).

If fisheries management does not change, then further decline and collapse of worldwide fisheries can be expected (Costello *et al.*, 2016). Such change would be specific to different areas and/or species, thus, in order to make changes there must be further research to increase knowledge for informing management strategies. With better strategies recovery can happen quickly, increasing fish abundance as well as security for maintaining populations and profits for the fisheries industry (Costello *et al.*, 2016).

1.3. Neutral vs. adaptive divergence

As well as structural analyses, signatures of selection can now also be explored with genomic methods (Riginos *et al.*, 2016). This knowledge combined with environmental factors helps to identify neutral or adaptive differentiation within/between populations (Riginos *et al.*, 2016).

Neutral genetic mutations will occur over time as the natural process of random drift transpires (Allen *et al.*, 2015). Also occurring is selection, which can over time lead to differentiation between populations and even speciation (Allen *et al.*, 2015). Through selection, rapid evolution can take place in response to environmental change (Bosse *et al.*, 2017). We can now identify the signatures of this selection through genomic research; linking our genetic understanding with ecological factors to predict potential drivers of selection (Bosse *et al.*, 2017). Using genomic techniques, we are able to investigate the adaptive diversity of populations through outlier loci (Valenzuela-Quiñonez, 2016). This opens the door to a new understanding of the potential effects of fishing and/or climate change on fish populations and stocks (Valenzuela-Quiñonez, 2016).

Although the great dispersal and connectivity of marine populations comes with an assumption of limited local adaptation, we are finding an increase in evidence for adaptation to different environments (Sherman *et al.*, 2016). With the nature of the marine environment allowing for large populations, we are now discovering that there is great genetic diversity within these populations (Kelley *et al.*, 2016). This aids towards adaptation to novel/different environments (Kelley *et al.*, 2016).

With genomic techniques, it is now possible for us to identify specific genes linked with differentiation and infer adaptation from these. For example, changes within opsin genes have been described within marine species in response to differing light levels (Rennison *et al.*, 2016; Pierotti *et al.*, 2017). Such an adaptation is thought to be commonplace (Rennison *et al.*, 2016). Because of this, and due to the rapid evolution of these genes allowing for great adaptability, it has thus been a useful tool in testing for adaptive divergence (Rennison *et al.*, 2016; Pierotti *et al.*, 2017).

As well as identifying genes associated with adaptive divergence, genomic methods can identify parallel evolution in geographically distant populations where similar forces are driving natural selection (Lamichhaney *et al.*, 2017). Such convergent evolution is not only of interest within different species who exhibit similar phenotypes, but also on a finer scale within a single species across populations. This has been seen in Atlantic herring, whose populations are geographically distant and genetically distinct, but share linking genetic factors with timing of reproduction (Lamichhaney *et al.*, 2017).

Adaptive evolution has become of particular interest in invasion biology. In these scenarios, adaptation occurs rapidly under the selective forces of a novel environment (Bernardi *et al.*, 2016). We have been able to study invasive species to gain knowledge on adaptive evolution (Bernardi *et al.*, 2016). This, and studies looking at fishery-induced evolution, go on to help us predict adaptability of marine species/populations in a changing global environment (Bernardi *et al.*, 2016; Foo and Byrne, 2016; Waples and Audzijonyte, 2016). With climate change comes oceans which are warmer, have a lesser oxygen content, and lower pH levels (Foo and Byrne, 2016; Waples and Audzijonyte, 2016). Marine species are now experiencing a great need to adapt to deal with the stressors brought on by the changing climate and through the pressures of fishing.

1.4. Next-Generation Sequencing

Over the past 50 years population genetics have been used to explore variation and evolution; to gain an understanding of populations and their interactions (Casillas and Barbadilla, 2017). New developments have now led to the production of genomic methods of analysis, such as Next-Generation Sequencing (NGS) which allows sequencing of the entire genome in a short

period of time (Toriello, 2016). Compared with genetics, genomics allows exploration of genome-wide effects rather than loci-specific, giving a much broader and informative understanding (Elmer, 2016). This has allowed for larger data sets for such studies, now growing in number as techniques improve (Selkoe *et al.*, 2016).

NGS can refer to any highly parallel or high output sequencing method which produces genomic data (Levy and Myers, 2016). There are many companies and technologies already specialising in NGS and this continues to amplify with the decreasing cost, increasing speed, and increasing high-performance (Levy and Myers, 2016; Kumar and Kocour, 2017). The expanding ability of NGS means it is now possible to obtain billions of reads from a sequencing experiment (Reinert *et al.*, 2015).

Until more recently, genomic studies were generally performed on model species due to associated costs and complexities (Greminger *et al.*, 2014). To battle this issue, a technique of reduced representation sequencing (RRS) was developed in which restriction enzymes are used to digest genomic DNA into fragments which can then be selected and sequenced (Greminger *et al.*, 2014). This would give a fraction of the genome but allow for identification of polymorphisms and genotype calling in species with no reference genome, and without having to sequence the genome of every individual (Davey *et al.*, 2011; Greminger *et al.*, 2014). If a reference genome is available, then RRS fragments can be mapped to this and SNPs can be called in this way (Davey *et al.*, 2011). Such a method is therefore applicable to both model and non-model organisms.

There are now multiple methods developed from this idea, including RAD-seq which utilises a single restriction enzyme and sequences essentially all fragments surrounding restriction sites associated with the enzyme (Davey *et al.*, 2011; Kagale *et al.*, 2016). The fragments in

this method are sheared to a desired length for sequencing (Davey *et al.*, 2011; Kagale *et al.*, 2016). Another technique further developing from this is that of GBS which has indirect size selection, so fragment lengths will vary (Andrews *et al.*, 2016). Other methods differ in approach by utilising more than one restriction enzyme but maintaining the fragment shearing, for example double-digest RAD (ddRAD) (Andrews *et al.*, 2016).

With a great shift from genetic to genomic research, NGS is now used in all manner of scientific disciplines and can be put forward to answer a plethora of questions (Kumar and Kocour, 2017). Notably, studies focused on ecology, evolution, and conservation genomics have benefited greatly from NGS (Andrews *et al.*, 2016). One area of research which NGS has proven an essential tool is that of fisheries science, looking at population structure, phylogenetics, signatures of selection, and so on (Kumar and Kocour, 2017). Developing SNP arrays for exploited species is an important aspect in fisheries science, as it creates a basis for future research within such species (Martínez *et al.*, 2017). This can then be geared towards stock assessment and management of species and marine areas.

Bioinformatics has developed alongside NGS as a solution to the growing amounts of genomic data produced in the lab (Ogbe, Ochalefu and Olaniru, 2016; Bolyen *et al.*, 2018). It allows for storing, retrieving, and organising such data, transforming it into that which is valuable for a number of analyses (Ogbe, Ochalefu and Olaniru, 2016). One of the common tasks performed through bioinformatics is calling of SNPs (Keane *et al.*, 2016; Mielczarek and Szyda, 2016). In fish, SNPs are found around every 100bp (Siccha-Ramirez *et al.*, 2018). This high density of SNPs makes them the most popular markers in genomic screening (Siccha-Ramirez *et al.*, 2018).

There is currently an array of bioinformatics software available for SNP calling. These pipelines allow for production of data with a reference genome available or can offer an alternative *de novo* method when no reference is available (Torkamaneh, Laroche and Belzile, 2016). This means that, although it is beneficial to have access to a reference genome when SNP calling, it is not essential, and we can still produce data for species lacking this. Many of these softwares have been tested out on GBS data in the past (Torkamaneh, Laroche and Belzile, 2016), but since such programs are ever-updated and improved along with lab techniques any comparisons can become out of date. One of the most popular programs for generation of SNPs is *STACKS* (Catchen *et al.*, 2013), which includes *de novo* and reference pipelines, and can support data produced from many different lab techniques (Paris, Stevens and Catchen, 2017).

An important element in SNP calling is that of filtering, as mistakes made during library preparation and/or bioinformatics may lead to incorrect interpretation of data in later analyses (O'Leary *et al.*, 2018). Like with SNP calling in bioinformatics software, there are many options to choose from when it comes to filtering a SNP dataset. *STACKS* offers its own filtering options, along with the many alternative programs including *TASSEL* (Bradbury *et al.*, 2007), *plink* (Purcell *et al.*, 2007), and *vcftools* (Danecek *et al.*, 2011). There are also multiple packages in *R* (R Core Team, 2017) which can be used to filter through a dataset.

Although there is a degree of reliability on SNPs called with a reference genome available, filtering is still required for this and *de novo* assemblies. Filters recommended include removing loci or individuals with a certain amount of missing data (O'Leary *et al.*, 2018). This may seem obvious, to get rid of data which is not informative enough and could give false patterns in results. Applying these restrictions to data will allow optimisation of the dataset.

It can, however, have a negative effect on results through removing informative loci. This can come about during filtering for minor allele frequency which, due to uncertainty of accuracy, can remove rare SNPs which would otherwise infer structure (Malomane *et al.*, 2018). This is also the case when filtering out SNPs which differ significantly from Hardy-Weinberg equilibrium. Due to this, the utmost care must be taken in the filtering process to find a balance in removing uncertain and inaccurate data, yet maintaining that which is informative for our analyses.

1.5. The common ling, *Molva molva*

The common ling *Molva molva* is a commercially fished species of the Lotidae family, which is a close relative to the cod (*Gadus sp.*) (Sustainable Fisheries Partnership, 2016; Luna, 2020b). It is the largest species of the Gadiformes, growing up to 200cm in length (Rowley, 2008). The distribution of the common ling ranges from the Barents Sea down to Morocco, and across the Northern Atlantic (Cohen *et al.*, 1990; Luna, 2020b). They may also be rarely found in the Mediterranean Sea (Cohen *et al.*, 1990). It can be distinguished by the length of the barbel compared to the diameter of the eye; the sensory barbel, protruding from the lower jaw, being longer than the diameter of the eye (Rowley, 2008; Luna, 2020b). This species can be found in depths of up to 600m (Rowley, 2008). It is generally a solitary, benthic species, found on rocky bottoms (Cohen *et al.*, 1990; Rowley, 2008; Luna, 2020b).

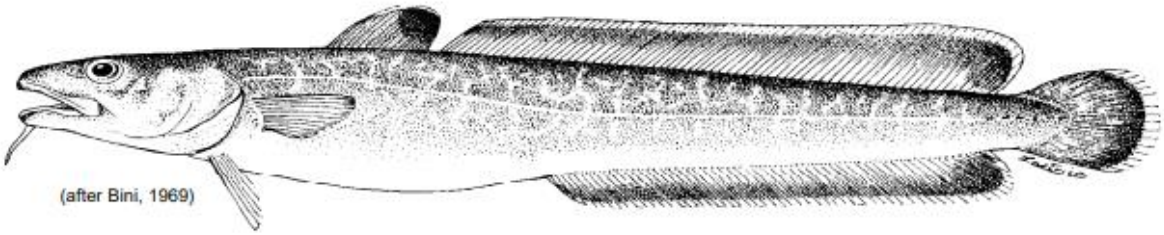

Figure 1: *Molva molva*, the common ling (Cohen *et al.*, 1990)

Dietary data indicates that the common ling feeds on other fish species, such as blue whiting, herring, cod, and flatfish, as well as some crustaceans, cephalopods, and starfish (Rowley, 2008; Luna, 2020b).

The common ling spawns offshore in the months of March to August at around 100-300m depths (Cohen *et al.*, 1990; Rowley, 2008). Females will grow faster and live longer than males, reaching maturity at around 90-100cm and living up to 14 years (Rowley, 2008). Males mature at around 80cm and can live up until 10 years (Rowley, 2008).

Fisheries are known to catch the common ling through longlines, gillnets, and bottom trawls (Sustainable Fisheries Partnership, 2016). In Iceland, where this species is particularly sought after, stock assessment is carried out using size- and age-structured models and results appear to indicate a current decline (Elvarsson *et al.*, 2018).

It has been stressed by the International Council for the Exploration of the Sea (ICES) that research investigating the population structure of this species is now required to help inform management, as very little is known about it (Ring *et al.*, 2009). Previous studies employing microsatellites suggested the possible presence of structure in the Northeast Atlantic (Ring *et al.*, 2009; Gonzalez *et al.*, 2015). Gonzalez *et al.* (2015) found that there was a split between the East and West, creating one group around Norway and another group consisting of Rockall

and Iceland. This prior knowledge can be used alongside any findings from this study to further inform management.

1.6. The blue ling, *Molva dypterygia*

The blue ling *Molva dypterygia* is also a commercially fished species of the Lotidae family and is the sister species of the common ling (Cohen *et al.*, 1990; Barnes, 2008). It is smaller than the common ling, growing up to 155cm in length, with a shorter barbel which does not exceed the diameter of the larger eye (Cohen *et al.*, 1990; Barnes, 2008; Papasissi, 2020). Distribution of the blue ling is similar to that of the common ling; ranging from the Barents Sea up to Spitsbergen, down to Morocco, across the Northern Atlantic, and into the Mediterranean Sea (Cohen *et al.*, 1990; Papasissi, 2020). The two species will often occur next to one another in the water column, with a little overlap in depth distribution. This species can be found at depths of up to 1000m, living on muddy bottoms (Cohen *et al.*, 1990; Papasissi, 2020). They are a demersal species and they are known to occur around seamounts and knolls (Stocks, 2009). The blue ling feed on crustaceans and fish; such as flatfishes, gobies, and rocklings (Cohen *et al.*, 1990; Papasissi, 2020).

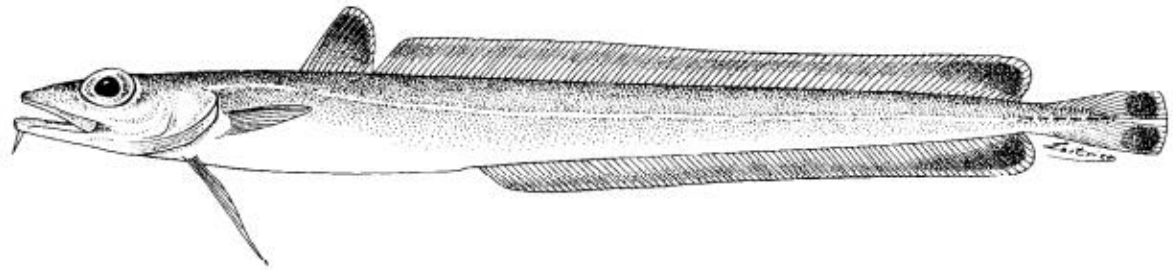

Figure 2: *Molva dypterygia*, the blue ling (Cohen *et al.*, 1990)

Spawning behaviour for this species occurs in April and May between 500-1000m depths across West Scotland to Norway and from the Faeroes to South Iceland; they spawn separately in the Mediterranean from late winter to early spring at 500-600m depths (Cohen *et al.*, 1990). Again, females are faster growing than males and will reach maturity at around 88cm, living up to 20 years (Cohen *et al.*, 1990). Males will reach maturity around 75cm and live to 17 years (Cohen *et al.*, 1990).

Like the common ling, the blue ling is commercially fished with longlines and bottom trawls, mainly in the North-Eastern Atlantic (Cohen *et al.*, 1990). It is known that the blue ling have been fished for decades, and with increasing exploitation stocks began to collapse through the 1990s (Helle *et al.*, 2019). Although there has been no genetic research for the blue ling, protection areas for spawning aggregations have been put in place to the West and North-West of the British Isles, following warnings from ICES that the species is susceptible to sequential depletion of spawning aggregations (Large *et al.*, 2010). Genetic studies are now desired for new and improved management strategies to be implemented to aid in the recovery of populations (Helle *et al.*, 2019).

Overall, the objectives of this study are to explore structural connectivity in two commercially exploited species using populations analyses. Connectivity and structure is explored using populations analyses, with F_{st}, PCA and DAPC, and Structure analyses. This can give us an insight into the population structure of such valuable deep-sea species, which have in the past been assumed as highly connected (Baco *et al.*, 2016). This could change current views of deep-sea connectivity and thus the approach to deep-sea fishing in the future. Results for the common ling can also be discussed alongside those findings from Gonzalez *et al.* (2015), further expanding our knowledge for this species.

Comparisons can also be made between the two species studied here, seeing how increasing depth can influence connectivity. Additionally, candidate gene analyses are performed to explore signs of selection, and the possibilities of adaptive divergence through the links found with outlier loci are discussed. This gives an insight for the possible drivers of divergence connected with the deep sea. Together, the outcomes of this study should help expand current knowledge of deep-sea connectivity and deep-sea species in general.

2. Materials and Methods

2.1. Data Collection/Lab work

Samples of *Molva molva* and *Molva dypterygia* were caught throughout their range in the Atlantic by research surveys and fishing vessels, between the years 2005-2015. Coordinates were recorded for points at which fish were captured, as well as the length of each individual. Tissue samples were obtained from gill, muscle, or fin clips.

Genotype-By-Sequencing (GBS) was used to produce Single Nucleotide Polymorphism (SNP) data (Elshire *et al.*, 2011). In total, 83 common ling and 190 blue ling were sequenced. The laboratory work was not part of the present study, which exclusively carried out the bioinformatic analysis starting from the raw data provided. DNA was extracted with the Blood and Tissue QIAGEN kit (QIAGEN, 2006), and its concentration was then standardised before using it to construct modified GBS libraries. GBS is a simple highly multiplexed system for constructing reduced representation libraries (Elshire *et al.*, 2011). It was digested at 75C for 2h with ApeKI, followed by adaptor and barcode ligation, purification, PCR, another purification, and quantification and only after this step, pooling of the samples. The libraries were then sent to the KU Leuven Genomics Core (Genomics Core Leuvin, 2020), where all five libraries were individually size selected on a Pippin Prep unit (Sage Science, 2020), checked for quantity using qPCR, and paired-end sequenced on one lane of a HiSeq 2500 platform (Illumina, 2020).

Taking the coordinates which were provided for catching points of individuals we could create maps to visualise the samples and populations. These were plotted in QGIS (QGIS Development Team, 2019). One map was produced for the common ling and one map for the blue ling (Figs. 3 & 4).

Table 1: Population names, central coordinates of catch area, and number of individuals sampled from each population for the common ling

Population	*Location*	*Coordinates (Lat, Long)*	*No. individuals*
BB15	Bay of Biscay	44, -3	2
BE08	Bergen - Sotra Bridge	60.39, 5.16	8
BO14	Bømlafjorden	60.06, 5.45	10
HA14	Hardangerfj., Steinstøberget	60.38, 6.28	10
IB13	Indre Boknafjord	59.28, 5.85	1
NY13	Nygrunnen	69.18, 14.5	10
RA08	Rockall	58.08, -13.35	8
RA14	Rockall	57.79, -13.44	6
RYF13	Ryfylke	59.13, 5.73	2
RYV14	SW Ryvingen	57.88, 7.21	10
SO14	Sørfjorden	60.43, 5.56	6
TF05	Tromsoflaket	69.02, 13.44	10

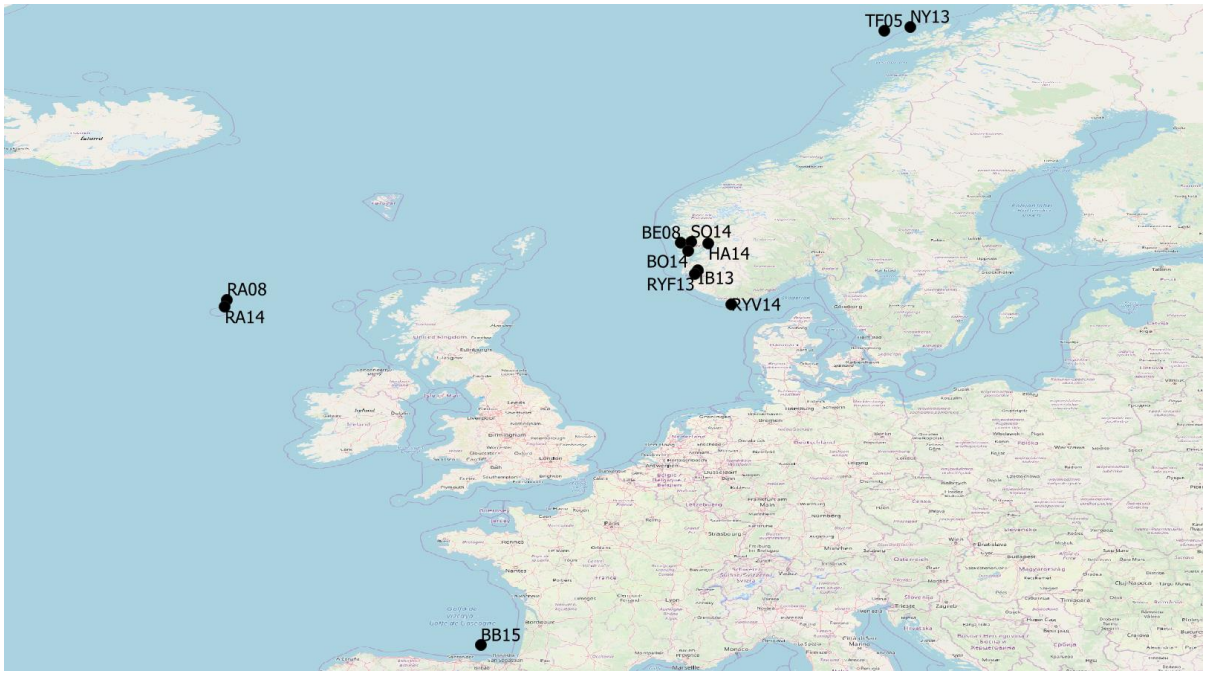

Figure 3: Map of common ling sample sites

Table 2: Population names, central coordinates of catch area, and number of individuals sampled from each population for the blue ling

Population	*Location*	*Coordinates (Lat, Long)*	*No. individuals*
AD07	Anton Dohrn	57.42, -11.22	1
BO14	Bømlafjorden	60.06, 5.45	4
HA14	Hardangerfj., Steinstøberget	60.39, 6.28	3
IB13	Indre Boknafjord	59.28, 5.85	1
NY13	Nygrunnen	69.18, 14.5	12
RA07	Rockall	56.95, -13.43	5
RA10	Rockall	56.01, -14.87	40
RA11	Rockall	58.20, -14.96	47
RS07	Rosemary Bank	59.10, -9.92	13
RYF13	Ryfylke	59.27, 5.72	2
RYV14	SW Ryvingen	57.88, 7.21	6
SL07	Slope	57.61, -9.63	19
SL11	Slope	58.50, -9	22
SL14	Slope	56.72, -9.13	4
SO13	Sørfjorden	60.43, 5.51	1
GRE15	Greenland NAFOXIVb	59, -44	10

Figure 4: Map of blue ling sample sites

2.2. Bioinformatics and filtering

Raw data was provided to process using bioinformatics, to produce the SNP data required for analysis. This was accomplished through *process_radtags, denovo_map.pl* and *ref_map.pl*, and *populations* pipelines from *STACKS v2* (Catchen *et al.*, 2013). First, coverage was tested to check the quality of the data using *FastQC* (Andrews, 2010). This is generally used to identify the ideal length to use when truncating sequences (parameter *-t*) in *STACKS*. Through *FastQC* it was found that all sequences were of 126bp in length, of which none were flagged as poor quality, and the per base sequence quality remained high throughout with no significant drop to the end of the reads. From this came the decision to keep all sequences at this length, and allow other filters applied later in the process to remove any undesirable data.

Using this information from *FastQC, process_radtags* could be informed to generate the stacks, feeding the program information using *-t* to indicate trim length deciphered from *FastQC* quality data (in this case leaving the length as is). Files were indicated to be paired, barcodes were given to identify individuals within the sequences, and the restriction enzyme used was given (ApeKI). The *-r*, *-c*, and *-q* functions were also given, which indicates to the program to rescue (*-r*) barcodes and RAD-Tags, clean (*-c*) data by removal of any uncalled bases, and to quality check (*-q*), discarding reads with low quality scores. These would help to quality control the data where trimming is not in place.

The next step in the pipeline is to call SNPs, which can be done either *de novo* or with a reference genome. The reference-based approach is known to be the best of the two; generally calling more SNPs with greater accuracy (Shafer *et al.*, 2017). It is suggested that this approach is taken with a closely related genome (Shafer *et al.*, 2017). Both approaches

were carried out in order to compare the two, and with the recently produced *Molva molva* genome (Malmstrøm *et al.*, 2017) this was therefore possible. There is no available genome for *M. dypterygia* yet, but being the two species of the same genus, we could assume that they were genetically related enough for us to use one genome as a reference for both (Nevado, Ramos-Onsins and Perez-Enciso, 2014; Shafer *et al.*, 2017). This comparison of methods allows exploration of the outcomes when SNPs are called on raw data, against when the data is first aligned to the reference genome. It was expected, as has been indicated, that using a reference genome would allow more SNPs to be called with certainty, and thus generate more data for later analysis and be the more suitable option of the two (Shafer *et al.*, 2017).

First, with *de novo*, parameters must be chosen. Various parameters were tested in *de novo* to optimise the calls of SNPs. This was performed on a subset of the data, and parameters chosen were used for both species for consistency and control. Values of *m* were tested between 3 and 7, as this should allow enough exploration of the parameter (Paris, Stevens and Catchen, 2017). With tests for *M* between 2 and 7, a value was chosen which was optimum at *r80* (Paris, Stevens and Catchen, 2017). Finally, to decide a value for n it is suggested to explore -*n* = -*M*, -*n* = -*M* - 1, and -*n* = -*M* + 1 (Paris, Stevens and Catchen, 2017). Optimum parameters chosen were -*m*=3, -*M*=4, and -*n* = 5. These chosen parameters could then be input to stacks with the full dataset.

To use reference-aligned data, reads were first aligned using *Bowtie2* (Langmead and Slazberg, 2013), rather than programs such as *BWA* (Li and Durbin, 2010) or *SOAP2* (R. Li *et al.*, 2009) as it is faster, more sensitive, and more accurate than those alternatives (Kagale *et al.*, 2016). This created sam files for each individual containing all of the aligned reads. These

files were then converted into bam files and sorted using *SAMtools* (H. Li *et al.*, 2009; Li, 2011). These were then ready to run through *ref_map.pl*.

In running the populations analysis in *STACKS*, minor filters were put in place to finally generate the initial SNP data. With *r* = 0.80 only SNPs in 80% of a population were kept, and with *write_single_snp* there was a single SNP written for each loci to minimise linking data. Output formats of vcf and genepop were generated. These formats could be used in *R* for filtering.

The files generated by *STACKS* could be then further filtered in *R* (R Core Team, 2017). It was decided that both the *de novo* and referenced data would be filtered to compare outcomes, and see which method used in *STACKS* is preferential. *R* packages required for filtering included; *poppr* (Kamvar, Tabima and Grünwald, 2014; Kamvar, Brooks and Grünwald, 2015), *adegenet* (Jombart, 2008; Jombart and Ahmed, 2011), *pegas* (Paradis, 2010), *vcfR* (Knaus and Grünwald, 2016, 2017), *hierfstat* (Goudet, 2005), and *diveRsity* (Keenan *et al.*, 2013).

Either the genepop or vcf files produced could be used, through a genepop to genind conversion in the *adegenet* package, or through a vcf to genind conversion using the radiator package (Gosselin, 2019). In this study the genepop to genind conversion was chosen for use in both species to maintain consistency. Taking this route, use of the multiple packages listed offered different filtering options for the SNPs. The main sections involved in filtering were as follows:

To filter the SNPs, focus was first directed towards missing data. With this, loci with a missing data content of over 10% were removed from the data. This was performed using the *missingno* function from *poppr*.

They were then filtered by F_{is}, removing loci with a F_{is} value greater than 0.999, or under -0.999. This meant removing those which were too close to the values of 1 and -1, assuming that there is not complete inbreeding, or complete outbreeding. These values were calculated with the *wc* function in the *hierfstat* package.

To filter for minor allele frequency (MAF), loci with a MAF value below the set threshold were removed. In order to retain the maximum number of informative loci, three MAF threshold we tested: 0.05 (the most widely used and also the most restrictive one), 0.025 and 0.01. These filters were applied to the data through *informloci* from *poppr*. The whole filtering process was run and three final datasets were generated in order to compare the outcomes.

Markers not in linkage equilibrium (LD) were filtered out using two thresholds, 0.7 and 0.5, to test the effect of this parameter on the number of SNPs retained in the final dataset. Values for LD were calculated using *pair.ia* of the *poppr* package, and any loci with values over the chosen cut off could then be removed from the data sets. In this step a threshold of 0.7 was chosen for continued analyses.

Then finally, filtering by Hardy-Weinberg equilibrium (HWE) was performed by removing loci with p-values of <0.05, as calculated by *hw.test* of the *pegas* package. The loci with undesirable values were removed using this data. This was the last step of the filtering process.

The final SNP dataset was explored to find how individuals were affected by the process.

2.3. Exploring signals of adaptive divergence

Neutrality tests were carried out using three different approaches with different assumptions: *pcadapt* (Luu, Bazin and Blum, 2017), *bayescan* (Foll and Gaggiotti, 2008; Foll *et al.*, 2010; Fischer *et al.*, 2011) and the function *lfmm* of the package *LEA* (Frichot and Francois, 2015). *Bayescan* was run with default parameters; setting the parameters of the chain with 5000 outputted iterations, a thinning interval size of 10, 20 pilot runs with a length of 5000, and an additional burn-in length of 50,000 (Foll, 2012). The parameters of the model were also set at default with the prior odds for neutral model set to 10, the F_{is} prior was set as uniform between 0 and 1, and a setting of 0.1 for the threshold for the recessive genotype as a fraction of maximum band density (Foll, 2012). *LEA* and *pcadapt* were run with an alpha value of 0.01 as a threshold for the outliers. For *pcadapt*, *K* was determined using a plot showing the proportion of explained variance as suggested by the authors. The value of *K* from *LEA* was determined using a plot of cross entropy values.

The loci which were found in both the *pcadapt* and *LEA* analyses were separated from the SNP datasets, and analyses were carried out on putatively neutral datasets.

To further investigate the outliers found, *BLAST* (Altschul *et al.*, 1997) searches were conducted on the loci. This step was used as an indication of whether these loci became outliers through adaptive evolution, identifiable by links found between these loci and genes, in relation with the environment (Jones *et al.*, 2012). It could also be indicative of neutral evolution through an absence of such links. Outlier locations were found by using the SNP number identified in the outlier calling process and tracing this back to the SNPs originally output from *STACKS*. Searching for these SNPs in the vcf files, which contain additional information on SNPs, the location of SNPs in the genome were found. This allowed sequences

for use in *BLAST* to be found. Since it is possible for SNPs to affect genes up and downstream within the genome, sequences 10,000bp in either direction were used. This would give a broad picture of the genes driving any change within these species.

Through locating these sequences within the *Molva molva* genome (assembled to scaffold level) the scaffold in which each outlier was located and their locations upon these scaffolds were revealed. Scaffolds are labelled OOFG01000001.1 to OOFG01111875.1. Scaffold labels could then be used in *BLAST* to search up and downstream from each SNP. Using Integrative Genomic Viewer (IGV) (Thorvaldsdóttir, Robinson and Mesirov, 2013), outlier loci were viewed within their scaffolds, and spanning across scaffolds it was possible to search 10,000bp up and down stream of the outliers, finding the full sequence areas which could be queried in *BLAST*.

MEGAblast and *blastn* were utilised to search the outlier loci against the NCBI database. With each search a cut-off e-value of $1x10^{-10}$ was put in place, (Humble, Martinez-Barrio, *et al.*, 2016; Humble, Thorne, *et al.*, 2016; Rohfritsch *et al.*, 2018; Cormier *et al.*, 2019).

2.4. Population Structure

In *R*, using the package *assigner*, Weir and Cockerham's Theta estimator (overall and pairwise) for F_{st} were computed, with pairwise confidence intervals estimated with 1000 permutations (Weir and Cockerham, 1984). Using these values, it could be determined whether F_{st} results were statistically significant.

Population structure was estimated initially with Principal Component Analysis (PCA) and Discriminant Analysis of Principal Components (DAPC). These analyses were performed in *R*,

using *dudi.pca* of the *ade4* package (Dray and Dufour, 2015; Bougeard and Dray, 2018) for PCA plots and *dapc* function of the *adegenet* package (Jombart, 2008; Jombart and Ahmed, 2011) for DAPC plots. Analyses were run on each dataset generated by different filtering values in order to estimate the influence of such settings on the results.

First, for the PCA, missing data was removed which meant taking any loci which had missing data in any individuals out of the dataset. Cross-validation was performed to choose the number of axes to retain in principal component and discriminant analyses steps. The plotting of this data was then straight forward, through *dudi.pca* and *s.class* functions for PCA, and the *dapc* and *scatter* functions for DAPC plots. Overall, three PCA plots and three DAPC plots were produced for each species.

The dataset was further analysed in *Structure*, another clustering approach with different underlying assumptions from PCA and DAPC (Pritchard, Stephens and Donnelly, 2000). This was performed with a burn-in of 20,000 and retaining 180,000 iterations. This was repeated five times for each value of *K* tested, from 1 to 10. The results of *Structure* were visualised with *Structure Harvester* (Earl and vonHoldt, 2012).

Further population structure analysis was performed using *fastStructure* (Raj, Stephens and Pritchard, 2014) in *Python* (Van Rossum and Drake Jr, 1995). Structure was tested for population assumptions up to 10, and the program was also used to estimate the most likely number of populations in the data given. This was performed using *chooseK*; returning *K* values for model complexity that maximises marginal likelihood, and for model components used to explain structure in data. These values give a range of values within which *K* is predicted to fall.

Using the *LEA* package (Frichot and Francois, 2015) in *R* it is possible to carry out Structure-like analyses, and even PCA and outlier analyses. This package can also run structure analyses at greater speeds than *Structure* itself, returning results within a time which would rival that of *fastStructure*. There were also no issues in obtaining these results, which made this package a very helpful tool. Performing the structure analyses in *LEA*, 10,000 iterations were performed, as has been suggested in past use of the function (Frichot and Francois, 2015). Again, *K* was measured from 1 to 10 with 5 repetitions to give results comparable to those of the *Structure* analysis.

3. Results

3.1. Data quality

FastQC showed that the sequencing data was of a very high quality. Drop off in quality towards the end of sequences was minimal, and no reads were flagged as 'poor'. This was the case across all data for both species.

Table 3: Basic statistics for sequences of the common ling, and the blue ling for lane1(L1) and lane 2 (L2)

	Common ling	*Blue ling L1*	*Blue ling L2*
Total sequences	501,736,552	488,705,838	531,172,060
Sequences flagged as poor quality	0	0	0
Sequence length	126	126	126
%GC	50	52	51

3.2. Bioinformatics

In demultiplexing the data, STACKS retained 480,039,207 reads for the common ling and 1,019,335,300 for the blue ling. When testing *de novo* parameters, SNPs decreased with increasing -m, and increased with increasing -M and -n, as is expected. These tests were performed on a subset of the data using individuals from both species.

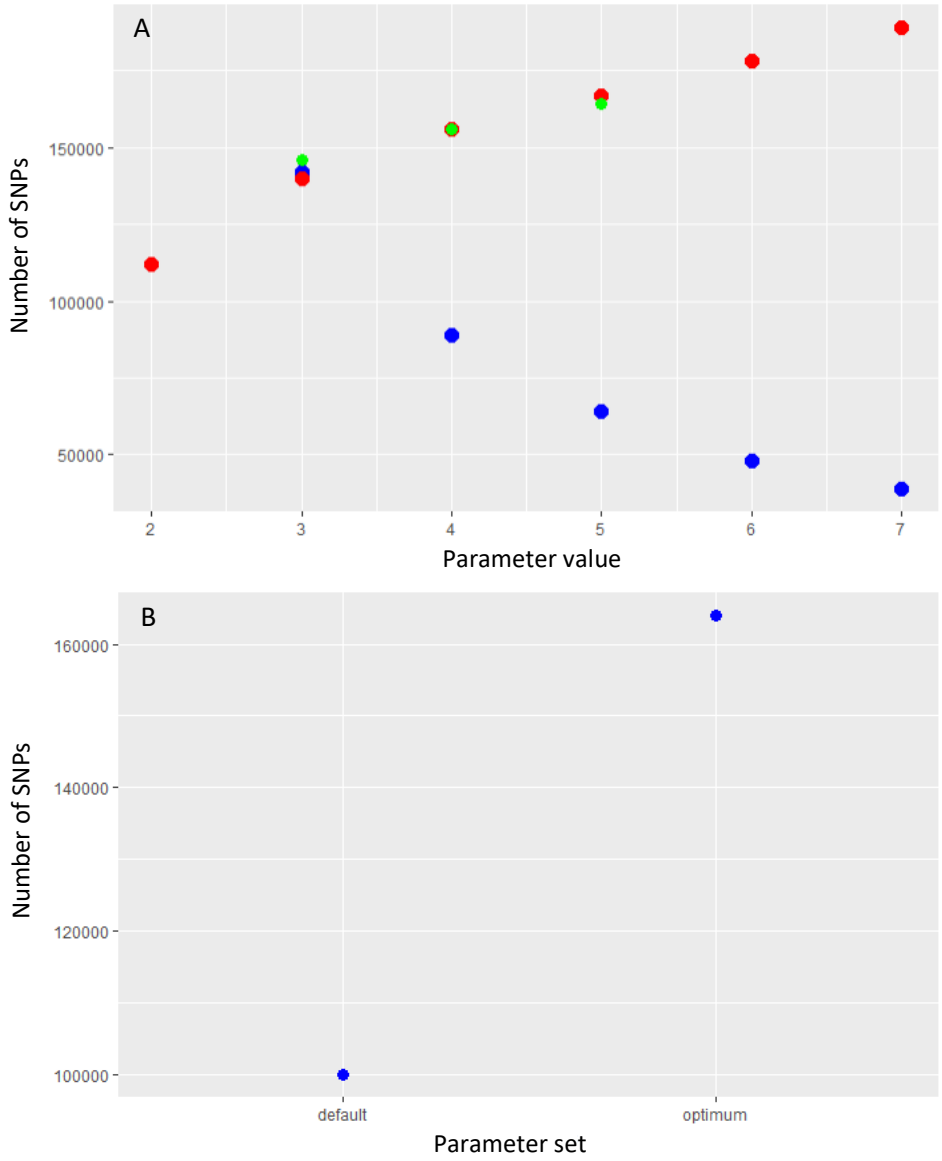

Figure 5: Tests carried out using a subset of data from the two species, a) Showing the number of SNPs retained with changing values of m in blue, M in red, and n in green, all at r80, and b) showing number of SNPs generated through default parameters (m=3 M=2 n=1) versus optimum parameters (m=3 M=4 n=5)

A value of *m* = 3 was chosen for the final production of *STACKS* (Fig. 5a). *M* = 4 was selected for the final dataset (Fig. 5a). Although higher than usual this appears to be optimum at r80 (Paris et al., 2017). This could be indicative of alleles at low frequency, which can later be filtered out through removing loci below a minor allele frequency (Paris et al., 2017). Once *M* was set at 4 *n* was then tested (Fig. 5a), finding that *n* = *M* + 1 obtained the highest number of SNPs for the data.

Comparing the default parameters for *STACKS* with the optimum parameters calculated for this dataset, there are a substantial amount more SNPs obtained through the optimum than through default (Fig. 5b).

3.3. Filtering

From the initial production of SNPs in *STACKS*, with the minor filtering used in the populations program (SNPs in 80% of individuals & one SNP per loci), the initial SNPs were produced ready for filtering.

After filtering through *R*, these decreased to more workable numbers. Although the blue ling produced more SNPs *de novo*, both species retained more SNPs from the reference-aligned data after the complete filtering. There was great difference seen between reference-aligned and *de novo* when filtering out missing data; with the aligned data maintaining around double the number of SNPs of those for the *de novo* dataset. Reference aligned common ling left 6,569 SNPs, while de novo left 2,983 (MAF = 0.05) (Table 4). Reference aligned blue ling generated 3,078 and *de novo* gave 2,118 (MAF = 0.05) (Table 4).

Table 4: Number of loci retained at each step of the filtering process in the various datasets

	Reference common ling	*de novo common ling*	*Reference blue ling*	*de novo blue ling*
Initial SNPs after STACKS	134,133	126,267	133,610	144,648
Missing loci filter	81,136	37,243	70,838	38,366
Fis filter	71,843	35,856	68,038	37,915
MAF filter	11,116	4,559	4,753	3,000
LD filter	7,288	3,385	3,666	2,502
HWE filter	6,569	2,983	3,078	2,118

Just the reference aligned data was used for further filtering tests. When changing MAF to 0.025, 4,177 SNPs were retained for the blue ling and 9,714 SNPs for the common ling (Table 5). With MAF at 0.01 this was 6,534 for the blue ling and 10,619 for the common ling (Table 5). These were then to be further analysed to explore the effects of the differing filtering parameters on the population structure detected within these populations.

Table 5: Changing values of MAF filtered for common ling and blue ling, and number of SNPs retained at each step after this filter

	Common ling MAF 0.05	*Common ling MAF 0.025*	*Common ling MAF 0.01*	*Blue ling MAF 0.05*	*Blue ling MAF 0.025*	*Blue ling MAF 0.01*
MAF filter	11,116	16,828	31,824	4,753	6,379	9,772
LD filter	7,288	10,708	10,619	3,666	4,928	7,553
HWE filter	6,569	9,714	10,619	3,078	4,177	6,534

3.4. Outlier loci and *BLAST*

Using *pcadapt,* eight loci were identified as outliers for the blue ling and 43 for the common ling. *Bayescan* did not detect any outliers in the common ling, and only one was flagged for the blue ling (not matching any of those found in the *pcadapt* approach). The *lfmm* function of the *LEA* package found 11 outlier loci for the blue ling and five from the common ling data.

For the sake of this study, we did not retain the *Bayescan* result, and we considered outliers only those loci that were identified as such in *LEA* and *pcadapt*. This meant there were three outlier loci for the common ling leaving 6,566 SNPs once removed from the dataset, and five outlier loci in the blue ling data which left 3,073 SNPs. These SNP putatively neutral datasets were then used for the population structure analyses.

The three loci were found to show up in both analyses. These were numbered 183129_18, 461445_111, and 660280_103 in the SNP dataset. This could then be traced back to the output from *STACKS*, which also held information for the location of these SNPs on the reference genome.

For the first common ling outlier (scaffold OOFG01008146.1 at position 1603) scaffold 00FG01008144.1 position 7590 to OOFG01008148.1 at position 5118 was used. The second common ling outlier (OOFG01029390.1 position 654) used the sequence from OOFG01029385.1 at position 1485 to OOFG01029393.1 position 2780, and the final search used OOFG01053503.1 at 908 to OOFG01053509.1 at 1296.

Through the same methods as those used on the common ling, five outlier loci were found within the blue ling data. These were numbered 51266_27, 367341_115, 494102_9, 714576_3, and 722729_54 in the dataset.

In the blue ling the 5 loci were searched starting from the first (OOFG01001639.1 position 2518) going from scaffold OOFG01001638.1 at position 4808 to OOFG01001639.1 at position 12518. The second loci (OOFG01015574.1 position 977) was investigated using the sequence between OOFG01015571.1 position 4169 and OOFG01015577.1 position 3007. For the third outlier (OOFG01024386.1 position 2108) utilising the sequence from OOFG01024382.1 position 1077 to OOFG01024388.1 position 3794. The fourth outlier (OOFG01048534.1 position 6703) used the sequence from OOFG01048532.1 position 13344 to OOFG01048538.1 position 2244. The final outlier (OOFG01048943.1 position 3987) for *M. dypterygia* had a search sequence spanning from OOFG01048940.1 position 6669 to OOFG01048946.1 position 44.

In the *BLAST* search, no significant matches were found for genes in relation to the common ling outliers. For the blue ling the search results found genes which matched significantly with sequences provided for the outlier loci. Because the search included 10,000bp up and downstream of SNP locations, it was possible to find multiple genes related to each SNP for the blue ling. These are outlined in the following tables (Tables 6 – 10), with one table produced per SNP. Tables are presented with the protein name, the associated gene, the species from which the gene matched with our sequence from, the E-value of this match, the hit length, and finally a description of the function of said gene. It should be noted that the hit length is longer than the 126bp of sequences which the SNPs are based off due to the searching up and downstream from the SNP locations. This allowed the hit length to be more variable, and sometimes exceed the 126bp length.

Table 6: *BLAST* hits for SNP 51266_27 located on the scaffold OOFG01001639.1 at the 2518th base

Protein	Gene	Species	E-Value	Hit length	Function
Short-wavelength sensitive opsin 2A and 2B	SWS2A and SWS2B	*Lota lota*	3e-22	220	Responsible for creating the protein opsin; proteins which allow for normal colour vision. Part of S cone cells; type 2 opsins in particular are responsible for picking up on the violet and blue spectrums.
Anti-apoptotic protein NR-13	NR-13 gene	*Gadus morhua*	7e-17	105	Encodes for the anti-apoptotic protein NR-13 which stops/inhibits cell growth.
Transmembrane growth hormone receptor 2 mRNA	Growth hormone receptor 2	*Epinephelus coioides*	3e-11	70	Produces a transmembrane receptor for growth hormones which allows these hormones to pass through cell membranes.

With the first outlier locus, *BLAST* returned three significant matches to sequences within our search query (Table 6). This gave three different genes from knowledge of three different species. The first, and most significant of these results, being from *Lota lota* (Burbot); a species of the same family as the ling (Lotidae) (Luna, 2020a), and it is also known as the freshwater ling. This gene encodes for short-wavelength sensitive opsin 2A and 2B and had an E- value of 3e-22. Opsin are the proteins which allow for normal colour vision, with type 2 in particular picking up on the violet and blue spectra (Marshall, Carleton and Cronin, 2015).

The second gene found relating to the first outlier came from *Gadus morhua* (Atlantic cod) with an E- value of 7e-17. This gene encodes for an anti-apoptotic protein called NR-13, which stops/inhibits cell death. This protein has been found at high levels in embryos, but then decreasing in number after birth (Lee *et al.*, 1999).

The third, but still significant, matching gene came from the species *Epinephelus coioides* (Orange-spotted grouper) and encodes for the growth hormone receptor 2. This is a transmembrane receptor for the growth hormone which is related to development or overall size of a species (Herrington and Carter-Su, 2001).

Table 7: *BLAST* hits for SNP 367341_115 located on the scaffold OOFG01015574.1 at the 977th base

Protein name(s)	Gene name(s)	Species	E-Value	Hit length	Function
Hepcidin precursor	HAMP	*Gadus morhua*	7e-46	312	Hepcidin is a hormone which controls regulation of iron absorption and distribution across tissues.As above.
Antifreeze glycoprotein polypeptide 8 and antifreeze glycoprotein polypeptide 9	AFGP8 and AFGP9	*Boreogadus saida*	2e-45	164	These polypeptides bind to ice crystals to inhibit growth and recrystallisation of ice.Hepcidin is a hormone which controls regulation of iron absorption and distribution across tissues.
Hepatoma-Derived Growth Factor	HDGFL3	*Esox lucius*	9e-38	140	Plays a role in cellular proliferation and differentiation and is also linked with growth of tumours when at high levels of expression.
Glucose transporter 1	GLUT1	*Gadus morhua*	1e-36	196	This transporter is responsible for the glucose level of uptake.
Hepatoma-Derived Growth Factor	HDGFL3	*Danio rerio*	1e-36	196	See column 6, row 3.
Apolipoprotein E mRNA	Apolipoprotein E	*Callorhinchus milii*	4e-11	73	A protein involved in the metabolism of fats in the body.

For the second outlier there were five matching genes (Table 7). The first match came from Atlantic cod with an E-value of 7e-46. The Hepcidin precursor gene links our outlier with the hormone which controls the regulation of iron absorption and distribution across tissues (Mirciov *et al.*, 2017).

The second matching gene had an E value of 2e-45 and was found in *Boreogadus saida* (Polar/Arctic cod). This gene codes for the antifreeze glycoprotein polypeptides 8 (AFGP8) and 9(AFGP9). These polypeptides bind to ice crystals to inhibit growth and recrystallisation of ice (Zhuang *et al.*, 2019; Tsuda *et al.*, 2020).

A gene known in *Esox lucius* (Northern pike) was the third found for the second outlier, with an E value of 9e-38. This was a Hepatoma-Derived Growth Factor like 3 (HGDFL3) gene, which may play a role in cellular proliferation and differentiation and is also linked with the growth of tumours when at high levels of expression (Yang *et al.*, 2016; Chung-Jung *et al.*, 2018). This same HGDFL3 gene was also matched to our outlier through the species *Danio rerio* (Zebrafish), with an E value of 1e-36.

The next gene was that of glucose transporter 1 (GLUT1) which again matched from Atlantic cod, with an E value of 1e-36. This transporter is responsible for the level of glucose uptake (L. Wang *et al.*, 2019).

The final hit for this outlier came from *Callorhinchus milii* (Elephant shark/Australian ghost shark) for a gene which codes for apolipoprotein E, matching with an E-value of 4e-11. This class of proteins are involved in the metabolism of fats in the body (Marais, 2019).

Table 8: *BLAST* hits for SNP 494102_9 located on the scaffold OOFG01024386.1 at the 2108th base

Protein name(s)	Gene name(s)	Species	E-Value	Hit length	Function
NMDA receptor subunit	NR2B	*Carassius auratus*	2e-50	146	When activated, by glutamate and glycine, allows positively charged ions to flow through the cell membrane; important for controlling synaptic plasticity and memory function.
Pregnancy-associated plasma protein A (Pappalysin)	PAPPA	*Salvelinus alpinus*	2e-50	196	Insulin-like growth factor-binding proteins involved in local proliferative processes such as wound healing and bone remodelling.

The third outlier loci gave a *BLAST* match for just two genes (Table 8). The first of these came from *Carassius auratus* (goldfish) with an E value of 2e-50. The gene encodes for NMDA receptor subunit NR2B which, when activated by glutamate and glycine, allows positively charged ions to flow through the cell membrane (Furukawa *et al.*, 2005). It is important for controlling synaptic plasticity and memory function (Li and Tsien, 2009).

The other gene came from *Salvelinus alpinus* (Arctic char) with pregnancy-associated plasma protein A; E value 2e-50. These proteins are thought to be involved in local proliferative processes such as wound healing and bone remodelling (Bulut *et al.*, 2018).

Table 9: *BLAST* hits for SNP 714576_3 located on the scaffold OOFG01048534.1 at the 6703rd base

Protein name(s)	Gene name(s)	Species	E-Value	Hit length	Function
Somatotropin-2	Growth hormone 2	*Oncorhynchus tshawytscha*	1e-168	441	Plays an important role in growth control and is involved in the regulation of several anabolic processes (implicated as an osmoregulatory substance important for seawater adaptation).
Raftlin	Raftlin-like pseudogene	*Oncorhynchus mykiss*	3e-54	210	Pseudogenes are genes which are related to working genes but have lost some functionality. This gene is linked to the RFTN1 gene which produced Raftlin; a protein which activates signalling of T cells and B cells upon bacterial lipopolysaccharide stimulation and plays an important role in formation and maintenance of lipid rafts.
Homeobox proteins	HOX-epsilon14; HOX-epsilon13; HOX-epsilon11	*Lethenteron camtschaticum*	6e-37	168	See column 7, row 2.
Homeobox proteins	HOX6IV; HOX5IV; HOX4IV; HOX3IV; HOX2IV	*Eptatetrus burgeri*	5e-14	86	Hox genes are a group of related genes that specify regions of the body plan of an embryo along the heal-tail axis of animals.

The 4 sequence matches for the fourth outlier indicated a larger number of possible genes affected. The first of those being a match to *Oncorhynchus tshawytscha* (Chinook salmon), informing us of a link with growth hormone 2, matching with an E value of 1e-168. This reaffirms strong connections of our outlier with growth-related functions. Growth hormones play an important role in growth control (Herrington and Carter-Su, 2001).

The second sequence match contained information from *Oncorhynchus mykiss* (Rainbow trout), with an E value of 3e-54. The gene found here was of a Raftlin-like protein. Raftlin activates signalling of T cells and B cells upon bacterial lipopolysaccharide stimulation and plays an important role in formation and/or maintenance of lipid rafts (Matsumoto and Tatematsu, 2017).

Eptatretus burgeri (Inshore hagfish) linked the outlier with 5 different genes; HOX6IV, HOX5IV, HOX4IV, HOX3IV, and HOX2IV. This scored an E value of 5e-14. The hox genes are a group of related genes that specify regions of the body plan of an embryo along the head-tail axis of animals (Song *et al.*, 2020).

The final sequence match for this outlier offered less of a diverse array of hox genes; HoxIV14, HoxIV13, and HoxIV11. These three genes came from a sequence from *Lethenteron camtschaticum* (Arctic lamprey), E value 6e-37.

Table 10: *BLAST* hits for SNP 722729_54 located on the scaffold OOFG01048943.1 at the 3987th base

Protein name(s)	Gene name(s)	Species	E-Value	Hit length	Function
Antifreeze glycoprotein polypeptides	AFGP1; AFGP2; AFGP3; AFGP4 pseudogene	*Microgadus tomcod*	8e-36	139	These polypeptides bind to ice crystals to inhibit growth and recrystallisation of ice.
Serine/threonine-protein kinase MAK and Ras-related protein Rab-14	MAK14 and RAB14 genes	*Boreogadus saida*	4e-33	139	MAK14 is part of a signalling cascade mediating cell growth, adhesion, survival, and differentiation through regulation of transcription, translation, and cytoskeletal rearrangements. RAB14 is involved in membrane trafficking between the Golgi complex and endosomes during early embryonic development.
Antifreeze glycoprotein polypeptides	AFGP1 pseudogene; AFGP2; AFGP3; AFGP4; AFGP5; AFGP6; AFGP7 pseudogene	*Gadus morhua*	1e-33	150	See column 6, row 1.
Antifreeze glycoprotein polypeptides	AFGP1 pseudogene; AFGP2; AFGP3; AFGP4; AFGP5 pseudogene; AFGP6 pseudogene; AFGP7 pseudogene	*Boreogadus saida*	1e-32	150	See column 6, row 1.
Serine/threonine-protein kinase MAK and Ras-related protein Rab-14	MAK14 and RAB14 genes	*Microgadus tomcod*	2e-31	124	See column 6, row2.
Serine/threonine-protein kinase MAK and Ras-related protein Rab-14	MAK14 and RAB14 genes	*Gadus morhua*	2e-30	139	See column 6, row 2.
Retinoic acid receptor gamma	RARG-like gene	*Pantodon buchholzi*	1e-26	102	Regulates gene expression, required for limb bud movement, required for skeletal growth/matrix homeostasis/growth plate function.
Short-wavelength sensitive opsin 2A and 2B	SWS2A; SWS2B	*Merluccius polli*	3e-16	141	Responsible for creating the protein opsin; proteins which allow for normal colour vision. Part of S cone cells; type 2 opsins in particular are responsible for picking up on the violet and blue spectrums.

The fifth and final outlier found the greatest number of genes of all of the searches (Table 10). The first gene found in our final outlier search came from *Microgadus tomcod* (Frostfish/Atlantic tomcod/Winter cod), giving two of the matches linking with the sequence, with multiple genes found within these matches. These were MAK14 and RAB14 genes, E value 2e-31, and various antifreeze glycoproteins with an E value of 8e-36. The first match appears to link with the regulation of membrane trafficking (Junutula *et al.*, 2004), while the latter with adapting to cold (Tsuda *et al.*, 2020). These same groupings of genes (of MAK14 and RAB14 with antifreeze glycoproteins) were also matched to our outlier through *Gadus morhua* (Atlantic cod) and *Boreogadus saida* (Polar cod/Arctic cod).

Pantodon buchholzi (Freshwater butterflyfish) finally gave a different genetic match, with an E value of 1e-26. Here it indicated the gene for a RARG-like protein, creating the RERG receptor for retinoic acid, regulating gene expression, required for limb bud movement, and required for skeletal growth/matric homeostasis/growth plate function (Pennimpede *et al.*, 2010; Ikami *et al.*, 2015; Shimo *et al.*, 2019).

Merluccius polli (Benguela hake) linked this outlier with short-wavelength sensitive opsin 2A and short-wavelength sensitive opsin 2B genes; E value 3e-16. As seen before, this links the outlier with genes required for normal colour vision (Marshall, Carleton and Cronin, 2015).

In the overall searches, many gene matches were found under the term “predicted”, which suggests that in the future there will be more knowledge to help us understand the influence of outlier loci in fish species.

For further comparison of the two species, the SNP locations for the outliers found in the blue ling were investigated in the common ling. None of these SNPs were produced in the common ling. This would be because there was a lack of any polymorphism at all found at these sites,

thus they were not SNPs. This shows that the effects allowing these minor alleles to exist within the blue ling are not having any influence on the common ling, which could be due to the environmental variation between the two species. This insight could have valuable inferences for the depth variation of the species, and the anthropogenic pressures which they face.

3.5. Population structure of the common ling

When performing the population analyses with the varied parameter datasets there was little difference found in the outputs. Because of this, the neutral datasets using the MAF filter of 0.05 were chosen as the best representation of the data as this is the strictest of the three filters tested. The 3 SNPs which were found in the outlier loci analyses were removed from the dataset to leave 6,566 SNPs for population analyses.

The highest pairwise F_{st} values are generally found between Atlantic-Fjord populations (Table 11). However, the strongest of F_{st} values are found between the Bay of Biscay (BB15) and Rockall (RA08, RA14) populations, which are both situated in the Atlantic.

Table 11: Pairwise F_{st} values for common ling populations (left/bottom) with lower and upper confidence intervals (right/top); ranges which do not cross 0 indicate F_{st} values which are significantly different from 0 and are highlighted as such

	BB15	*BE08*	*BO14*	*HA14*	*IB13*	*NY13*	*RA08*	*RA14*	*RYF13*	*RYV14*	*SO14*	*TF05*
BB15		0.0165 – 0.0330	0.0149 – 0.0300	0.0155 – 0.0311	0 – 0	0.0165 – 0.0331	0.0198 – 0.0373	0.0334 – 0.0495	0 – 0	0.0162 – 0.0320	0.0124 – 0.0293	0.0152 – 0.0311
BE08	0.0243		0.0015 – 0.0058	0.0023 – 0.0069	0 – 0	0.0015 – 0.0062	0.0092 – 0.0152	0.0102 – 0.0162	0.0010 – 0.0140	0.0029 – 0.0071	0 – 0	5.07e-05 – 0.0048
BO14	0.0228	0.0036		0.0007 – 0.0050	0 – 0.0160	0.0013 – 0.0055	0.0060 – 0.0111	0.0062 – 0.0118	0 – 0.0090	0.0011 – 0.0051	0.0003 – 0.0058	0.0004 – 0.0045
HA14	0.0237	0.0048	0.0029		0 – 0	9.06e-05 – 0.0041	0.0069 – 0.0121	0.0078 – 0.0137	0 – 0.0074	0.0004 – 0.0045	0 – 0.0045	0 – 0.0031
IB13	0	0	0.0047	0		0 – 0.0059	0.0031 – 0.0284	0.0272 – 0.0491	0 – 0.0065	0 – 0.0172	0 – 0.0057	0 – 0.0106
NY13	0.0245	0.0037	0.0034	0.0020	0		0.0059 – 0.0110	0.0075 – 0.0133	0 – 0.0093	0.0008 – 0.0051	0 – 0.0046	0 – 0.0021
RA08	0.0282	0.0121	0.0086	0.0095	0.0150	0.0084		0.0052 – 0.0111	0.0075 – 0.0210	0.0077 – 0.0132	0.0052 – 0.0121	0.0054 – 0.0105
RA14	0.0415	0.0133	0.0090	0.0107	0.0384	0.0104	0.0081		0.0166 – 0.0288	0.0079 – 0.0136	0.0115 – 0.0184	0.0062 – 0.0115
RYF13	0	0.0074	0.0033	0.0013	0	0.0031	0.0145	0.0228		0.0009 – 0.0135	0.0008 – 0.0148	0 – 0.0067
RYV14	0.0238	0.0052	0.0031	0.0024	0.0053	0.0029	0.0105	0.0108	0.0069		0.0004 – 0.0059	0.0008 – 0.0051
SO14	0.0208	0	0.0031	0.0017	0	0.0017	0.0087	0.0148	0.0079	0.0031		0 – 0.0040
TF05	0.0229	0.0022	0.0025	0.0009	0	0.0002	0.0080	0.0087	0.0002	0.0031	0.0013	

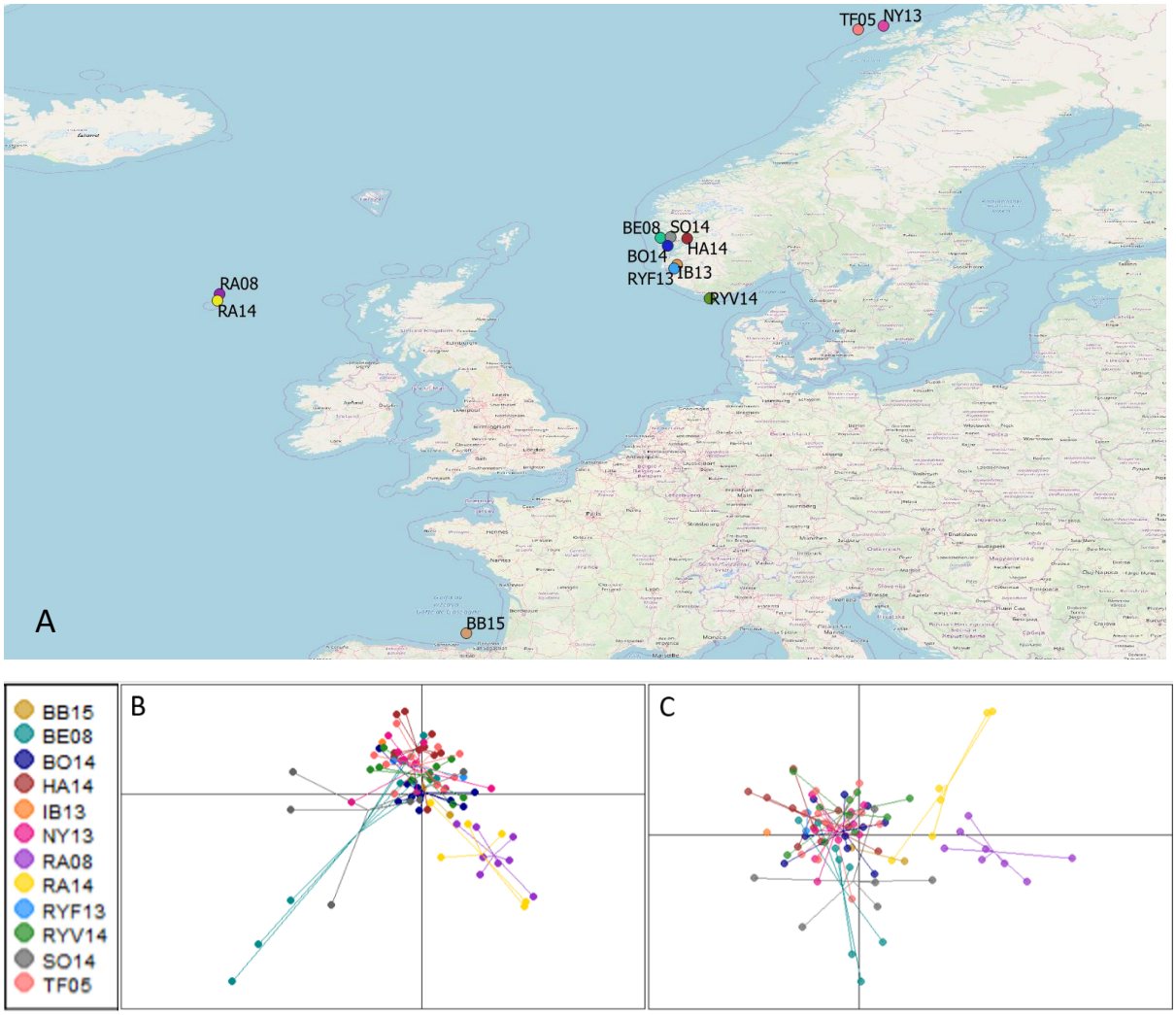

Figure 6: Population structure of the common ling with; a) map showing coordinates for the locations of each population sample site; b) PCA; and c) DAPC

In the PCA of the common ling there appears to be outlying populations stretching out from a clustered centre (Fig. 6b). The clearly differentiating populations shown in the PCA plots are the two Rockall population samples, and two fjord populations located around Bergen and Sørfjorden (BE08/turquoise and SO14/grey).

The further analysis using DAPC shows the same populations creating the main cluster, with again the same populations deviating from this (Fig. 6c). The two Rockall populations (RA08/purple and RA14/yellow) are more strongly differentiated here, which corresponds to

the distance found between these individuals and those found in the rest of the populations (fjords). The two fjord populations deviating are the same two seen in the PCA, found around Bergen.

Population analyses of the outlier SNPs did not add to this data. Analyses were also carried out on a non-neutral dataset, which provided the same genetic structure as seen here. Both analyses of the outlier SNPs and of the non-neutral dataset can be found in the appendix as supplementary materials. A table of pairwise F_{st} values calculated using the outlier data for the common ling populations can also be found in the supplementary material. The values calculated were generally higher than that of the neutral data, but this bias and probable inaccuracy could be expected when reducing the markers to such a small number and thus this data is not being presented here (Willing, Dreyer and van Oosterhout, 2012).

3.6. Population structure of the blue ling

Again, the neutral dataset using the filter of MAF 0.05 was chosen for the best representation of the blue ling data. This was due to little difference again found between the analyses using the more relaxed filtering values and kept the analyses consistent between species. The 5 SNPs which were found in the outlier loci analyses were removed from the dataset to leave 3,073 SNPs for population analyses.

The majority of the pairwise F_{st} values differed significantly from 0, indicated by the 95% confidence intervals, and most of the highest F_{st} values were recorded between Atlantic and fjord populations (Table 12). The only Atlantic population which differed significantly from

others was Anton Dohrn (AD07), differing from the Rockall (RA07, RA10 and RA11) and Slope (SL07, SL11 and SL14) populations (Table 12). Within the fjords, the Indre Boknafjord (IB13) and Bømlafjorden (BO14) populations differed from one another, and Nygrunnen (NY13) differed from all but Bømlafjorden (Table 12). The strongest values (>0.025) were all shared between Atlantic and fjord population pairs (Table 12).

Notably the Rockall (RA) populations all have a value of 0, along with the Greenland population (GRE15) and most of slope populations (SL) as well (Table 12).

Table 12: Pairwise F_{st} values for blue ling populations (left/bottom) with lower and upper confidence intervals (right/top); ranges which do not cross 0 indicate F_{st} values which are significantly different from 0 and are highlighted as such

	AD07	*BO14*	*HA14*	*IB13*	*NY13*	*RA07*	*RA10*	*RA11*	*RS07*	*RYF13*	*RYV14*	*SL07*	*SL11*	*SL14*	*SO13*	*GRE15*
AD07		0.0163 – 0.0610	0 – 0.0450	0 - 0	0.0170 – 0.0485	0.0030 – 0.0408	0.0031 - 0.0343	0.0017 - 0.0319	0 – 0.0321	0.0076 - 0.0670	0.0104 - 0.0481	0.0059 - 0.0360	0.0025 - 0.0329	0 – 0.0316	0 – 0	0 – 0.0250
BO14	0.0379		0 – 0.0078	0.0052 - 0.0505	0 – 0.0066	0.0038 - 0.0197	0.0076 - 0.0173	0.0064 - 0.0154	0.0013 - 0.0118	0 – 0.0249	0 – 0.0031	0.0072 - 0.0177	0.0072 - 0.0173	0 – 0.0077	0 – 0.0205	0.0022 - 0.0148
HA14	0.0222	0		0 – 0.0318	0.0039 - 0.0166	0 – 0.0168	0.0063 - 0.0172	0.0057 - 0.0172	0.0035 - 0.0168	0 – 0.0171	0 – 0.0124	0.0083 - 0.0204	0.0067 - 0.0196	0 – 0	0 – 0.0253	0.0003 - 0.0137
IB13	0	0.0272	0.0048		0.0026 - 0.0328	0 – 0.0335	0.0065 - 0.0381	0.0051 - 0.0363	0 – 0.0337	0 – 0.0516	0 – 0.0275	0.0061 - 0.0377	0.0063 - 0.0389	0 – 0.0327	0 – 0	0 – 0.0311
NY13	0.0336	0.0015	0.0101	0.0172		0.0098 - 0.0189	0.0112 - 0.0162	0.0116 - 0.0163	0.0098 - 0.0156	0.0053 - 0.0225	0.0026 - 0.0099	0.0121 - 0.0174	0.0130 - 0.0179	0.0038 - 0.0145	0.0001 - 0.0296	0.0085 - 0.0148
RA07	0.0225	0.0117	0.0085	0.0132	0.0143		0 – 0	0 – 0	0 – 0.0002	0.0132 - 0.0372	0.0047 - 0.0168	0 – 0.0012	0 – 0.0013	0 – 0.0034	0 – 0.0386	0 – 0.0015
RA10	0.0185	0.0126	0.0116	0.0222	0.0137	0		0 – 0.0002	0 – 0.0004	0.0171 - 0.0346	0.0086 - 0.0156	0 – 0.0012	0 – 0.0013	0 – 0.0012	0.0098 - 0.0397	0 – 0.0008
RA11	0.0164	0.0109	0.0115	0.0198	0.0139	0	0		0 – 0	0.0140 - 0.0310	0.0095 - 0.0168	0 – 0.0016	0 – 0.0005	0 – 0	0.0075 - 0.0368	0 – 0
RS07	0.0153	0.0064	0.0099	0.0153	0.0127	0	0	0		0.0125 - 0.0316	0.0062 - 0.0144	0 – 0.0017	0 – 0.0012	0 – 0.0009	0 – 0.0309	0 – 0
RYF13	0.0387	0.0108	0.0028	0.0218	0.0140	0.0247	0.0260	0.0226	0.0212		0 – 0.0149	0.0189 - 0.0370	0.0162 - 0.0332	0.0078 - 0.0329	0 – 0.0407	0.0085 - 0.0286
RYV14	0.029	0	0.0043	0.0096	0.0063	0.0109	0.0120	0.0131	0.0102	0.0034		0.0115 - 0.0187	0.0090 - 0.0170	0.0004 - 0.0147	0 – 0.0283	0.0072 - 0.0166
SL07	0.021	0.0123	0.0144	0.0221	0.0147	0	0.0002	0.0003	0	0.0278	0.0149		0 – 0.0024	0 – 0.0029	0.0043 - 0.0355	0 – 0.0015
SL11	0.0176	0.0123	0.0132	0.0222	0.0156	0	0.0003	0	0	0.0245	0.0130	0.0009		0 – 0	0.0045 - 0.0381	0 – 0.0014
SL14	0.0123	0	0	0.0123	0.0092	0	0	0	0	0.0211	0.0074	0	0		0 – 0.0433	0 – 0
SO13	0	0.0001	0	0	0.0150	0.0198	0.0246	0.0223	0.0146	0.0114	0.0107	0.0196	0.0210	0.0207		0 – 0.0318
GRE15	0.0093	0.0084	0.0070	0.0135	0.0116	0	0	0	0	0.0187	0.0117	0	0	0	0.0150	

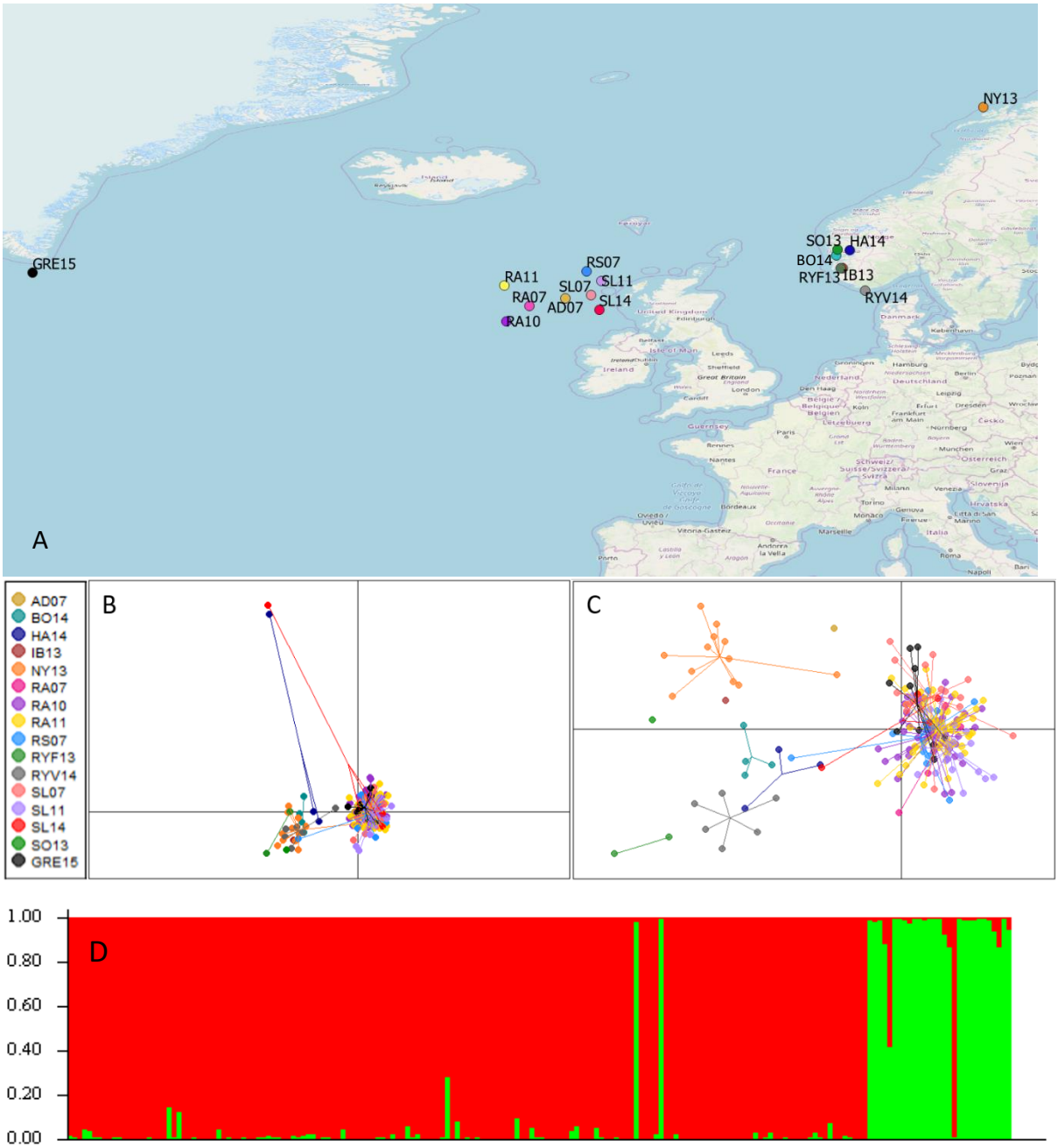

Figure 7: Population structure of the blue ling with; a) map showing coordinates for the locations of each population sample site; b) PCA; c) DAPC; and d) structure plot

Two clear clusters are formed in the PCA plot (Fig 7b). Within these fjord samples are grouped together in one cluster and Atlantic samples in the other. Deviating from these clusters there are two individuals stretching out as outliers from the two groups (Fig. 7b). One individual from Hardangerfjord (HA14/dark blue), located in the fjords, and the other located on the Rockall slope (SL14/red).

The DAPC also exhibits a similar pattern; here there is a very clear cluster of Atlantic populations formed, with the fjord samples separated from this (Fig. 7c). These fjords are then separated out from one another (Fig. 7c), which differs from the PCA (Fig. 7c). Possibly the most clearly separated is the Nygrunnen (NY13/orange) group, a fjord found a great distance from the rest in the North of Norway.

Outlying individuals within the DAPC plot should also be noted. From the Atlantic population there are two individuals clearly branching out towards the fjord populations to the bottom of the plot; these individuals originating from Rosemary Bank (RS07/light blue) and the Rockall slope (SL14/red) (Fig. 5c). This is the same Rockall slope individual seen branching out in the PCA, and although this same behaviour is not seen from the Hardangerfjord (HA14/dark blue) individual like in the PCA, the slope individual is stretching out to sit with this same individual in the DAPC plot.

Other deviating individuals are seen towards the top of the plot, with one individual stretching out from Nygrunnen (NY13/orange) more towards the Atlantic cluster, and with the one Anton Dohrn (AD07/beige) individual, an Atlantic population, sitting on its own between the Atlantic and Fjord groupings (Fig 7c). This may fit in with the F_{st} analysis (Table 12) in which this was the only Atlantic population found to significantly differ from the others.

With the *Structure* analysis of the blue ling there is more clear structure with the plot exhibiting a clear split between the Atlantic and the fjords, the likes of which are seen in the PCA and DAPC plots (Fig. 7d). In this plot the Atlantic group is to the left with most individuals here indicated by being mainly red in colour, and the fjords are found to the right of the plot with a mainly green colouring (Fig. 7d). There are also a couple of outliers in the *Structure* plot (Fig. 7d). The same two Atlantic outliers as those seen in the DAPC analysis, one individual from Rosemary Bank (RS07) and one individual from the slope (SL14), appear to fit better with the fjord individuals in this plot. These are seen as the two mostly green individuals sitting in the red Atlantic section of the plot (Fig. 7d). The fjord outliers are the Hardangerfjord (HA14) outlier seen in the PCA plots, and Ryvingen (RYV14), which doesn't show up as a deviation in any other analyses. These two outliers are indicated in the plot as the two redder individuals within the green fjord area of the plot (Fig. 7d).

Again, non-neutral analyses indicated the same structure and the analyses for the outlier SNPs did not add to this data. Both can be found in the appendix as supplementary material, along with a pairwise F_{st} table calculated using the outlier data for blue ling populations. This data, like the common ling, gave generally higher values of F_{st} but with so few outliers a high level of inaccuracy can be expected (Willing, Dreyer and van Oosterhout, 2012). Because of this the data is not presented in the main results.

4. Discussion

In this study we explore patterns of population structure of common and blue ling in the North East Atlantic. For this purpose, we used a panel of 6,566 SNPs for the common ling, and 3,073 SNPs for the blue ling, identified within this research project. Two distinct genetic groups were discovered within the blue ling which expanded on the weaker differentiation seen in the common ling, separating an open water Atlantic group with a coastal Norwegian fjords group. This appears to divide further in the blue ling, where weak differentiation is exhibited between fjords. It was possible to then further investigate the blue ling dataset and identify genes which could be driving adaptation between these groups. Our results show the presence of strong population structure in deep sea fish species using novel genomic markers; an important insight which contrasts with the belief of strong connectivity in the deep sea, and thus has implications for their suspected resilience (Cowen *et al.*, 2007; Baco *et al.*, 2016).

4.1. Common ling population structure

Multivariate analyses performed on the common ling samples show that there is some structure within this species (Fig. 6b & 6c). The Rockall populations (RA08 and RA14) deviate a little from the main cluster, indicating some differentiation. This is the main population sample taken from the Northern Atlantic, while the samples from the Bay of Biscay (BB15) are located within the main cluster falling in with the fjord populations. This could be indicative of a coastal species now expanding out into more open waters; indicated by the differentiation exhibited between the populations of the fjords and the Bay of Biscay (the coastal populations) clustering together versus the differentiated Rockall populations found further out into the Atlantic.

F_{st} values appear low (Table 11), as is seen often in marine relationships, so it can be difficult to infer from this data (Picq, Mcmillan and Puebla, 2016). We do find that the Bay of Biscay population (BB15) has the most consistently high (in relative terms) F_{st} values shared between other populations. This population falls into the main cluster with the fjords in the other analyses, so it is interesting that F_{st} picks up on something different here. Despite the low values across this data, confidence intervals calculate that most of these values are significantly different from zero. This may infer that each population, or most populations, are distinct from one another, if only at a minor level.

Although we find low values of F_{st} across our data, it is shown that genetic differentiation can still be found in marine populations with values such as these (Cano *et al.*, 2008). For example, a study on Atlantic salmon has shown low but significant genetic differentiation; finding significance in values of F_{st} = 0.018 (Aykanat *et al.*, 2015). This brings more significance to our data here, in which many populations share an F_{st} higher than this. It is also stated that management decisions based solely on these differentiation values would produce poor fisheries policies (Cano *et al.*, 2008). It is important that we combine this data with our other analyses, to create a broader understanding of structure to infer onto management.

These results can be directly compared to Gonzalez *et al.* (2015), with both studies indicating some differentiation between coastal and open water populations. For Gonzalez *et al.* (2015), a PCA indicated strong population structure splitting the coast and the open Atlantic, stronger than that found here. Although they did not find most pairwise F_{st} values to be significantly different from 0, they found their Rockall sample groups did significantly differ from other populations; mainly Norwegian coastal populations (Gonzalez *et al.*, 2015). The results from

this study back up the decision by Gonzalez *et al.* (2015) to reject the hypothesis of a single stock in the Northeast Atlantic.

The results of this study do not indicate structure as strong as that seen by Gonzalez *et al.* (2015) and this could be due to some shortcomings. There could be uncertainty here due to the low number of individuals overall. With only 83 individuals investigated, compared to the 647 of the Gonzalez *et al.* (2015) study, it is likely that this study is hindered by a low sample number, especially in a marine species. Genetic diversity can be underestimated if a sample size is too small, which could be the case with this data (Street *et al.*, 1998). Despite these weaknesses, the data is still able to show a weak split between the two groups and can be used for future management decisions alongside the Gonzalez *et al.* (2015) study.

4.2. Blue ling population structure

Here we present data for the first genetic study of the blue ling. Patterns of population structure are much clearer in this species. Looking at the PCA and DAPC results, there appears to be a clear split between the Atlantic and Norwegian fjord populations, into two separate genetic clusters (Fig. 7b & 7c). This would appear to indicate strong differentiation between the two groups. Moving on to the *Structure* plot we find the same pattern here, with a very clear split of the Atlantic and fjords (Fig. 7d).

It should also be noted that there are some outlying individuals found in these analyses. In the first PCA there are two very strong outliers who make it harder to see the overall pattern displayed in this plot (Fig. 7b). These individuals consist of one member of the Hardangerfjord population (HA14 in the fjords) and one from the slope population (SL14 in the Atlantic). Both

can be seen again grouped together in the DAPC (mainly the slope individual deviating), and again deviate from their clusters in *Structure* and *LEA* (Fig. 7c & 7d). Another notable outlier is an individual from Rosemary Bank in the Atlantic (RS07) who appears in the DAPC, *Structure*, and *LEA* plots (Fig. 7c & 7d).

Again, the F_{st} values calculated are all low as we can expect in marine species (Table 12). Despite finding these low values again, the data does appear to concur with the previous analyses. The stronger values are generally exhibited between Atlantic and fjord population relationships. This is also the case when it comes to the confidence intervals, with mostly Atlantic – fjord relationships being significantly different from 0. This data can again be combined with that of our other analyses to create a better understanding of the populations structure in this species. Many of these values are above the 0.018 found to be statistically significant (Aykanat *et al.*, 2015), and thus this could be taken into consideration when informing stakeholders.

Looking at the DAPC analysis (Fig. 7c), there appears to be some fine-scale structure found between the fjord sample sites. With fjords fairly separated out from one another, we could expect that there may be more here than can be seen in the broader analysis. Although less so than the Atlantic/fjord split, this indicates a degree of divergence between these areas, potentially indicating that isolation (due to genetic drift and low connectivity) has been acting on these populations for several generations. Most notable is the fjord located furthest to the North, Nygrunnen (NY13), which separates out from the other fjords (Fig. 7c). Geographically this site is a great distance from the other fjord sample locations, which are more grouped together.

These results are comparable to Atlantic cod, where increasing SNP variation was found with increasing geographic separation of fjords along the Norwegian coast (Kirubakaran *et al.*, 2016). Such differentiation is not found in every species, for example European sprat (*Sprattus sprattus*) has been found to show a striking lack of genetic divergence across the Norwegian fjords (Quintela *et al.*, 2020). This highlights the significance of this finding in the blue ling. This data suggests that there may be some differentiation between the fjords, just not at the same kind of scale as that seen between the fjords and the Atlantic. Further study looking into this would be beneficial, and help inform future management efforts for fjord populations.

Overall, our analyses join to reveal a clear structural pattern, in which the Atlantic samples cluster together to form one population, and fjord samples cluster to create another. This data gives our first insight into the population genetic structure of this species, and it is important that this is now taken into consideration for the species' management. To improve on this knowledge, further analysis within the fjords should be done to determine the extent of isolation between these locations and the individuals within them.

4.3. Comparison of the two species

These results are of particular interest considering that we find more substantial structure in the deeper dwelling of our two species. It is thought to be more common that deeper species will exhibit greater connectivity (Baco *et al.*, 2016), and here we see the opposite of this. The structural pattern shown indicates a stronger separation of the fjords and the Atlantic for the deeper of the two species. A similar pattern to this has also been found in Pacific cod (Drinan *et al.*, 2018). Here, coastal samples were compared with inland samples and two

differentiated populations were exhibited, with no hybridisation between the coastal and inland individuals (Drinan *et al.*, 2018). This could be suggestive of local adaptation, as the lack of mixing among these individuals in very different environments could allow for adaptive evolution to take place.

These results could also be suggesting that the blue ling has moved into a new niche/environment, adapting to this. Typically seen as a deep-sea species, they may have started inhabiting the fjords and began adapting to its many environmental differences to the ocean. As this presents a completely different environment to the species, adaptation will thus occur through natural selection (Bernatchez, 2016; Wilson *et al.*, 2019).

With the common ling we may be seeing hints of the opposite behaviour; A more coastal species moving out to deeper/more open waters. The individuals found at Rockall deviating from the other populations hints at this, and again it may be that these individuals are adapting to a different environment.

For both species we have this differentiation between coastal and open water populations, comparable to that which was found with the common ling in the past (Gonzalez *et al.*, 2015). The differentiation exhibited from the Rockall samples is particularly notable, and fits with that found by Gonzalez *et al.* (2015) for the common ling. It also follows a trend found in other species, for example saithe (*Pollachius virens*) was found to split into 4 genetic cluster across the Northern Atlantic; one of which being located around Rockall (Saha *et al.*, 2015). This is also seen in tusk (*Brosme brosme*) where the Rockall population is considered a stock unit (ICES, 2014).

It should also be noted, however, that in the blue ling analyses the individuals around Rockall fall into the same cluster as those from Greenland, which is a great distance to the West and

thus they are greatly geographically separated from one another. This implies that Rockall does not stick out as a singular stock in the blue ling, but falls into a broader North Atlantic stock. Knowing this, it would be of great interest to further investigate the common ling and see if this pattern continues into the sister species. This information would prove valuable for assessing stocks.

These results are extremely important as they give a greater insight into the interactions of deep-sea species which are currently understudied. Combining such studies creates stronger evidence that there is differentiation within the deep sea, particularly between more coastal populations and those found in more open water (Gonzalez *et al.*, 2015). We see here that we cannot just assume that all deep-sea populations are completely connected. It is best to be cautionary with the exploitation of these species and carry out more studies to increase our understanding of the deep sea.

4.4. Outliers and adaptation

Such a low number of outlier loci found in the data may be partially explained by the nature of marine populations, and the great connectivity experienced by them (Nickols *et al.*, 2015). This could be seen as an advantage or a disadvantage, as lower levels of differentiation may mean that the outliers selected are only the strongest, most informative outliers. As a disadvantage, it could indicate a lack of adaptative divergence and not be so informative. A low number of outliers, however, is common in marine populations and should not be considered in the context of terrestrial studies (Lal, Southgate, Jerry, Bosserelle, *et al.*, 2016; Carreras *et al.*, 2017; Feutry *et al.*, 2017; Jansson *et al.*, 2017; Mathiesen *et al.*, 2017).

To explore outliers within the datasets we used a combination of methods to ensure each SNP picked out was a true outlier, with no false positives (Milano *et al.*, 2014; Lal, Southgate, Jerry, Bosserelle, *et al.*, 2016; Coscia *et al.*, 2019). Choosing out only those SNPs which were calculated by both methods gave few SNPs for either species, but these could then be explored further. Outlier loci are believed to be influenced by selective processes and to have an effect on the fitness of an individual (Nayfa and Zenger, 2016). By identifying outliers, we are able to get a better insight into the micro-evolutionary forces affecting populations, and possibly driving differentiation between them (Nayfa and Zenger, 2016). Using outliers we can identify genes which are possible targets of local selection, but environmental data is needed to confirm such data and give a greater insight (Pujolar *et al.*, 2014).

The outliers found here act as highly informative loci and may be associated with adaptive divergence between the populations found in this study (Russello *et al.*, 2011; Funk *et al.*, 2016). This research is limited due to a lack of environmental data, which means that local adaptation cannot be identified in this case. It can, however, be speculated and here the adaptive possibilities are explored and discussed through the genes found to be linked with outlier loci. Future studies should make use of landscape genomics in order to identify local adaptation.

There are three main factors which are believed to influence successful adaptation; generation time, population size, and population structure (Bernatchez, 2016). Here we may expect to find adaptive evolution linked with population structure, causing the variation found between the Atlantic and the fjords.

Any genetic links found are greatly important as they can inform us of adaptive evolution occurring within the species/populations. In the past, local adaptation has mostly only been

possible to investigate in model organisms (Savolainen, Lascoux and Merilä, 2013). Thanks to improving genomic tools, it is now possible to investigate adaptive divergence more reliably (Savolainen, Lascoux and Merilä, 2013). With this expectation for signals of adaptation to be found within populations who are living in different habitats, it can be assumed that differing depths will have an influence on this. Such distinct environments are known to drive local adaptation, causing reproductive isolation and ecological speciation (Daníelsdóttir *et al.*, 2008). Coastal/depth related adaptation has been discussed in multiple species (Fyhn *et al.*, 1994; Daníelsdóttir *et al.*, 2008; Cadrin *et al.*, 2010).

4.5. Adaption to varying light environments

When searching for genes linked with our outliers, there were significant matches found with all of the blue ling loci and none with the common ling. With a total of 24 genes we found some trends which could be indicative of adaptation within this species. The first interesting link found is that with the opsin protein, found twice in our searches (Tables 6 & 10). This is a gene commonly seen to adapt to varying light environments, affecting the sensitivity of photoreceptors (Pampoulie *et al.*, 2015; Luehrmann *et al.*, 2018). In fish these adaptations can be due to depth or water colour/turbidity variations and can be short-term (altering gene expression within the life of an individual) or long-term (affecting the evolution of the species) (Luehrmann *et al.*, 2018). In this study there are two possibilities, in that environmental variations in the Atlantic and fjords could be driving the change, or it may be more depth related as this is the deeper of our two species. Depth changes may still be linked with the Atlantic/fjord split as we know the fjords to be more restricting with depth, for example Hardangerfjord with depths up to 890m (Buhl-mortensen and Buhl-mortensen, 2014),

Boknafjord limited to 550m depths (Stensvold and Minoretti, 2015), and Sørfjorden with a maximum depth of 390m (Beyer *et al.*, 1996). Since the blue ling has a maximum depth range of around 1000m (Cohen *et al.*, 1990; Papasissi, 2020), these fjords therefore limit this species. Links with depth seem likely, as only the short wavelength spectrum reaches down through 200m to 1000m depths (Pampoulie *et al.*, 2015; de Busserolles and Marshall, 2017; Lin *et al.*, 2017), and short wavelength sensitive opsins being those found connected with our outliers.

Research has shown that, in fishes, opsin genes have evolved in three different ways to adapt to the dimly lit marine environment (Chang and Yan, 2019). These include gene duplication, mutations, and plasticity of expression (Chang and Yan, 2019). The fact that we find a match for the same gene for two of our outliers located at different areas of the genome can thus be explained by gene duplication (Cortesi *et al.*, 2014; Chang and Yan, 2019). This has been previously witnessed for short wavelength sensitive opsin in percomorph fishes (Cortesi *et al.*, 2014). These duplications allow gained sensitivity to different wavelengths and can also be differentially expressed through development (Cortesi *et al.*, 2014). It is possible that the two environments in which we find our blue ling (Atlantic and fjords) require adaptations to different light sensitivities. There are many variables in an environment which can influence these needs; as well as light conditions relating to depth, individuals must look out for visual cues such as those for foraging or avoiding predators (Luehrmann *et al.*, 2018). For example, environmental colour can affect skin pigmentation (Costa *et al.*, 2017), and thus another visual cue which is specific to varying habitats is born. It is possible that such factors would vary between the Atlantic and fjords, thus altering the visual needs of the populations within and leading to adaptive divergence. This could especially be the case for the *Molva* genus,

who are more dependent on visual stimuli in their feeding strategies than other gadoid species (Løkkeborg, Skajaa and Fernö, 2000).

Recently, a study identified a trend in deep sea fish to independently adapt and expand their number of opsin genes (Musilova *et al.*, 2019). This was found after inspecting 101 fish genomes and discovering this development within 3 different deep sea species (Musilova *et al.*, 2019). This makes it even more likely that the blue ling may also have adapted to this environment and expanded its opsin genes.

Similarly, local adaptation for the light pigment rhodopsin has been witnessed in Atlantic cod (Pampoulie *et al.*, 2015). This has been linked with behaviours which exposed the divergent groups to differing light levels (Pampoulie *et al.*, 2015). Combining this with the prior mentioned knowledge it seems likely that alterations to the opsin genes are driven by differences in the two environments our samples are found in. Whether this is through direct changes to light levels via depth, or from behavioural changes coming from environmental differences such as food availability or predators, is unknown.

4.6. Adaption to colder environments

Another trend seen throughout our outlier loci was a link with genes relating to growth and development. This could suggest morphological and size variations between the two populations of blue ling. One of the recurring genetic matches is with Hox clusters (Table 9), which decipher the body plan of a species at the embryonic stages (Pan *et al.*, 2016). An example of one of the functions these are responsible for is that of forming limb buds (Kelley *et al.*, 2016). This also links with another gene found (responsible for producing retinoic acid

receptor gamma) which has been linked with limb bud movement (Table 10). This could indicate that differing environmental variables within the two populations habitats are selecting for morphological traits suited more specifically to them. These environments could require different swimming capabilities (Kirk *et al.*, 2016). This has similarly been found in hamlet fish species, where Hox genes appear to be indicative of differentiation and speciation (Picq, Mcmillan and Puebla, 2016).

Temperature is possibly one of the most important effectors of growth and development in marine fish (Laurel *et al.*, 2016). At a species' optimal temperature, high growth rates have been recorded (Sandersfeld, Mark and Knust, 2017). This has inferences for changing temperatures in the future, which could affect marine species by deviating from the temperatures which they have evolved to live in. It has been suggested that Arctic cod are highly vulnerable to changing temperatures, with negative impacts on growth when increasing this factor (Laurel *et al.*, 2016). There has also been a pattern witnessed in Arctic charr showing higher growth and metabolic rates in colder, more Northernly waters (Niloshini Sinnatamby *et al.*, 2015). This is thought to be linked with phenotypic plasticity and genetic variation (Niloshini Sinnatamby *et al.*, 2015). It seems likely from this that the genetic variation linking the many growth genes with our outlier SNPs is due to differences in environmental temperatures.

Metabolic rates are often linked with growth (Munday, Donelson and Domingos, 2017; Sandersfeld, Mark and Knust, 2017). Apolipoprotein E, involved in the metabolism of fats, is another protein affected by our outliers which may also have links with colder climates (Table 7). There is evidence that links changing metabolic rates, and thus differing energetic costs, with adaptation to habitat temperatures (Sandersfeld, Mark and Knust, 2017). This is even

more relevant to our data considering another genetic link with one of our outliers, with a gene encoding for glucose transporter 1 (Table 7). With these 2 linking genes affected by our outlier SNPs it would seem to back up the theory that there may be divergence occurring within the species due to thermal differences in the environment.

This potential change of lipid metabolism may also be linked back to depth adaptations. It was found in a deep-sea snailfish that changes to genes linking with fatty acid metabolism could be aiding the membrane fluidity/flexibility of cells; helping the species to cope with the high pressures of the deep ocean (Wang *et al.*, 2019). This could again suggest an environmentally forced change in depth of the species between the Atlantic and fjords.

Again, linking with colder environments, we have outliers matching with genes encoding for multiple antifreeze glycoproteins (Tables 7 & 10). These proteins are already a very specified adaptation for organisms living in colder environments, which have evolved independently in Northern and Southern hemisphere fishes (Baalsrud *et al.*, 2018). This adaptation is essential for fish species to survive in sub-zero waters (Yamazaki *et al.*, 2018). Activity levels of these genes will differ depending on the environment, with great variability exhibited among species within a genus (Yamazaki *et al.*, 2019). It seems plausible that this can then go to population level, considering that the influencing factor on this variable is the environmental temperature/condition. Mutations within the genes which produce antifreeze glycoproteins would suggest that our species is living within varying temperatures, which is driving selection of these genes. This could come about due to the colder waters around Norway, compared to warmer deep Atlantic waters (Gordon, 2001). It is therefore likely that this is one of the factors driving divergence between our 2 populations.

There are even more inferences to depth adaptation from these links with antifreeze glycoproteins. In species living in deeper waters there is a reduction in antifreeze activity due to the absence of ice nuclei (Yamazaki *et al.*, 2019). If the blue ling located in the fjords are adapting to shallower waters, then selection would drive for increased activity of these genes.

Together these give us an indication for climate driven adaptation in the blue ling. Considering the population pattern exhibited in our analyses, this would suggest that there is a difference in temperature for the Atlantic and fjords which is driving some divergence of these genes. Warming of the ocean is happening fast, for example there are areas of the Arctic with surface temperatures increasing by 0.5°C per decade (Laurel *et al.*, 2016). This has inferences for the future; with climate change looming causing temperatures to rise, it would appear possible for fish to adapt to changing temperatures and survive. Despite this ability to adapt, the rate of warming seen in the ocean may be too rapid to allow for such changes to occur resulting in a great potential threat to the species. Increasing habitat temperatures can affect growth and reproduction, which may in turn lead to altered population structures and divergence (Sandersfeld, Mark and Knust, 2017). It is essential that we study these effects more to understand how changing temperatures in the future will impact fish species in the marine environment (Laurel *et al.*, 2016)

It is important that these do not appear to be phenotypically plastic changes to our populations, as they are altering the genetic code. This shows that the two populations are genetically deviating from one another. It appears that the environment of the Atlantic and fjords are largely diverse from one another; enough to drive adaptation in the blue ling.

4.7. Linking temperature with immune adaptations

Antifreeze glycoproteins have also been shown to have anti-infective properties (Fikrig *et al.*, 2016), which ties in with some of our other genes linking with outliers. Here we found the PAPPA gene (Table 8), encoding for pappalysin which links with wound healing and bone remodelling, and a Raftlin-like pseudogene (Table 9), the functional gene of which produces Raftlin which activates signalling of T and B cells upon contact with bacterial lipopolysaccharides. These could be linked, and indicative of increased damage inflicted on individuals, and thus increased susceptibility to bacterial infection. This could come about through increased infection of parasites; something which has been linked with warmer temperatures (Schade, Raupach and Wegner, 2016; Franke *et al.*, 2017; Strepparava *et al.*, 2018). Linking with earlier theories of cold adaptation, this may be a case of differences between the Atlantic and the fjords. It also seems possible that the more enclosed nature of the fjords would allow for increased infection of parasites, and thus possible selection to adapt to this.

There is also the possibility of local adaptation within the parasites themselves, which can allow for increased compatibility to a host (Weber *et al.*, 2017). Host specificity is high in parasites, as they evolve with their host species (Klapper *et al.*, 2017). This is the case with Monogenean species, who have shown higher success in colder habitats (Klapper *et al.*, 2017), contrasting to the higher temperatures found to be a better fit for endoparasites (Franke *et al.*, 2017; Strepparava *et al.*, 2018). This seems the more likely route to drive such adaptation, as Monogenean ectoparasites leave flesh wounds on the skin of fish which are then open to infection.

Any variation in this area could also be depth linked. It has been found that ectoparasite diversity and host specificity declines with increasing depths (Quattrini and Demopoulos, 2016). This, again, points to differences found between the Atlantic and fjords. There could be multiple factors causing a need for change and adaptation towards healing and immune defences, specified for different environments.

The immune system of Teleost fishes has been found to be highly adaptive and links immune-related genes to speciation rates within these species (Malmstrøm *et al.*, 2016). This flexibility can be linked with gene losses and expansions, acting as responses to changing environments (Solbakken *et al.*, 2016, 2017). Such rapid adaptation has been witnessed in response to increased pollution levels (Reid *et al.*, 2016). It has been suggested that a high nucleotide diversity is likely essential to drive this swift adaptation (Reid *et al.*, 2016). This is another possibility within our species, in that pollution levels may vary between the Atlantic and the fjords.

4.8. Memory adaptations

Adaptations towards memory requirements have been described in Teleost species, for example that of long-term social memory (Madeira and Oliveira, 2017). Here, we have found a possible adaptation within memory requirements with a gene responsible for controlling synaptic plasticity and memory function (Table 8). Knowing that social recognition is present in these fishes could suggest an increased/decreased need for this function in one of our populations. It could be indicative of an increased memory capacity within the fjords, where a more confined area may allow for increased repeat interactions between individuals.

Another possibility is for different functional requirements between the environments due to varying factors such as that of survival (Andersen *et al.*, 2016). These changes can be adapted to through natural selection for instincts and responses, or through learning and remembering from experience (Andersen *et al.*, 2016). It could be possible that one of the environments in which we find our populations is more demanding in this sense, and thus leading to a change in requirements.

Memory is also used in a spatial and navigational sense, and fish live in complex and unstable environments in which strong spatial memory is essential (Mcaroe, Craig and Holland, 2016). The capability for this across species is varied (Mcaroe, Craig and Holland, 2016) and thus probably adapted to each individually. This is another aspect which could be driving adaption within the blue ling between our two environments.

Significant differences in learning and memory have been witnessed across fish populations, and it has been suggested that various interacting factors drive these functions (Roy and Bhat, 2018). This brings together the different ideas for what could drive such an adaptation and indicates that many variances found between our environments may lead to adaptation within memory functions. More understanding and greater research is needed here, where few studies have investigated memory in fish species, and greater focus has been applied to mammal or bird species (Mcaroe, Craig and Holland, 2016; Madeira and Oliveira, 2017).

4.9. Adaptation to seawater

Linking back to adaptations affecting growth of our species, there are further functions of the growth hormone genes found linked with our outliers. As well as their role in growth control,

the somatotropin proteins (Table 9) are also known to improve salinity tolerance (Bystriansky *et al.*, 2017). Higher levels of this hormone have been found when increasing salinity levels under experimental conditions (Semenova, Pritvorova and Krayushkina, 2018). Therefore, adaptation involving these genes could indicate differing salinity of the environment these fish are living in. We may expect that our fjords, being more inland, have a lesser salt content and thus do not require as high a tolerance to this factor. This could lead to variation between our populations; with a selection for this in the Atlantic individuals and not in fjord individuals.

4.10. Adaptation comparisons between species

All of the blue ling outlier SNPs were investigated in the common ling; to check the levels of polymorphism and see if the same genes affected in the blue ling are also experiencing selection in the common ling. Sequencing data was found for 4 out of 5 of the blue ling SNPs in the common ling, and there was a complete lack of polymorphism at every site. This shows clearly that the genes found to be potentially driving differentiation within the blue ling are not experiencing these same effects within the common ling. This may link back to population structure and be explained by the less distinct structure exhibited in the common ling data.

This also has inferences for adaptation at different depths. This data would suggest that greater adaptive divergence is found in deeper-dwelling species, as has been seen elsewhere in deep sea species where an increased role of adaptation has been found with occurrence at a faster rate than would be expected elsewhere (Gaither *et al.*, 2016; Porter, Roberts and Partridge, 2016; K. Wang *et al.*, 2019; Sutton and Milligan, 2019). This could be due to the extreme environmental pressures experienced by deep sea species, pushing them to their physical limits (Sutton and Milligan, 2019). Much of the evidence points towards depth

differences between the Atlantic and fjords, which would then be further driven by a lack of connectivity experienced between the 2 populations in these locations.

4.11. Conclusion

This study seems to tell us a story of changing movements in two sister species, who are possibly moving into new environments and adapting to this. We can confidently say that there is strong differentiation between the Atlantic and fjord populations of the blue ling, with a lack of connectivity between these coastal versus open water individuals. Structure is less clear within the common ling, but suggests differentiation is seen from the more open water individuals from the Rockall populations, compared with the remaining coastal populations.

Marine sampling is often opportunistic (Miller *et al.*, 2016), and on this occasion the sampling numbers for the common ling are not quite as high. This data then fails to find strong indicators for structure within the species. From this I would suggest future research, with greater numbers. This would allow further exploration of the structural patterns within the common ling, hinted at within this study. This would clarify the length to which coastal and open water populations truly differentiate from one another.

Despite this, we do have the Gonzalez *et al*. (2015) study to compare with this common ling data. In the prior study there was a pattern indicated which showed similarities to that which was found here in the common ling and the blue ling; with the more open water Atlantic populations differing from individuals located along the Norwegian coast (Gonzalez *et al.*, 2015). This strengthens the results found here in which the Rockall populations differ from

the remaining groups. It is comparable to studies of *Maccullochella peelii* (Murray cod), and brown trout, in which SNP and microsatellite analyses detected similar structural patterns (Harrisson *et al.*, 2017; Saint-pé *et al.*, 2019).

Genetic and genomic analyses have been useful in informing management and conservation plans for wild populations over the last half century (Mamoozadeh, Graves and Mcdowell, 2019). This data is therefore of great importance for management and can now be taken into consideration for future planning. For example, SNP analyses of the European hake were able to confirm a split between Northern and Southern populations, which were the current management units for the species (Leone *et al.*, 2019). It also went further to identify additional structure within the Northern group, finding Norwegian individuals to be differentiated from the rest of the group (Leone *et al.*, 2019). Due to these findings, it was suggested that stock units should be redefined to include the Norwegian population as its own unit (Leone *et al.*, 2019). These findings are comparable to those found in this study, especially for the blue ling.

Due to less studies concerning structure of deep sea species, there is a lack of examples in which such data has been incorporated into their management. Because of this, and the poorly understood nature of deep sea species, more effective management is needed in the deep sea (Gonçalves da Silva *et al.*, 2019). For the commercially exploited deep sea orange roughy, genomic analyses identified population structuring which could be used for consideration of deep sea management in an area where the species was previously believed to be panmictic (Gonçalves da Silva *et al.*, 2019). Finding such structure in deep sea species has greater inferences for fishing in the deep, as it shows that such populations may not be as connected as previously thought. Combining the orange roughy data with that of the

common ling and the blue ling shows clearly that structure in the deep sea is underestimated. These studies should now aid in the development of proper management of deep sea resources.

We should use this as an indicator for further research and look at more commercially fished deep sea species to understand their structure and the impacts that fishing may have on them. More caution is needed when fishing in the deep sea and we should continue to study this area. Stock assessments for the common ling and the blue ling should be carried out with consideration of this data, and management enforced in correspondence with this.

From this data we suggest that the blue ling should now be considered as at least 2 separate stock units in the Northeast Atlantic; one for the Norwegian fjords and one for the rest of the Northeast Atlantic. We suggest that differentiation between the fjords should be further investigated to determine any further structuring, and that the fjords to the North should be taken into consideration for a separate management plan. We agree with Gonzalez *et al.* (2015) that the common ling should be considered as 2 stock units; one coastal and one open water. Both species would benefit from continued research, assessing the species across the Atlantic to the West to determine further stock units. This would also be beneficial to determine whether or not the common ling Rockall population is a separate stock in its own right, or whether these individuals fall in with the rest of the Atlantic. This is of particular interest considering the separation seen in other studies (ICES, 2014; Gonzalez *et al.*, 2015; Saha *et al.*, 2015).

5. References

Albaina, A. *et al.* (2013) 'Single nucleotide polymorphism discovery in albacore and Atlantic bluefin tuna provides insights into worldwide population structure', *Animal Genetics*, 44(6), pp. 678–692. doi: 10.1111/age.12051.

Allen, B. *et al.* (2015) 'The molecular clock of neutral evolution can be accelerated or slowed by asymmetric spatial structure', *PLoS Computational Biology*, 11(2), pp. 1–32. doi: 10.1371/journal.pcbi.1004108.

Allendorf, F. W. (2017) 'Genetics and the conservation of natural populations: allozymes to genomes', *Molecular Ecology*, 26(2), pp. 420–430. doi: 10.1111/mec.13948.

Altschul, S. F. *et al.* (1997) 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic Acids Research*, 25(17), pp. 3389–3402. doi: 10.1093/nar/25.17.3389.

Andersen, B. S. *et al.* (2016) 'The proximate architecture for decision-making in fish', *Fish and Fisheries*, 17, pp. 680–695. doi: 10.1111/faf.12139.

Andrews, K. R. *et al.* (2016) 'Harnessing the power of RADseq for ecological and evolutionary genomics', *Nature Reviews Genetics*, 17(2), pp. 81–92. doi: 10.1038/nrg.2015.28.Harnessing.

Andrews, S. (2010) 'FastQC: A quality control tool for high throughput sequence data.' Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Aykanat, T. *et al.* (2015) 'Low but significant genetic differentiation underlies biologically meaningful phenotypic divergence in a large Atlantic salmon population', *Molecular Ecology*, 24(20), pp. 5158–5174. doi: 10.1111/mec.13383.

Baalsrud, H. T. *et al.* (2018) 'De novo gene evolution of antifreeze glycoproteins in codfishes revealed by whole genome sequence data', *Molecular Biology and Evolution*, 35(3), pp. 593–606. doi: 10.1093/molbev/msx311.

Baco, A. R. *et al.* (2016) 'A synthesis of genetic connectivity in deep-sea fauna and implications for marine reserve design', *Molecular ecology*, 25(14), pp. 3276–3298. doi: 10.1111/mec.13689.

Bagley, M. J., Lindquist, D. G. and Geller, J. B. (1999) 'Microsatellite variation, effective population size, and population genetic structure of vermilion snapper, *Rhomboplites aurorubens*, off the southeastern USA', *Marine Biology*, 134(4), pp. 609–620. doi: 10.1007/s002270050576.

De Barba, M. *et al.* (2016) 'High-throughput microsatellite genotyping in ecology: Improved accuracy, efficiency, standardisation and success with low-quantity and degraded DNA', *Molecular Ecology Resources*, 17(3), pp. 492–507. doi: 10.1111/1755-0998.12594.

Barnes, M. K. S. (2008) *Molva dypterygia Blue ling*, *Marine Life Information Network: Biology and Sensitive Key Information Reviews*. Available at: https://www.marlin.ac.uk/species/detail/101 (Accessed: 14 January 2020).

Baxter, J. M. (2001) 'Establishing management schemes on marine special areas of conservation in Scotland', *Aquatic Conservation: Marine and Freshwater Ecosystems*, 11(4), pp. 261–265.

Bernardi, G. *et al.* (2016) 'Genomic signatures of rapid adaptive evolution in the bluespotted cornetfish, a Mediterranean Lessepsian invader', *Molecular Ecology*, 25, pp. 3384–3396. doi: 10.1111/mec.13682.

Bernatchez, L. (2016) 'On the maintenance of genetic variation and adaptation to environmental change: Considerations from population genomics in fishes', *Journal of Fish Biology*, 89(6), pp. 2519–2556. doi: 10.1111/jfb.13145.

Bernatchez, L. *et al.* (2017) 'Harnessing the power of genomics to secure the future of seafood', *Trends in Ecology & Evolution*. Elsevier Ltd, 32(9), pp. 665–680. doi: 10.1016/j.tree.2017.06.010.

Beyer, J. *et al.* (1996) 'Contaminant accumulation and biomarker responses in flounder (*Platichthys flesus L*.) and Atlantic cod (*Gadus morhua L*.) exposed by caging to polluted sediments in Sørfjorden, Norway', *Aquatic Toxicology*, 36, pp. 75–98.

Bolyen, E. *et al.* (2018) 'An introduction to applied bioinformatics: A free, open, and interactive text.', *Journal of Open Source Education*, 1(5), pp. 1–10. doi: 10.1109/EMBC.2016.7590696.Upper.

Bosse, M. *et al.* (2017) 'Recent natural selection causes adaptive evolution of an avian polygenic trait', *Science*, 358, pp. 365–368.

Bougeard, S. and Dray, S. (2018) 'Supervised multiblock analysis in *R* with the **ade4** package', *Journal of Statistical Software*, 86(1). doi: 10.18637/jss.v086.i01.

Bradbury, I. R. *et al.* (2013) 'Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish', *Evolutionary Applications*, 6(3), pp. 450–461. doi: 10.1111/eva.12026.

Bradbury, P. J. *et al.* (2007) 'TASSEL: Software for association mapping of complex traits in diverse samples', *Bioinformatics*, 23(19), pp. 2633–2635. doi: 10.1093/bioinformatics/btm308.

Buhl-mortensen, P. and Buhl-mortensen, L. (2014) 'Diverse and vulnerable deep-water biotopes in the Hardangerfjord', *Marine Biology Research*. Taylor & Francis, 10(3), pp. 253–267. doi: 10.1080/17451000.2013.810759.

Bulut, I. *et al.* (2018) 'Pregnancy-associated plasma protein-A (PAPP-A) levels in patients with severe allergic asthma are reduced by omalizumab', *Journal of Asthma*. Taylor & Francis, 55(10). doi: 10.1080/02770903.2017.1396471.

de Busserolles, F. and Marshall, N. J. (2017) 'Seeing in the deep-sea: Visual adaptations in lanternfishes', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1717). doi: 10.1098/rstb.2016.0070.

Bystriansky, J. S. *et al.* (2017) 'Salinity acclimation and advanced parr–smolt transformation in growth-hormone transgenic coho salmon (*Oncorhynchus kisutch*)', *Canadian Journal of Zoology*, 95(9), pp. 633–643. doi: 10.1139/cjz-2016-0201.

Cadrin, S. X. *et al.* (2010) 'Population structure of beaked redfish, *Sebastes mentella*: Evidence of divergence associated with different habitats', *ICES Journal of Marine Science*, 67(8), pp. 1617–1630. doi: 10.1093/icesjms/fsq046.

Cano, J. M. *et al.* (2008) 'Genetic differentiation, effective population size and gene flow in marine fishes: Implications for stock management', *Journal of integrated field science*, 5, pp. 1–10.

Carreras, C. *et al.* (2017) 'Population genomics of an endemic Mediterranean fish: Differentiation by fine scale dispersal and adaptation', *Scientific Reports*, 7. doi: 10.1038/srep43417.

Casillas, S. and Barbadilla, A. (2017) 'Molecular population genetics', *Genetics*, 205(3), pp.

1003–1035. doi: 10.1534/genetics.116.196493.

Catchen, J. M. *et al.* (2013) 'Stacks: An analysis tool set for population genomics', *Molecular Ecology*, 22(11), pp. 3124–3140. doi: 10.1111/mec.12354.Stacks.

Chang, C.-H. and Yan, H. Y. (2019) 'Plasticity of opsin gene expression in the adult red shiner (*Cyprinella lutrensis*) in response to turbid habitats', *Plos One*, 14(4), p. e0215376. doi: 10.1371/journal.pone.0215376.

Chung-Jung, L. *et al.* (2018) '*Helicobacter pylori* infection-induced hepatoma-derived growth factor regulates the differentiation of human mesenchymal stem cells to myofibroblast-like cells', *Cancers*, 10(12). doi: 10.3390/cancers10120479.

Clarke, J. *et al.* (2015) 'A scientific basis for regulating deep-sea fishing by depth', *Current Biology*, 25(18), pp. 2425–2429. doi: 10.1016/j.cub.2015.07.070.

Cohen, D. M. *et al.* (1990) *Vol.10. Gadiform fishes of the world (Order Gadiformes)*. Rome, Italy: FAO.

Coll, M. *et al.* (2016) 'Ecological indicators to capture the effects of fishing on biodiversity and conservation status of marine ecosystems', *Ecological Indicators*, 60, pp. 947–962. doi: 10.1016/j.ecolind.2015.08.048.

Collie, J. *et al.* (2017) 'Indirect effects of bottom fishing on the productivity of marine fish', *Fish and Fisheries*, 18(4), pp. 619–637. doi: 10.1111/faf.12193.

Cormier, F. *et al.* (2019) 'A reference high-density genetic map of greater yam (*Dioscorea alata L.*)', *Theoretical and Applied Genetics*, 132(6), pp. 1733–1744. doi: 10.1007/s00122-019-03311-6.

Cortesi, F. *et al.* (2014) 'Ancestral duplications and highly dynamic opsin gene evolution in percomorph fishes', *Proceedings of the National Academy of Sciences*, 112(5), pp. 1493–1498. doi: 10.1073/pnas.1417803112.

Coscia, I. *et al.* (2019) 'Fine-scale seascape genomics of an exploited marine species, the common cockle', *bioRxiv*, pp. 1–44.

Costa, D. C. *et al.* (2017) 'The effect of environmental colour on the growth, metabolism, physiology and skin pigmentation of the carnivorous freshwater catfish *Lophiosilurus alexandri*', *Journal of Fish Biology*, 90(3), pp. 922–935. doi: 10.1111/jfb.13208.

Costello, C. *et al.* (2016) 'Global fishery prospects under contrasting management regimes', *Proceedings of the National Academy of Sciences*, 113(18), pp. 5125–5129. doi: 10.1073/pnas.1520420113.

Cowen, R. K. *et al.* (2000) 'Connectivity of marine populations: Open or closed?', *Science*, 287(5454), pp. 857–859. doi: 10.1126/science.287.5454.857.

Cowen, R. K. *et al.* (2007) 'Population connectivity in marine systems: An overview', *Oceanography*, 20(3), pp. 14–21.

Cuéllar-pinzón, J. *et al.* (2016) 'Genetic markers in marine fisheries: Types, tasks and trends', *Fisheries Research*, 173(2016), pp. 194–205.

Danecek, P. *et al.* (2011) 'The variant call format and VCFtools', *Bioinformatics*, 27(15), pp. 2156–2158. doi: 10.1093/bioinformatics/btr330.

Daníelsdóttir, A. K. *et al.* (2008) 'Population structure of deep-sea and oceanic phenotypes of deepwater redfish in the Irminger Sea and Icelandic continental slope: Are they cryptic species?', *Transactions of the American Fisheries Society*, 137(6), pp. 1723–1740. doi:

10.1577/t07-240.1.

Davey, J. W. *et al.* (2011) 'Genome-wide genetic marker discovery and genotyping using next-generation sequencing', *Nature Reviews Genetics*, 12(7), pp. 499–510. doi: 10.1038/nrg3012.

Dray, S. and Dufour, A.-B. (2015) ' The ade4 package: Implementing the duality diagram for ecologists ', *Journal of Statistical Software*, 22(4). doi: 10.18637/jss.v022.i04.

Drinan, D. P. *et al.* (2018) 'Population assignment and local adaptation along an isolation-by-distance gradient in Pacific cod (*Gadus macrocephalus*)', *Evolutionary Applications*, 11(8), pp. 1448–1464. doi: 10.1111/eva.12639.

Duforet-Frebourg, N., Bazin, E. and Blum, M. G. B. (2014) 'Genome scans for detecting footprints of local adaptation using a Bayesian factor model', *Molecular Biology and Evolution*, 31(9), pp. 2483–2495. doi: 10.1093/molbev/msu182.

Earl, D. A. and vonHoldt, B. M. (2012) 'STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method', *Conservation Genetics Resources*, 4(2), pp. 359–361. doi: 10.1007/s12686-011-9548-7.

Elmer, K. R. (2016) 'Genomic tools for new insights to variation, adaptation, and evolution in the salmonid fishes: A perspective for charr', *Hydrobiologia*. Springer International Publishing, 783(1), pp. 191–208. doi: 10.1007/s10750-015-2614-5.

Elshire, R. J. *et al.* (2011) 'A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species', *PLoS ONE*, 6(5), pp. 1–10. doi: 10.1371/journal.pone.0019379.

Elvarsson, B. Þ. *et al.* (2018) 'Pushing the limits of a data challenged stock: A size- and age-structured assessment of ling (*Molva molva*) in Icelandic waters using Gadget', *Fisheries*

Research, 207, pp. 95–109. doi: 10.1016/j.fishres.2018.06.005.

Etter, R. J. and Bower, A. S. (2015) 'Dispersal and population connectivity in the deep North Atlantic estimated from physical transport processes', *Deep-Sea Research Part I*, 104, pp. 159–172. doi: 10.1016/j.dsr.2015.06.009.

Fernández-Pérez, J. *et al.* (2018) 'Mitochondrial DNA analyses of *Donax trunculus* (Mollusca: Bivalvia) population structure in the Iberian Peninsula, a bivalve with high commercial importance', *Aquatic Conservation: Marine and Freshwater Ecosystems*, 28(5), pp. 1139–1152. doi: 10.1002/aqc.2929.

Feutry, P. *et al.* (2017) 'Inferring contemporary and historical genetic connectivity from juveniles', *Molecular Ecology*, 26(2), pp. 444–456. doi: 10.1111/mec.13929.

Fikrig, E. *et al.* (2016) 'Anti-infective properties of antifreeze proteins', *United States Patent Application Publication*, 1(19).

Fischer, M. C. *et al.* (2011) 'Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*)', *Molecular ecology*, 20, pp. 1450–1462. doi: 10.1111/j.1365-294X.2011.05015.x.

Foll, M. *et al.* (2010) 'Estimating population structure from AFLP amplification intensity', *Molecular Ecology*, 19(21), pp. 4638–4647. doi: 10.1111/j.1365-294X.2010.04820.x.

Foll, M. (2012) 'BayeScan v2.1 user manual', *Ecology*, pp. 1450–1462.

Foll, M. and Gaggiotti, O. (2008) 'A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective', *Genetics*, 180(2), pp. 977–993. doi: 10.1534/genetics.108.092221.

Foo, S. A. and Byrne, M. (2016) *Acclimatization and Adaptive Capacity of Marine Species in a Changing Ocean*., *Advances in Marine Biology*. 1st edn. Elsevier Ltd. doi: 10.1016/bs.amb.2016.06.001.

Franke, F. *et al.* (2017) 'Environmental temperature variation influences fitness trade-offs and tolerance in a fish-tapeworm association', *Parasites and Vectors*, 10(252), pp. 1–11. doi: 10.1186/s13071-017-2192-7.

Frichot, E. and Francois, O. (2015) 'LEA: An R package for landscape and ecological association studies', *Methods in Ecology and Evolution*, 6, pp. 925–929. doi: 10.1111/2041-210X.12382.

Funk, W. C. *et al.* (2016) 'Adaptive divergence despite strong genetic drift: Genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (*Urocyon littoralis*)', *Molecular Ecology*, 25(10), pp. 2176–2194. doi: 10.1111/mec.13605.

Furukawa, H. *et al.* (2005) 'Subunit arrangement and function in NMDA receptors', *Nature*, 438, pp. 185–192. doi: 10.1038/nature04089.

Fyhn, U. E. H. *et al.* (1994) 'New variants of the haemoglobins of Atlantic cod: A tool for discriminating between coastal and Arctic cod populations', *ICES Marine Science Symposia*, 198(0), pp. 666–670.

Gagnaire, P. A. *et al.* (2015) 'Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era', *Evolutionary Applications*, 8(8), pp. 769–786. doi: 10.1111/eva.12288.

Gaither, M. R. *et al.* (2016) 'Molecular phylogenetics and evolution depth as a driver of evolution in the deep sea: Insights from grenadiers (Gadiformes : Macrouridae) of the genus

Coryphaenoides', *Molecular Phylogenetics and Evolution*, 104, pp. 73–82. doi: 10.1016/j.ympev.2016.07.027.

Genomics Core Leuvin (2020) *Genomics Core Leuvin*. Available at: genomicscore.be (Accessed: 22 June 2020).

Gonçalves da Silva, A. *et al.* (2019) 'Genomic data suggest environmental drivers of fish population structure in the deep sea: A case study for the orange roughy (*Hoplostethus atlanticus*)', *Journal of Applied Ecology*, pp. 1–11. doi: 10.1111/1365-2664.13534.

Gonzalez, E. B. *et al.* (2015) 'Genetic analyses of ling (*Molva molva*) in the Northeast Atlantic reveal patterns relevant to stock assessments and management advice', *ICES Journal of Marine Science*, 72(2), pp. 635–641.

Gordon, J. D. M. (2001) 'Deep-water fisheries at the Atlantic Frontier', *Continental Shelf Research*, 21, pp. 987–1003.

Gosselin, T. (2019) 'radiator: RADseq data exploration, manipulation and visualization using R.' Available at: https://thierrygosselin.github.io/radiator/.

Goudet, J. (2005) 'Hierfstat, a package for R to compute and test hierarchical F-statistics', *Molecular Ecology Notes*, 5, pp. 184–186. doi: 10.1111/j.1471-8278.

Greminger, M. P. *et al.* (2014) 'Generation of SNP datasets for orangutan population genomics using improved reduced-representation sequencing and direct comparisons of SNP calling algorithms', *BMC Genomics*, 15(1), pp. 1–15. doi: 10.1186/1471-2164-15-16.

Harrisson, K. A. *et al.* (2017) 'Signatures of polygenic adaptation associated with climate across the range of a threatened fish species with high genetic connectivity', *Molecular Ecology*, 26, pp. 6253–6269. doi: 10.1111/mec.14368.

Hedgecock, D., Barber, P. H. and Edmands, S. (2007) 'Genetic approaches to measuring connectivity', *Oceanography*, 20(3), pp. 70–79.

Helle, K. *et al.* (2019) 'Development of SNP for the deep-sea fish blue ling, *Molva dypterygia* (Pennant , 1784) from ddRAD sequencing data', *Conservation Genetics Resources*. Springer Netherlands, 7. doi: 10.1007/s12686-019-01107-w.

Herrington, J. and Carter-Su, C. (2001) 'Signaling pathways activated by the growth hormone receptor', *Trends in Endocrinology and Metabolism*, 12(6), pp. 252–257.

Humble, E., Martinez-Barrio, A., *et al.* (2016) 'A draft fur seal genome provides insights into factors affecting SNP validation and how to mitigate them', *Molecular ecology resources*, 16(4), pp. 909–921. doi: 10.1111/1755-0998.12502.

Humble, E., Thorne, M. A. S., *et al.* (2016) 'Transcriptomic SNP discovery for custom genotyping arrays: Impacts of sequence data, SNP calling method and genotyping technology on the probability of validation success', *BMC Research Notes*, 9(1), pp. 1–12. doi: 10.1186/s13104-016-2209-x.

ICES (2014) *Advice May 2014 Widely distributed and migratory stocks Tusk (Brosme brosme) in the Northeast Atlantic*.

ICES (2017) *Blue ling (Molva dypterygia) in subareas 1, 2, 8, 9, and 12, and in divisions 3.a and 4.a (other areas)*. doi: 10.17895/ices.pub.3056.

Ikami, K. *et al.* (2015) 'Hierarchical differentiation competence in response to retinoic acid ensures stem cell maintenance during mouse spermatogenesis', *Development*, 142(9), pp. 1582–1592. doi: 10.1242/dev.118695.

Illumina (2020) *Illumina*. Available at: illumina.com (Accessed: 22 June 2020).

Janjua, S. *et al.* (2020) 'Improving our conservation genetic toolkit: ddRAD-seq for SNPs in snow leopards', *Conservation Genetics Resources*, 12(2), pp. 257–261. doi: 10.1007/s12686-019-01082-2.

Jansson, E. *et al.* (2017) 'Genetic analysis of goldsinny wrasse reveals evolutionary insights into population connectivity and potential evidence of inadvertent translocation via aquaculture', *ICES Journal of Marine Science*, 74(8), pp. 2135–2147. doi: 10.1093/icesjms/fsx046.

Jombart, T. (2008) 'Adegenet: A R package for the multivariate analysis of genetic markers', *Bioinformatics*, 24(11), pp. 1403–1405. doi: 10.1093/bioinformatics/btn129.

Jombart, T. and Ahmed, I. (2011) 'adegenet 1.3-1: New tools for the analysis of genome-wide SNP data', *Bioinformatics*, 27(21), pp. 3070–3071. doi: 10.1093/bioinformatics/btr521.

Jones, F. C. *et al.* (2012) 'Report a genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks', *Current Biology*, 22(1), pp. 83–90. doi: 10.1016/j.cub.2011.11.045.

Junge, C. *et al.* (2019) 'Comparative population genomics confirms little population structure in two commercially targeted carcharhinid sharks', *Marine Biology*, 166(2), pp. 1–15. doi: 10.1007/s00227-018-3454-4.

Junutula, J. R. *et al.* (2004) 'Rab14 is involved in membrane trafficking between the Golgi Complex and Endosomes', *Molecular Biology of the Cell*, 15(5), pp. 2218–2229. doi: 10.1091/mbc.E03.

Kagale, S. *et al.* (2016) 'Analysis of Genotype-by-Sequencing (GBS) data', in *Plant Bioinformatics*. doi: 10.1007/978-3-319-67156-7.

Kamvar, Z. N., Brooks, J. C. and Grünwald, N. J. (2015) 'Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality', *Frontiers in Genetics*, 6, pp. 1–10. doi: 10.3389/fgene.2015.00208.

Kamvar, Z. N., Tabima, J. F. and Grünwald, N. J. (2014) 'Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction ', *PeerJ*, 2, p. e281. doi: 10.7717/peerj.281.

Keane, J. A. *et al.* (2016) 'SNP-sites: Rapid efficient extraction of SNPs from multi-FASTA alignments', *Microbial Genomics*, 2(4), pp. 1–5. doi: 10.1099/mgen.0.000056.

Keenan, K. *et al.* (2013) 'DiveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors', *Methods in Ecology and Evolution*, 4(8), pp. 782–788. doi: 10.1111/2041-210X.12067.

Kelley, J. L. *et al.* (2016) 'The life aquatic: Advances in marine vertebrate genomics', *Nature Reviews Genetics*, 17(9), pp. 523–534. doi: 10.1038/nrg.2016.66.

Kirk, M. A. *et al.* (2016) 'Effects of water velocity, turbulence and obstacle length on the swimming capabilities of adult Pacific lamprey', *Fisheries Management and Ecology*, 23, pp. 356–366. doi: 10.1111/fme.12179.

Kirubakaran, T. G. *et al.* (2016) 'Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod', *Molecular Ecology*, 25(10), pp. 2130–2143. doi: 10.1111/mec.13592.

Klapper, R. *et al.* (2017) *Biodiversity and Evolution of Parasitic Life in the Southern Ocean*. Edited by S. Klimpel. Frankfurt/Main, Germany: Springer International Publishing Switzerland. doi: 10.1007/978-3-319-46343-8_1.

Knaus, B. J. and Grünwald, N. J. (2017) 'VcfR: A package to manipulate and visualize variant call format data in R', *Molecular Ecology Resources*, 17(1), pp. 44–53. doi: 10.1111/1755-0998.12549.

Knutsen, H. *et al.* (2009) 'Bathymetric barriers promoting genetic structure in the deepwater demersal fish tusk (*Brosme brosme*)', *Molecular Ecology*, 18(15), pp. 3151–3162.

Kumar, G. and Kocour, M. (2017) 'Applications of next-generation sequencing in fisheries research: A review', *Fisheries Research*, 186, pp. 11–22. doi: 10.1016/j.fishres.2016.07.021.

Lal, M. M., Southgate, P. C., Jerry, D. R., Bosserelle, C., *et al.* (2016) 'A parallel population genomic and hydrodynamic approach to fishery management of highly-dispersive marine invertebrates: The case of the Fijian black- lip pearl oyster *Pinctada margaritifera*', *PLoS ONE*, 11(8). doi: 10.1371/journal.pone.0161390.

Lal, M. M., Southgate, P. C., Jerry, D. R. and Zenger, K. R. (2016) 'Marine Genomics Fishing for divergence in a sea of connectivity: The utility of ddRADseq genotyping in a marine invertebrate, the black-lip pearl oyster', *Marine Genomics*, 25, pp. 57–68. doi: 10.1016/j.margen.2015.10.010.

Lamichhaney, S. *et al.* (2017) 'Parallel adaptive evolution of geographically distant herring populations on both sides of the North Atlantic Ocean', *Proceedings of the National Academy of Sciences*, 114(17), pp. 3452–3461. doi: 10.1073/pnas.1617728114.

Langmead, B. and Slazberg, S. L. (2013) 'Fast gapped-read alignmnet with Bowtie 2', *Nature methods*, 9(4), pp. 357–359. doi: 10.1038/nmeth.1923.Fast.

Large, P. A. *et al.* (2010) 'Spatial and temporal distribution of spawning aggregations of blue ling (*Molva dypterygia*) west and northwest of the British Isles', *ICES Journal of Marine*

Science, 67, pp. 494–501.

Laurel, B. J. *et al.* (2016) 'Temperature-dependent growth and behavior of juvenile Arctic cod (*Boreogadus saida*) and co-occurring North Pacific gadids', *Polar Biology*. Springer Berlin Heidelberg, 39(6), pp. 1127–1135. doi: 10.1007/s00300-015-1761-5.

Lee, R. *et al.* (1999) 'Role of Nr13 in regulation of programmed cell death in the bursa of Fabricius', *Genes and Development*, 13, pp. 718–728.

Leone, A. *et al.* (2019) 'Genome-wide SNP based population structure in European hake reveals the need for harmonizing biological and management units', *ICES Journal of Marine Science*, 76(7), pp. 2260–2266. doi: 10.1093/icesjms/fsz161.

Levy, S. E. and Myers, R. M. (2016) 'Advancements in Next-Generation Sequencing', *Annual Review of Genomics and Human Genetics*, 17(1), pp. 95–115. doi: 10.1146/annurev-genom-083115-022413.

Li, F. and Tsien, J. Z. (2009) 'Memory and the NMDA Receptors', *The New England Journal of Medicine*, 361, pp. 302–303.

Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.

Li, H. (2011) 'A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data', *Bioinformatics*, 27(21), pp. 2987–2993. doi: 10.1093/bioinformatics/btr509.

Li, H. and Durbin, R. (2010) 'Fast and accurate long-read alignment with Burrows-Wheeler transform', *Bioinformatics*, 26(5), pp. 589–595. doi: 10.1093/bioinformatics/btp698.

Li, R. *et al.* (2009) 'SOAP2: An improved ultrafast tool for short read alignment', *Bioinformatics*, 25(15), pp. 1966–1967. doi: 10.1093/bioinformatics/btp336.

Lin, J. J. *et al.* (2017) 'The rises and falls of opsin genes in 59 ray-finned fish genomes and their implications for environmental adaptation', *Scientific Reports*, 7(1), pp. 1–13. doi: 10.1038/s41598-017-15868-7.

Løkkeborg, S., Skajaa, K. and Fernö, A. (2000) 'Food-search strategy in ling (*Molva molva L.*): Crepuscular activity and use of space', *Journal of Experimental Marine Biology and Ecology*, 247(2), pp. 195–208. doi: 10.1016/S0022-0981(00)00148-9.

Luehrmann, M. *et al.* (2018) 'Short-term colour vision plasticity on the reef: Changes in opsin expression under varying light conditions differ between ecologically distinct fish species', *The Journal of Experimental Biology*, 221(22), p. jeb175281. doi: 10.1242/jeb.175281.

Luna, S. M. (2020a) *Lota lota*, *Fishbase*.

Luna, S. M. (2020b) *Molva molva*, *Fishbase*. Available at: fishbase.se/summary/Molva-molva.html (Accessed: 14 January 2020).

Luu, K., Bazin, E. and Blum, M. G. B. (2017) 'pcadapt: An R package to perform genome scans for selection based on principal component analysis', *Molecular Ecology Resources*, 17(1), pp. 67–77. doi: 10.1111/1755-0998.12592.

Madeira, N. and Oliveira, R. F. (2017) 'Long-Term Social Recognition Memory in Zebrafish', *Zebrafish*, 14(4), pp. 305–310. doi: 10.1089/zeb.2017.1430.

Malmstrøm, M. *et al.* (2016) 'Evolution of the immune system influences speciation rates in teleost fishes', *Nature Genetics*, 48(10), pp. 1204–1210. doi: 10.1038/ng.3645.

Malmstrøm, M. *et al.* (2017) 'Data descriptor: Whole genome sequencing data and de novo draft assemblies for 66 teleost species', *Scientific Data*, 4, pp. 1–13. doi: 10.1038/sdata.2016.132.

Malomane, D. K. *et al.* (2018) 'Efficiency of different strategies to mitigate ascertainment bias when using SNP panels in diversity studies', *BMC Genomics*, 19(22), pp. 1–16. doi: 10.1186/s12864-017-4416-9.

Mamoozadeh, N. R., Graves, J. E. and Mcdowell, J. R. (2019) 'Genome-wide SNPs resolve spatiotemporal patterns of connectivity within striped marlin (*Kajikia audax*), a broadly distributed and highly migratory pelagic species', *Evolutionary Applications*, pp. 1–22. doi: 10.1111/eva.12892.

Mangi, S. C. *et al.* (2016) 'The economic implications of changing regulations for deep sea fishing under the European Common Fisheries Policy: UK case study', *Science of the Total Environment*. Elsevier B.V., 562, pp. 260–269. doi: 10.1016/j.scitotenv.2016.03.218.

Marais, A. D. (2019) 'Apolipoprotein E in lipoprotein metabolism, health and cardiovascular disease', *Pathology*, 51(2), pp. 165–176. doi: 10.1016/j.pathol.2018.11.002.

Marshall, J., Carleton, K. L. and Cronin, T. (2015) 'Colour vision in marine organisms', *Current Opinion in Neurobiology*, 34, pp. 86–94. doi: 10.1016/j.conb.2015.02.002.

Martínez, J. G. *et al.* (2017) 'SNPs markers for the heavily overfished tambaqui *Colossoma macropomum*, a Neotropical fish, using next-generation sequencing-based de novo genotyping', *Conservation Genetics Resources*, 9(1), pp. 29–33. doi: 10.1007/s12686-016-0610-3.

Mathiesen, S. S. *et al.* (2017) 'Genetic diversity and connectivity within *Mytilus spp.* in the

subarctic and Arctic', *Evolutionary Applications*, 10(1), pp. 39–55. doi: 10.1111/eva.12415.

Matsumoto, M. and Tatematsu, M. (2017) 'Cell type-specific role of raftlin in the regulation of endosomal TLR signaling', *Inflammation and Cell Signalling*, 4, pp. 1–8. doi: 10.14800/ics.1326.

Mcaroe, C. L., Craig, C. M. and Holland, R. A. (2016) 'Place versus response learning in fish: A comparison between species', *Animal Cognition*, 19(1), pp. 153–161. doi: 10.1007/s10071-015-0922-9.

Mielczarek, M. and Szyda, J. (2016) 'Review of alignment and SNP calling algorithms for next-generation sequencing data', *Journal of Applied Genetics*, 57(1), pp. 71–79. doi: 10.1007/s13353-015-0292-7.

Milano, I. *et al.* (2014) 'Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*)', *Molecular Ecology*, 23, pp. 118–135. doi: 10.1111/mec.12568.

Miller, A. D. *et al.* (2016) 'Contrasting patterns of population connectivity between regions in a commercially important mollusc *Haliotis rubra*: Integrating population genetics, genomics and marine LiDAR data', *Molecular ecology*, 25(16), pp. 3845–3864. doi: 10.1111/mec.13734.

Miller, K. J. and Gunasekera, R. M. (2017) 'A comparison of genetic connectivity in two deep sea corals to examine whether seamounts are isolated islands or stepping stones for dispersal', *Nature Scientific Reports*, 7(46103), pp. 1–14. doi: 10.1038/srep46103.

Mirciov, C. S. G. *et al.* (2017) 'Characterization of putative erythroid regulators of hepcidin in mouse models of anemia', *PLoS ONE*, 12(1). doi: 10.1371/journal.pone.0171054.

Mullins, R. B. *et al.* (2018) 'Genomic analysis reveals multiple mismatches between biological and management units in yellowfin tuna (*Thunnus albacares*)', *ICES Journal of Marine Science*, 75(6), pp. 2145–2152. doi: 10.1093/icesjms/fsy102.

Munday, P. L., Donelson, J. M. and Domingos, J. A. (2017) 'Potential for adaptation to climate change in a coral reef fish', *Global Change Biology*, 23(1), pp. 307–317. doi: 10.1111/gcb.13419.

Musilova, Z. *et al.* (2019) 'Vision using multiple distinct rod opsins in deep-sea fishes', *Science*, 364(6440), pp. 588–592.

Nayfa, M. G. and Zenger, K. R. (2016) 'Marine genomics unravelling the effects of gene fl ow and selection in highly connected populations of the silver-lip pearl oyster (*Pinctada maxima*)', *Marine Genomics*, 28, pp. 99–106. doi: 10.1016/j.margen.2016.02.005.

Nevado, B., Ramos-Onsins, S. E. and Perez-Enciso, M. (2014) 'Resequencing studies of nonmodel organisms using closely related reference genomes: Optimal experimental designs and bioinformatics approaches for population genomics', *Molecular Ecology*, 23(7), pp. 1764–1779. doi: 10.1111/mec.12693.

Nickols, K. J. *et al.* (2015) 'Marine population connectivity: Reconciling large-scale dispersal and high self-retention', *The American Naturalist*, 185(2), pp. 196–211. doi: 10.1086/679503.

Niloshini Sinnatamby, R. *et al.* (2015) 'Latitudinal variation in growth and otolith-inferred field metabolic rates of Canadian young-of-the-year Arctic charr', *Ecology of Freshwater Fish*, 24(3), pp. 478–488. doi: 10.1111/eff.12166.

O'Leary, S. J. *et al.* (2018) 'These aren't the loci you'e looking for: Principles of effective SNP

filtering for molecular ecologists', *Molecular Ecology*, pp. 3193–3206. doi: 10.1111/mec.14792.

Ogbe, R. J., Ochalefu, D. O. and Olaniru, O. B. (2016) 'Bioinformatics advances in genomics – A review', *International journal of current pharmaceutical review and research*, 8(10), pp. 5–11. Available at: http://www.scopemed.org/?mno=231738.

Pampoulie, C. *et al.* (2015) 'Rhodopsin gene polymorphism associated with divergent light environments in Atlantic cod', *Behavior genetics*, 45(2), pp. 236–244. doi: 10.1007/s10519-014-9701-7.

Pan, H. *et al.* (2016) 'The genome of the largest bony fish, ocean sunfish (*Mola mola*), provides insights into its fast growth rate', *GigaScience*, pp. 1–12. doi: 10.1186/s13742-016-0144-3.

Papasissi, C. (2020) *Molva dypterygia*, *Fishbase*. Available at: fishbase.se/summary/Molva-dypterygia.html (Accessed: 14 January 2020).

Paradis, E. (2010) 'Pegas: An R package for population genetics with an integrated-modular approach', *Bioinformatics*, 26(3), pp. 419–420. doi: 10.1093/bioinformatics/btp696.

Paris, J. R., Stevens, J. R. and Catchen, J. M. (2017) 'Lost in parameter space: A road map for stacks', *Methods in Ecology and Evolution*, 8(10), pp. 1360–1373. doi: 10.1111/2041-210X.12775.

Pauly, D. and Zeller, D. (2016) 'Catch reconstructions reveal that global marine fisheries catches are higher than reported and declining', *Nature Communications*, 7, pp. 1–9. doi: 10.1038/ncomms10244.

Pecoraro, C. *et al.* (2018) 'The population genomics of yellowfin tuna (*Thunnus albacares*) at

global geographic scale challenges current stock delineation', *Scientific Reports*, 8(1), pp. 1–10. doi: 10.1038/s41598-018-32331-3.

Pennimpede, T. *et al.* (2010) 'Analysis of Cyp26b1/Rarg compound-null mice reveals two genetically separable effects of retinoic acid on limb outgrowth', *Developmental Biology*. Elsevier Inc., 339(1), pp. 179–186. doi: 10.1016/j.ydbio.2009.12.024.

Picq, S., Mcmillan, W. O. and Puebla, O. (2016) 'Population genomics of local adaptation versus speciation in coral reef fishes (*Hypoplectrus spp*, Serranidae)', *Ecology and Evolution*, 6(7), pp. 2109–2124. doi: 10.1002/ece3.2028.

Pierotti, M. E. R. *et al.* (2017) 'Rapid and parallel adaptive evolution of the visual system of neotropical midas cichlid fishes', *Molecular Biology and Evolution*, 34(10), pp. 2469–2485. doi: 10.1093/molbev/msx143.

Porter, M. L., Roberts, N. W. and Partridge, J. C. (2016) 'Evolution under pressure and the adaptation of visual pigment compressibility in deep-sea environments', *Molecular Phylogenetics and Evolution*. doi: 10.1016/j.ympev.2016.08.007.

Pritchard, J. K., Stephens, M. and Donnelly, P. (2000) 'Inference of population structure using multilocus genotype data', *Genetics*, 155(2).

Pujolar, J. M. *et al.* (2014) 'Genome-wide single-generation signatures of local selection in the panmictic European eel', *Molecular Ecology*, 23(10), pp. 2514–2528. doi: 10.1111/mec.12753.

Purcell, S. *et al.* (2007) 'PLINK: A tool set for whole-genome association and population-based linkage analyses.', *American journal of human genetics*, 81(3), pp. 559–75. doi: 10.1086/519795.

QGIS Development Team (2019) ‘QGIS geographic information system. Open source geospatial foundation project.’ Available at: http://qgis.osgeo.org.

QIAGEN (2006) *DNeasy Blood & Tissue Handbook*. doi: 10.1111/ele.12937.

Quattrini, A. M. and Demopoulos, A. W. J. (2016) ‘Ectoparasitism on deep-sea fishes in the western North Atlantic: In situ observations from ROV surveys’, *International Journal for Parasitology: Parasites and Wildlife*, 5(3), pp. 217–228. doi: 10.1016/j.ijppaw.2016.07.004.

Quintela, M. *et al.* (2020) ‘Genetic analysis redraws the management boundaries for the European sprat’, *Evolutionary Applications*, Early View. doi: 10.1111/eva.12942.

R Core Team (2017) ‘R: A language and environment for statistical computing’. Vienna, Austria: R Foundation for Statistical Computing.

Raj, A., Stephens, M. and Pritchard, J. K. (2014) ‘FastSTRUCTURE: Variational inference of population structure in large SNP data sets’, *Genetics*, 197(2), pp. 573–589. doi: 10.1534/genetics.114.164350.

Reid, N. M. *et al.* (2016) ‘The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish’, *Science*, 354(6317), pp. 1305–1309.

Reinert, K. *et al.* (2015) ‘Alignment of Next-Generation Sequencing Reads’, *Annual Review of Genomics and Human Genetics*, 16(1), pp. 133–151. doi: 10.1146/annurev-genom-090413-025358.

Rennison, D. J. *et al.* (2016) ‘Rapid adaptive evolution of colour vision in the threespine stickleback radiation’, *Proceedings of the Royal Society B*, 283.

Riginos, C. *et al.* (2016) ‘Navigating the currents of seascape genomics: How spatial analyses

can augment population genomic studies', *Current Zoology*, 62(6), pp. 581–601. doi: 10.1093/cz/zow067.

Ring, A. K. *et al.* (2009) 'Development of 10 microsatellite loci in the ling (*Molva molva*)', *Molecular Ecology Resources*, 9(5), pp. 1401–1403. doi: 10.1111/j.1755-0998.2009.02677.x.

Rodríguez-Ezpeleta, N. *et al.* (2016) 'Population structure of Atlantic mackerel inferred from RAD-seq-derived SNP markers: Effects of sequence clustering parameters and hierarchical SNP selection', *Molecular ecology resources*, 16, pp. 991–1001. doi: 10.1111/1755-0998.12518.

Rohfritsch, A. *et al.* (2018) 'Preliminary insights into the genetics of bank vole tolerance to *Puumala hantavirus* in Sweden', *Ecology and Evolution*, 8(22), pp. 11273–11292. doi: 10.1002/ece3.4603.

Rosenberg, A. A. *et al.* (2018) 'Applying a new ensemble approach to estimating stock status of marine fisheries around the world', *Conservation Letters*, 11(1), pp. 1–9. doi: 10.1111/conl.12363.

Van Rossum, G. and Drake Jr, F. L. (1995) 'Python reference manual'. Centrum voor Wiskunde en Informatica Amsterdam.

Rowley, S. J. (2008) *Molva molva Ling.*, *Marine Life Information Network: Biology and Sensitive Key Information Reviews*. Available at: https://www.marlin.ac.uk/species/detail/10 (Accessed: 14 January 2020).

Roy, T. and Bhat, A. (2018) 'Divergences in learning and memory among wild zebrafish: Do sex and body size play a role?', *Learning and Behaviour*, 46, pp. 124–133. doi: 10.3758/s13420-017-0296-8.

Ruperao, P. and Edwards, D. (2015) 'Bioinformatics: Identification of markers from next-generation sequence data', *Plant Genotyping: Methods and Protocols*, pp. 29–47. doi: 10.1007/978-1-4939-1966-6.

Russello, M. A. *et al.* (2011) 'Detection of outlier loci and their utility for fisheries management', *Evolutionary Applications*, 5(1), pp. 39–52. doi: 10.1111/j.1752-4571.2011.00206.x.

Sá-Pinto, A. *et al.* (2012) 'Barriers to gene flow in the marine environment: Insights from two common intertidal limpet species of the Atlantic and Mediterranean', *PLoS ONE*, 7(12). doi: 10.1371/journal.pone.0050330.

Sage Science (2020) *Sage Science*. Available at: sagescience.com (Accessed: 22 June 2020).

Saha, A. *et al.* (2015) 'Seascape genetics of saithe (*Pollachius virens*) across the North Atlantic using single nucleotide polymorphisms', *ICES Journal of Marine Science*, 72(9), pp. 2732–2741.

Saha, A. *et al.* (2017) 'Geographic extent of introgression in *Sebastes mentella* and its effect on genetic population structure', *Evolutionary Applications*, 10(1), pp. 77–90. doi: 10.1111/eva.12429.

Saint-pé, K. *et al.* (2019) 'Development of a large SNPs resource and a low-density SNP array for brown trout (*Salmo trutta*) population genetics'. BMC Genomics, pp. 1–13.

Sandersfeld, T., Mark, F. C. and Knust, R. (2017) 'Temperature-dependent metabolism in Antarctic fish: Do habitat temperature conditions affect thermal tolerance ranges?', *Polar Biology*, 40(1), pp. 141–149. doi: 10.1007/s00300-016-1934-x.

Santos, S. *et al.* (2012) 'Isolation by distance and low connectivity in the peppery furrow

shell *Scrobicularia plana* (Bivalvia)', *Marine Ecology Progress Series*, 462, pp. 111–124. doi: 10.3354/meps09834.

Savolainen, O., Lascoux, M. and Merilä, J. (2013) 'Ecological genomics of local adaptation', *Nature Reviews Genetics*, 14, pp. 807–820.

Schade, F. M., Raupach, M. J. and Wegner, K. M. (2016) 'Seasonal variation in parasite infection patterns of marine fish species from the Northern Wadden Sea in relation to interannual temperature fluctuations', *Journal of Sea Research*, 113, pp. 73–84. doi: 10.1016/j.seares.2015.09.002.

Selkoe, K. A. *et al.* (2016) 'A decade of seascape genetics: Contributions to basic and applied marine connectivity', *Marine Ecology Progress Series*, 554, pp. 1–19. doi: 10.3354/meps11792.

Semenova, O. G., Pritvorova, A. V and Krayushkina, L. S. (2018) 'Changes of somatotropin concentration in blood serum of juvenile Russian sturgeon *Acipenser gueldenstaedtii* (Acipenseriformes) during adaptation to hyperosmotic medium', *Journal of Ichthyology*, 58(2), pp. 265–268. doi: 10.1134/S0032945218020133.

Shafer, A. B. A. *et al.* (2017) 'Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference', *Methods in Ecology and Evolution*, 8, pp. 907–917. doi: 10.1111/2041-210X.12700.

Sherman, C. D. H. *et al.* (2016) 'What are we missing about marine invasions? Filling in the gaps with evolutionary genomics', *Marine Biology*, 163(10), pp. 1–24. doi: 10.1007/s00227-016-2961-4.

Shimo, T. *et al.* (2019) 'Chondrocyte-specific gene expression', *In Vivo*, 33(1), pp. 85–91. doi:

10.21873/invivo.11443.

Siccha-Ramirez, Z. R. *et al.* (2018) 'SNP identification and validation on genomic DNA for studying genetic diversity in *Thunnus albacares* and *Scomberomorus brasiliensis* by combining RADseq and long read high throughput sequencing', *Fisheries Research*, 198, pp. 189–194. doi: 10.1016/j.fishres.2017.09.002.

Solbakken, M. H. *et al.* (2016) 'Successive losses of central immune genes characterize the Gadiformes' alternate immunity', *Genome Biology and Evolution*, 8(11), pp. 3508–3515. doi: 10.1093/gbe/evw250.

Solbakken, M. H. *et al.* (2017) 'Linking species habitat and past palaeoclimatic events to evolution of the teleost innate immune system', *Proceedings of the Royal Society B*, 284(20162810).

Song, J. Y. *et al.* (2020) 'Hox genes maintain critical roles in the adult skeleton', *Proceedings of the National Academy of Sciences*, 117(13), pp. 7296–7304. doi: 10.1073/pnas.1920860117.

Stensvold, B. and Minoretti, A. (2015) *Long Span Bridge in Norway*.

Stocks, K. (2009) *Seamounts Online: an online information system for seamount biology*. Available at: http://seamounts.sdsc.edu (Accessed: 14 January 2020).

Street, G. T. *et al.* (1998) 'Reduced genetic diversity in a meiobenthic copepod exposed to a xenobiotic', *Journal of Experimental Marine Biology and Ecology*, 222, pp. 93–111.

Strepparava, N. *et al.* (2018) 'Temperature-related parasite infection dynamics: The case of proliferative kidney disease of brown trout', *Parasitology*, 145(3), pp. 281–291. doi: 10.1017/S0031182017001482.

Sustainable Fisheries Partnership (2016) *Ling NE Atlantic nei*, *FishSource*. Available at: https://www.fishsource.org/stock_page/2014 (Accessed: 14 January 2020).

Sutton, T. and Milligan, R. J. (2019) 'Aquatic ecology: Deep-sea ecology', in Fath, B. (ed.) *Encyclopedia of Ecology*. 2nd edn. Amsterdam, Netherlands: Elsevier, pp. 35–45.

Thorvaldsdóttir, H., Robinson, J. T. and Mesirov, J. P. (2013) 'Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration', *Briefings in Bioinformatics*, 14(2), pp. 178–192. doi: 10.1093/bib/bbs017.

Toriello, H. V. (2016) 'What is new in genetics and genomics?', in *Health Care for People with Intellectual and Developmental Disabilities Across the Lifespan*, pp. 1–2307. doi: 10.1007/978-3-319-18096-0.

Torkamaneh, D., Laroche, J. and Belzile, F. (2016) 'Genome-wide SNP calling from genotyping by sequencing (GBS) data: A comparison of seven pipelines and two sequencing technologies', *PLoS ONE*, 11(8), pp. 1–14. doi: 10.1371/journal.pone.0161333.

Tsuda, S. *et al.* (2020) 'Fish-derived antifreeze proteins and antifreeze glycoprotein exhibit a different ice-binding property with increasing concentration', *Biomolecules*, 10(3).

Valenzuela-Quiñonez, F. (2016) 'How fisheries management can benefit from genomics?', *Briefings in Functional Genomics*, 15(5), pp. 352–357. doi: 10.1093/bfgp/elw006.

Varela, A. I., Ritchie, P. A. and Smith, P. J. (2013) 'Global genetic population structure in the commercially exploited deep-sea teleost orange roughy (*Hoplostethus atlanticus*) based on microsatellite DNA analyses', *Fisheries Research*. Elsevier B.V., 140, pp. 83–90. doi: 10.1016/j.fishres.2012.12.011.

Verhelst, P. *et al.* (2016) 'Acoustic telemetry as a tool for cod stock assessment', in Degraer,

S. et al. (eds) *North Sea Open Science Conference 2016: Abstract Booklet*. Brussels, Belgium: Royal Belgian Institute of Natural Sciences; Belgian Biodiversity Platform.

Victorero, L. *et al.* (2018) 'Out of sight, but within reach: A deep-sea fisheries from >400m global history of bottom-trawled depth', *Frontiers in Marine Science*, 5, pp. 1–17. doi: 10.3389/fmars.2018.00098.

de Villemereuil, P. and Gaggiotti, O. E. (2015) 'A new FST-based method to uncover local adaptation using environmental variables', *Methods in Ecology and Evolution*, 6(11), pp. 1248–1258. doi: 10.1111/2041-210X.12418.

Wang, K. *et al.* (2019) 'Morphology and genome of a snailfish from the Mariana Trench provide insights into deep-sea adaptation', *Nature Ecology and Evolution*, 3, pp. 823–833. doi: 10.1038/s41559-019-0864-8.

Wang, L. *et al.* (2019) 'Glucose transporter 1 critically controls microglial activation through facilitating glycolysis', *Molecular Neurodegeneration*, 14(2).

Waples, R. S. and Audzijonyte, A. (2016) 'Fishery-induced evolution provides insights into adaptive responses of marine species to climate change', *Frontiers in Ecology and the Environment*, 14(4), pp. 217–224. doi: 10.1002/fee.1264.

Weber, J. N. *et al.* (2017) 'Resist globally, infect locally: A transcontinental test of adaptation by stickleback and their tapeworm parasite', *The American Naturalist*, 189(1), pp. 43–57. doi: 10.1086/689597.

Weir, B. S. and Cockerham, C. C. (1984) 'Estimating F-statistics for the analysis of population structure', *Evolution*, 38(6), p. 1358. doi: 10.2307/2408641.

Willing, E. M., Dreyer, C. and van Oosterhout, C. (2012) 'Estimates of genetic differentiation

measured by Fst do not necessarily require large sample sizes when using many SNP markers', *PLoS ONE*, 7(8), pp. 1–7. doi: 10.1371/journal.pone.0042649.

Wilson, K. L. *et al.* (2019) 'History variation along environmental and harvest clines of a northern freshwater fish: Plasticity and adaptation', *Journal of Animal Ecology*, 88(5), pp. 717–733. doi: 10.1111/1365-2656.12965.

Yamazaki, A. *et al.* (2018) 'Gene expression of antifreeze protein in relation to historical distributions of Myoxocephalus fish species', *Marine Biology*, 165(11), pp. 1–11. doi: 10.1007/s00227-018-3440-x.

Yamazaki, A. *et al.* (2019) 'Freeze tolerance in sculpins (Pisces; Cottoidea) inhabiting north Pacific and Arctic oceans: Antifreeze activity and gene sequences of the antifreeze protein', *Biomolecules*, 9(4), p. 139. doi: 10.3390/biom9040139.

Yang, G. *et al.* (2016) 'Hepatoma-derived growth factor promotes growth and metastasis of hepatocellular carcinoma cells', *Cell Biochemistry and Function*, 34(4), pp. 274–285.

Zhuang, X. *et al.* (2019) 'Molecular mechanism and history of non-sense to sense evolution of antifreeze glycoprotein gene in northern gadids', *Proceedings of the National Academy of Sciences*, 116(10). doi: 10.1073/pnas.1817138116.

6. Supplementary material

Before deciding on the use of the MAF = 0.05 dataset for the study, test analyses were run to compare filtering values of MAF at 0.05, 0.025, and 0.01, to ensure that the most suitable data was used. This was carried out on the dataset before removal of outlier SNPs, as a preliminary study to decide on which dataset to use for the main study. These results are presented here as supplementary material to show how this conclusion came about.

Also covered here is some additional work on the dataset from the main study. Included are analyses in which the blue ling data was split into its apparent populations; the Atlantic and the fjords. Population analyses were repeated on these new datasets to see if any fine-scale structure could be found which was hidden in the main analysis.

The supplementary material then goes onto the outlier dataset, for which the population analyses were again repeated. This has not been included in the main study as there were no findings of interest from this. With only a small number of outliers (3 for the common ling and 5 for the blue ling) the results had some resemblance to the original analyses, but some individuals were dropped from the analyses due to missing data.

6.1. Population structure of the common ling

In the *structure* plot for MAF = 0.05 (Fig. 6d) we see some individual outliers coming from the BE08 and SO14 populations; the same two populations around Bergen which are shown to deviate in the PCA and DAPC analyses. This adds to the possibility of structure here, inferring that there is some stronger differentiation at work in this particular area/fjord. The corresponding plot from *harvester* (Fig. 6e) struggles to allocate one value to *K*, but upon

closer inspection *K* = 1 comes out as the best fit. However, it seems that *K* could equal anywhere between 1 and 5 from this data.

The structural analyses carried out using *LEA* in *R* also give a general pattern to indicate one population overall (Fig. 6f), but again with some outlying individuals. This time they correspond with the PCA and DAPC (Fig. 6b & 6c) differently, indicating one of the Rockall populations to be differentiated from the rest of the population (RA14 – the purple population in PCA and DAPC plots). This is interesting that the different structural analysis methods pick up on different aspects of patterns found in PCA and DAPC.

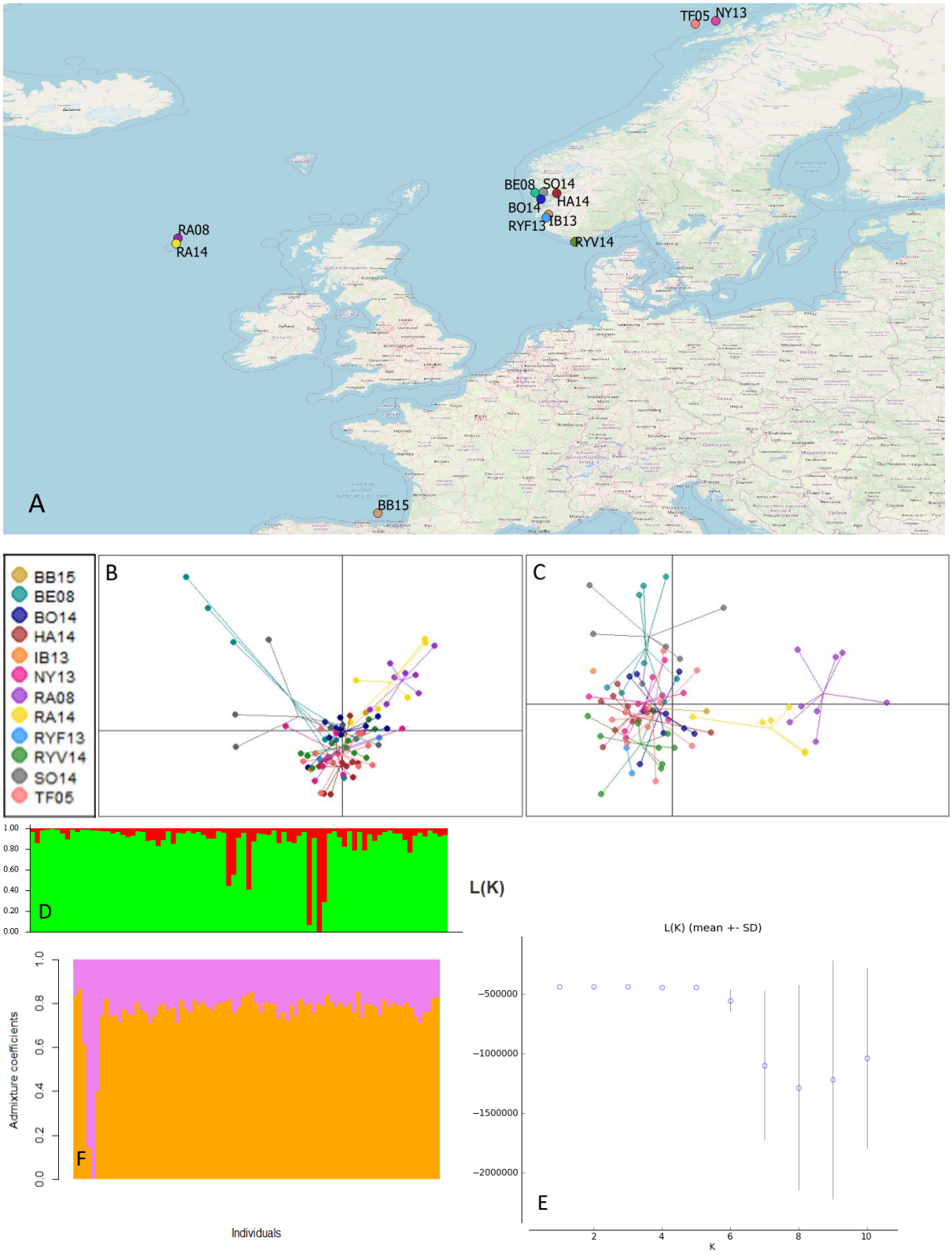

Figure 6: Population structure of the common ling with; a) map showing coordinates for the locations of each population sample site; b) principle component analysis plot; c) discriminant analysis of principal components; d) structure plot; e) the corresponding plot produced in structure harvester of L(k) (mean +- SD) for each value of k tested (both produced using the dataset for MAF = 0.05); and f) structure plot produced in the LEA package in R for the common ling dataset MAF = 0.05

The cross-entropy values given (Table 13) are similar to the pattern seen in *structure harvester*, in that it isn't completely clear what the best fit for *K* is from these values. Again, *K* = 1 comes out as best, but it is very close between that and *K* = 2. The pattern pointing toward *K* = 1 is also backed up by analysis in *fastStructure* (Table 14) which also indicated 1 as best fit.

Table 13: the corresponding cross entropy values produced for LEA structure data (both produced using the dataset for MAF = 0.05)

	K=1	*K=2*	*K=3*	*K=4*	*K=5*	*K=6*	*K=7*	*K=8*	*K=9*	*K=10*
Min	0.655	0.663	0.672	0.684	0.702	0.715	0.732	0.742	0.768	0.776
Mean	0.662	0.670	0.679	0.692	0.707	0.720	0.737	0.750	0.772	0.786
Max	0.668	0.677	0.686	0.701	0.713	0.728	0.745	0.759	0.778	0.797

Table 14: FastStructure results for best fit of k calculated using the dataset for MAF = 0.05

	K
Model complexity that maximises marginal likelihood	1
Model components used to explain structure in data	1

With the MAF = 0.025 data the PCA and DAPC plots maintain the same patterns as those produced with the MAF = 0.05 data (Fig. 7b & 7c). The analyses then begin to differ, as we lose the outliers in the *structure* plot (Fig. 7d). This plot seems to more closely point towards *K* = 1, which again comes out as best fit in *harvester* (Fig. 7e). Again, we see that *K* could equal a number of values from this plot, falling anywhere between 1 and 6.

With the *LEA* plot from the MAF = 0.025 dataset we get the same Rockall population deviating from the norm, and also a slight deviation from one individual in the RYF13 population (fjord)

(Fig. 7f). This deviation does not seem to correspond to any pattern found anywhere else in the analyses.

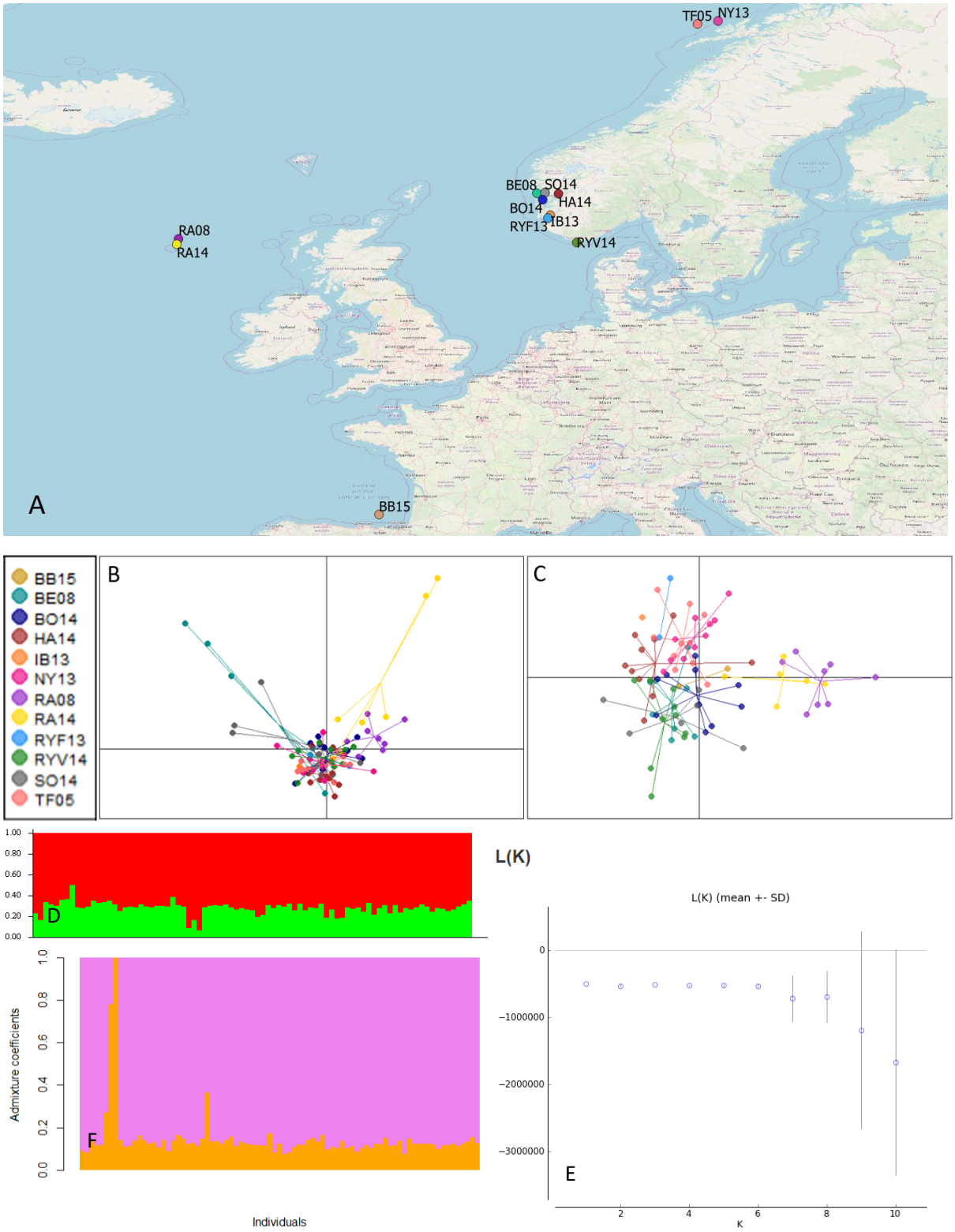

Figure 7: Population structure of the common ling at MAF 0.025 with; a) map showing coordinates for the locations of each population sample site; b) principle component analysis plot; c) discriminant analysis of

principal components; d) structure plot; e) LEA plot; and f) structure harvester plot corresponding to the structure plot

Again, *K* = 1 appears to come out as best fit from the cross-entropy values (Table 15) and this is also found in *fastStructure* (Table 16).

Table 15: the corresponding cross entropy values produced for LEA structure data (both produced using the dataset for MAF = 0.025)

	K=1	*K=2*	*K=3*	*K=4*	*K=5*	*K=6*	*K=7*	*K=8*	*K=9*	*K=10*
Min	0.532	0.538	0.548	0.560	0.572	0.582	0.600	0.613	0.627	0.638
Mean	0.534	0.540	0.552	0.562	0.576	0.587	0.602	0.619	0.632	0.649
Max	0.537	0.544	0.555	0.565	0.581	0.593	0.609	0.625	0.637	0.658

Table 16: FastStructure results for best fit of k calculated using the dataset for MAF = 0.025

	K
Model complexity that maximises marginal likelihood	1
Model components used to explain structure in data	1

Again, as may be expected, with MAF at the value of 0.01 we find PCA and DAPC plots (Fig. 8b & 8c) exhibiting the same patterns as those produced from MAF = 0.05 and MAF = 0.025. With the *structure* analysis we find that the outliers seen earlier appear again, along with a couple of extra individuals (Fig. 8d). The added individuals are from the populations HA14 and TF05, two more of the fjords, but this data does not appear to correspond with any deviations found anywhere else in the analyses. Maybe this is linked with the increased ability to find the fine-scale structural variations with an increased number of SNPs.

Harvester no longer gives the flatline pattern, and it is much clearer that *K* = 1 is best fit (Fig. 8e). This could indicate that relaxing the filtering on minor alleles and retaining more SNPs allows us to find the true structure in the analyses. It seems that this becomes unclear with the loss of SNPs, and with the stricter datasets our analyses are unable to pick out one suitable value.

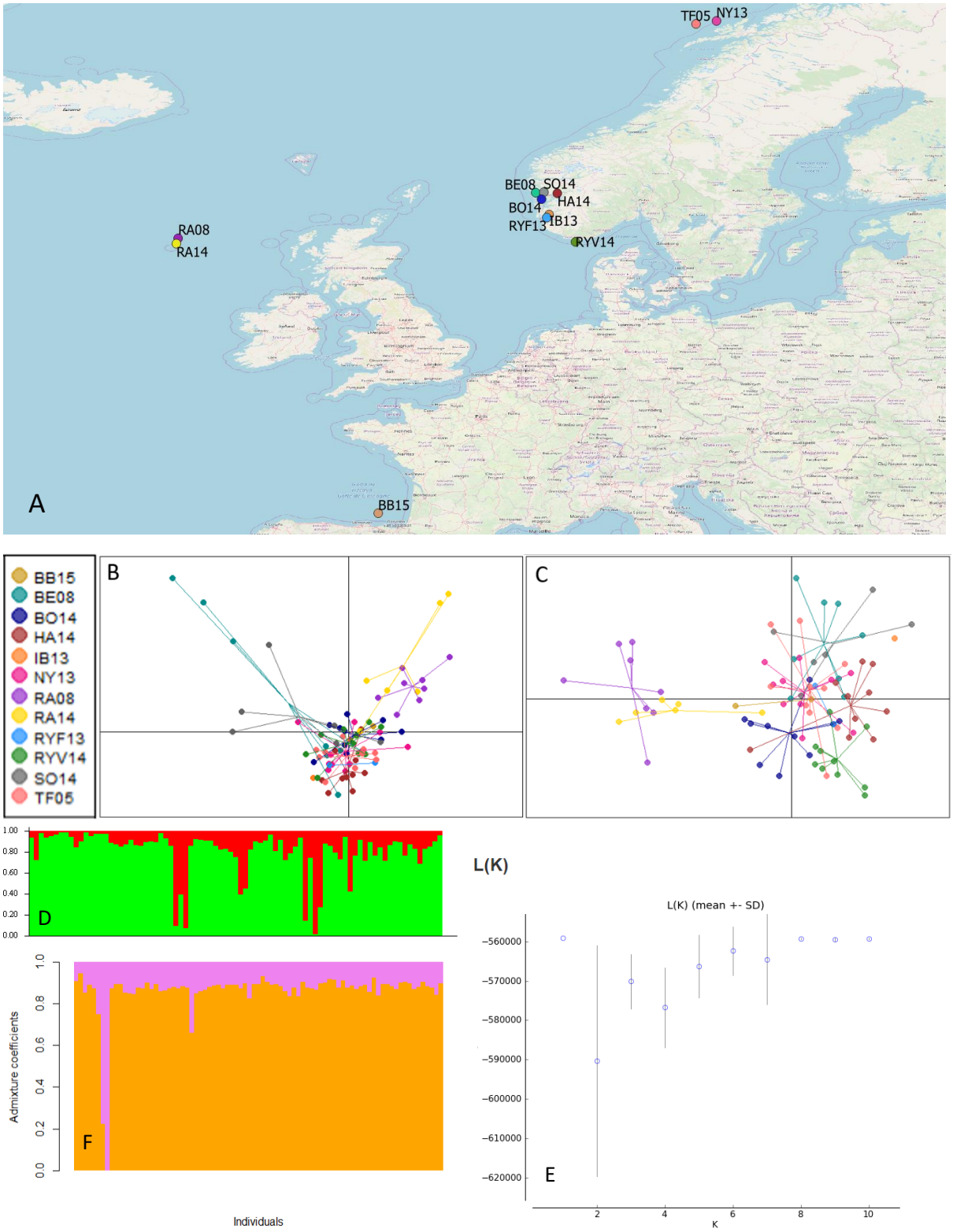

Figure 8: Population structure of the common ling at MAF 0.01 with; a) map showing coordinates for the locations of each population sample site; b) principle component analysis plot; c) discriminant analysis of principal components; d) structure plot; e) LEA plot; and f) structure harvester plot corresponding to the structure plot

Although *harvester* is more certain of the calculations for *K* in the MAF = 0.01 dataset, the cross-entropy values found by *LEA* do not seem to become any clearer (Table 17). Values remain consistently close, and *K* still seems to best fit as 1 population. The *LEA* plot is also very similar to that of MAF = 0.025, with the same outliers indicated (Fig. 8f). Again, *fastStructure* indicates 1 population overall (Table 18).

Table 17: the corresponding cross entropy values produced for LEA structure data (both produced using the dataset for MAF = 0.01)

	K=1	*K=2*	*K=3*	*K=4*	*K=5*	*K=6*	*K=7*	*K=8*	*K=9*	*K=10*
Min	0.526	0.531	0.542	0.556	0.569	0.576	0.592	0.603	0.621	0.641
Mean	0.538	0.544	0.555	0.568	0.581	0.591	0.609	0.619	0.637	0.654
Max	0.547	0.554	0.564	0.580	0.590	0.602	0.623	0.627	0.649	0.661

Table 18: FastStructure results for best fit of k calculated using the dataset for MAF = 0.01

	K
Model complexity that maximises marginal likelihood	1
Model components used to explain structure in data	1

From these first analyses we do not find many differences between the results produced from the different filtered datasets. Slight changes in weight of differentiation for some of the populations are exhibited, which may indicate that some of the SNPs lost in stricter filtering are holding structural information; or it could be the opposite and increased SNPs could be hiding the true structure of the populations. Overall, from the (mostly) consistent results throughout we decided it was safe to go with a stricter filter and lose more SNPs, as structure should not be lost through this.

6.2. Population structure for the blue ling

The analysis of the MAF = 0.05 filtered dataset gives an overall pattern of two populations, seeing a clear split of the Atlantic and fjords for the *structure* plot in Figure 9d, and the same signal to a lesser degree for the *LEA* plot in Figure 9f. Though this pattern appears quite clear in these plots, the *harvester* plot (Fig. 9e) still exhibits a flatline pattern across K = 1 to 4 and cross-entropy values from *LEA* (Table 19) are again unclear. This time, when looking closer at the *harvester* plot we find K = 2 to be best fit, although cross-entropy still finds a value of 1. The values we find in *fastStructure* seem to correspond more with the *harvester* values, finding that K could equal between 2 and 5 (Table 20).

We also see a couple of outliers in our plots. That produced in the *Structure* program (Fig. 9d) shows the same two Atlantic outliers as those seen in the DAPC analysis (Fig. 9c), one individual from RS07 and one individual from SL14. The fjord outliers are the HA14 outlier seen in the PCA plot (Fig. 9b), and RYV14, which doesn't show up as a deviation in any other analyses.

The *LEA* plot (Fig. 9f), although the structure is less clear, exhibits the same two Atlantic outliers as the *Structure* plot and DAPC. From the fjords, we again find HA14 deviating, and also one individual from NY13.

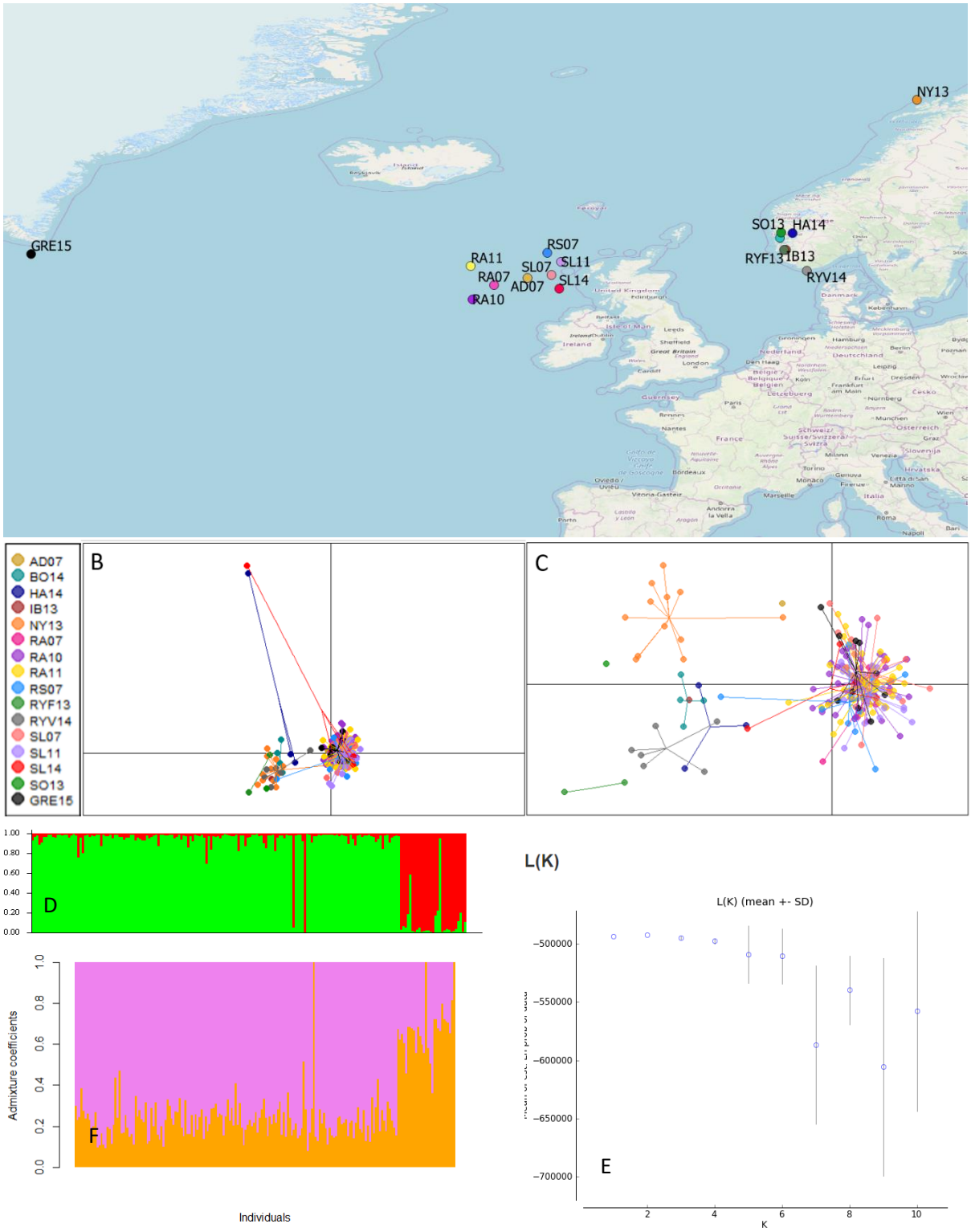

Figure 9: Population structure of the blue ling with; a) map showing coordinates for the locations of each population sample site; b) principle component analysis plot; c) discriminant analysis of principal components; d) structure plot; e) the corresponding plot produced in structure harvester of L(k) (mean +- SD) for each value of k tested (both produced using the dataset for MAF = 0.05); and f) structure plot produced in the LEA package in R for the common ling dataset MAF = 0.05

Table 19: the corresponding cross entropy values produced for LEA structure data (both produced using the dataset for MAF = 0.05)

	K=1	*K=2*	*K=3*	*K=4*	*K=5*	*K=6*	*K=7*	*K=8*	*K=9*	*K=10*
Min	0.681	0.683	0.690	0.692	0.693	0.699	0.705	0.712	0.715	0.715
Mean	0.685	0.686	0.692	0.695	0.697	0.702	0.707	0.713	0.718	0.722
Max	0.687	0.688	0.695	0.699	0.700	0.706	0.714	0.714	0.723	0.725

Table 20: FastStructure results for best fit of k calculated using the dataset for MAF = 0.05

	K
Model complexity that maximises marginal likelihood	2
Model components used to explain structure in data	5

With the MAF = 0.025 dataset (Fig. 10) we get the same patterns as in our population analysis at MAF = 0.05. All outlying individuals remain the same in all plots, and we still see this clear split between the Atlantic and fjord populations. *Harvester* again finds *K* = 2 (Fig. 10e), and cross-entropy indicates *K* = 1 (Table 21). This time *fastStructure* gives a best value of *K* to be between 1 and 3 (Table 22).

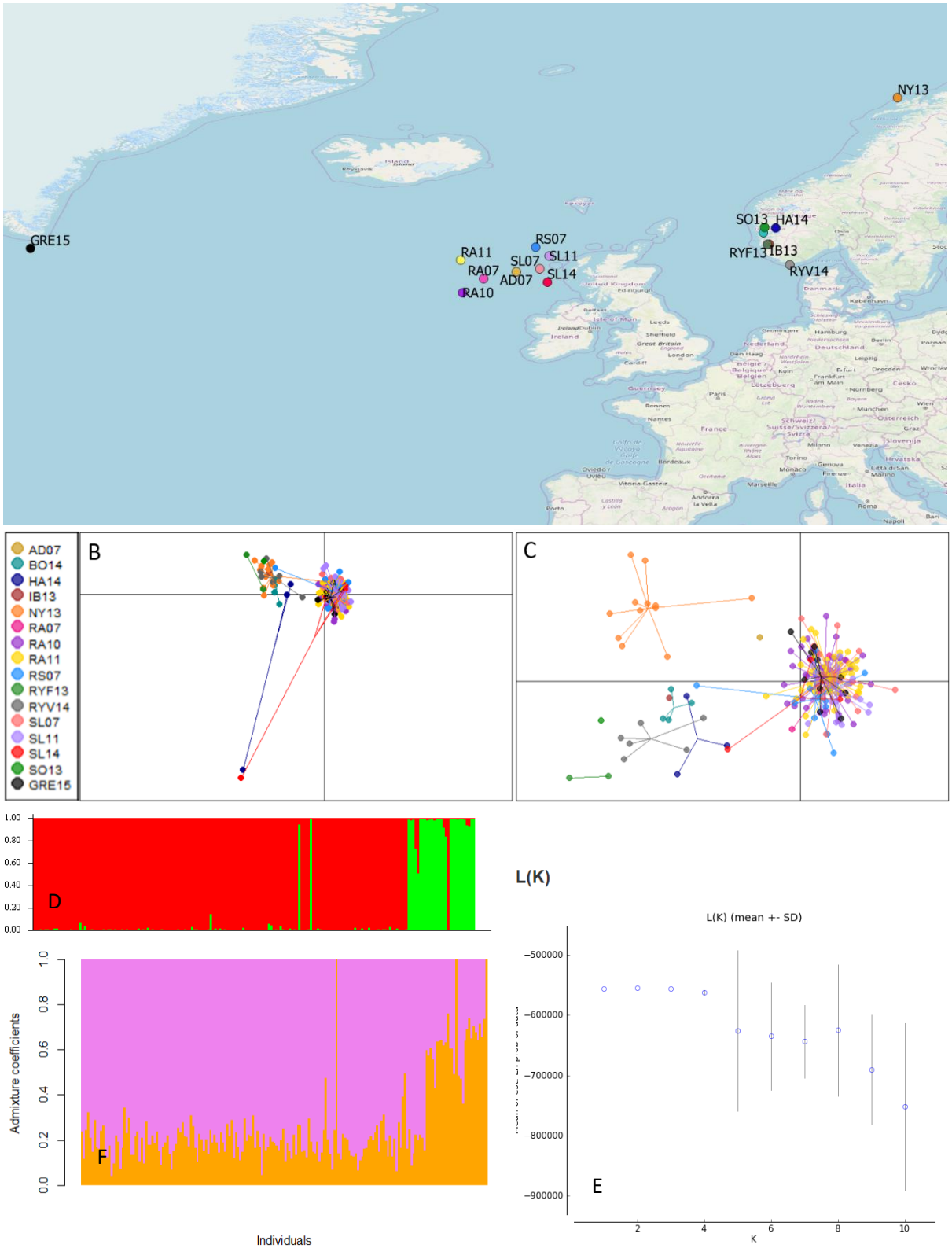

Figure 10: Population structure of the blue ling at MAF = 0.025 with; a) map showing coordinates for the locations of each population sample site; b) principle component analysis plot; c) discriminant analysis of principal components; d) structure plot; e) LEA plot; and f) structure harvester plot corresponding to the structure plot

Table 21: the corresponding cross entropy values produced for LEA structure data (both produced using the dataset for MAF = 0.025)

	K=1	*K=2*	*K=3*	*K=4*	*K=5*	*K=6*	*K=7*	*K=8*	*K=9*	*K=10*
Min	0.569	0.569	0.574	0.572	0.576	0.582	0.587	0.591	0.595	0.598
Mean	0.571	0.572	0.576	0.578	0.581	0.585	0.588	0.592	0.596	0.600
Max	0.573	0.575	0.580	0.581	0.584	0.589	0.591	0.597	0.599	0.603

Table 22: FastStructure results for best fit of k calculated using the dataset for MAF = 0.025

	K
Model complexity that maximises marginal likelihood	1
Model components used to explain structure in data	3

Again, there is little difference found from the MAF = 0.01 data (Fig. 11) compared with that of MAF = 0.05 and MAF = 0.025. However, with this data it is difficult to pick a value for *K* from *harvester* (Fig. 11e), which seems to be a best fit of either 1 or 3. Cross-entropy still indicates *K* = 1 (Table 23) and *fastStructure* finds that *K* is best fit with 2 or 1 (Table 24). The outlying individuals again follow the same pattern in both Atlantic and fjord populations in all of the plots.

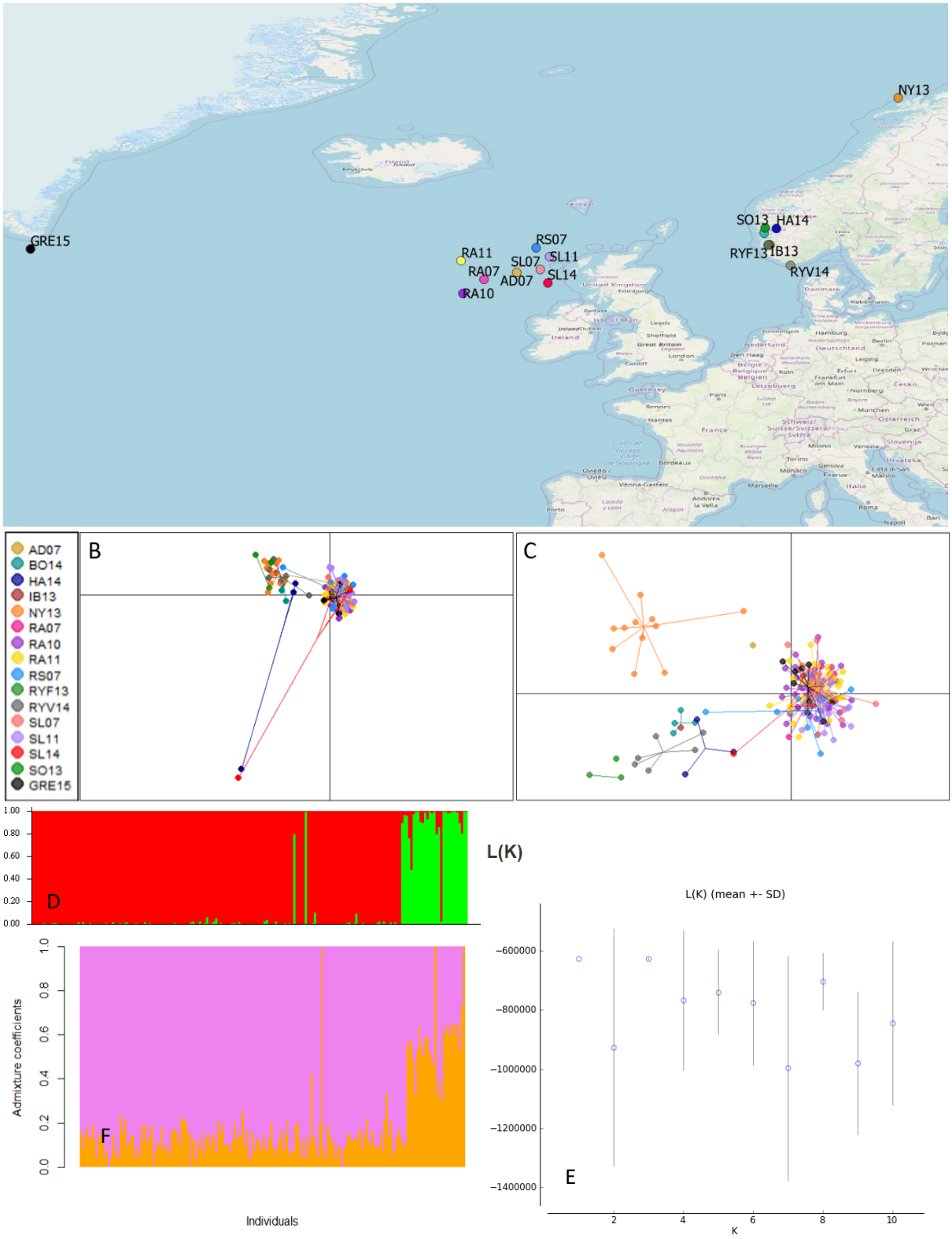

Figure 11: Population structure of the blue ling at MAF = 0.01 with; a) map showing coordinates for the locations of each population sample site; b) principle component analysis plot; c) discriminant analysis of principal components; d) structure plot; e) LEA plot; and f) structure harvester plot corresponding to the structure plot

Table 23: the corresponding cross entropy values produced for LEA structure data (both produced using the dataset for MAF = 0.01)

	K=1	*K=2*	*K=3*	*K=4*	*K=5*	*K=6*	*K=7*	*K=8*	*K=9*	*K=10*
Min	0.413	0.414	0.413	0.417	0.420	0.422	0.426	0.429	0.432	0.437
Mean	0.418	0.418	0.419	0.422	0.425	0.428	0.432	0.435	0.438	0.441
Max	0.422	0.423	0.427	0.429	0.430	0.434	0.437	0.440	0.443	0.446

Table 24: FastStructure results for best fit of k calculated using the dataset for MAF = 0.01

	K
Model complexity that maximises marginal likelihood	2
Model components used to explain structure in data	1

In the PCA and DAPC across the different filtering datasets for the blue ling we find the same consistency in results as found with the common ling data. Very little change is seen, possibly with even less effect within this species. Two clear clusters are formed in all plots; within these clusters we see fjord samples grouping together in one and Atlantic samples in the other.

Drawing away from these clusters are the same two individuals in all PCA plots. One individual from the HA14 sample group, taken from one of the fjords, and the other from the SL14 sample which comes from the Rockall Slope. These appear to greatly deviate from the rest of the individuals.

We find the same scenario with the DAPC in which all plots across the datasets are almost identical. They also exhibit a similar pattern; we see a very clear cluster of Atlantic populations formed, with the fjord samples separated from this. These fjords are then separated out from

one another, which differs from the PCA. Possibly the most clearly separated is the NY13 group, a fjord found much further North from the rest at Nygunnen in the North of Norway.

Outlying individuals within these plots should also be noted. From the Atlantic population there are two individuals clearly branching out towards the fjord populations to the bottom of the plot; these individuals originating from RS07 and SL14. This is the same SL14 individual seen branching out in the PCA, and although we don't see this same behaviour from HA14 like in the PCA the SL14 individual is stretching out to sit with this same individual in the DAPC plots.

Other deviating individuals are seen towards the top of the plot, with one individual stretching out from NY13 more towards the Atlantic cluster, and with the one AD07 individual (an Atlantic population) sitting on its own between the Atlantic and Fjord groupings.

In all *Structure* plots we see the same two Atlantic individuals, RS07 and SL14, and the same two fjord individuals, HA14 and RYV14, acting as outliers. All *LEA* plots exhibit these same Atlantic outliers and from the fjords we see HA14 and NY13 deviating from the norm. The consistency of this throughout all analyses indicates the strength of these patterns, and confirms the presence of two distinct populations within the blue ling throughout the sampled area. Again, it was clear that the stricter filtering dataset of MAF = 0.05 was sufficient to detect structure in this study and thus results from those analyses were presented.

6.3. Atlantic population analysis in the blue ling

To investigate the results of the blue ling analyses further, we split the populations into the Atlantic and the fjords and repeated our analyses – excluding F_{st} as with it being pairwise it

should not change when excluding different populations. This should give us a deeper look into the two population clusters inferred in the previous analyses. Again, the analyses were done with the three different filtering datasets for comparison.

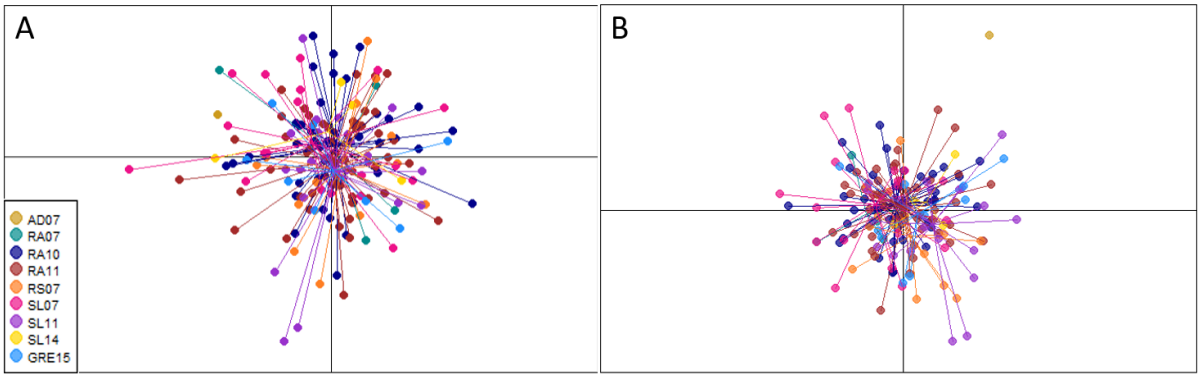

Figure 12: Analyses for the blue ling Atlantic populations using the filtered dataset with MAF = 0.05; a) principal component analysis and b) discriminant analysis of principal components

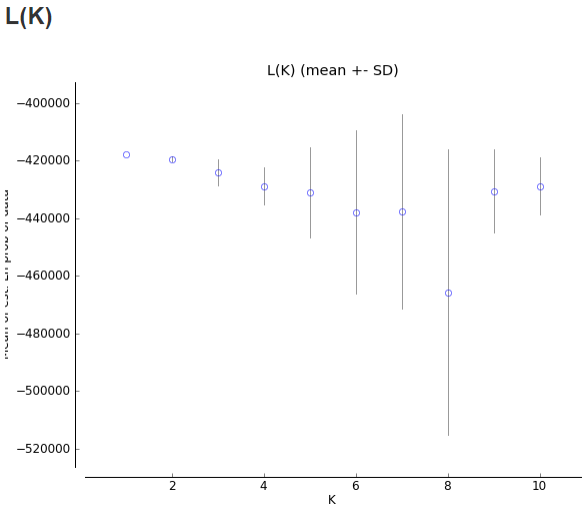

Figure 13: Structure harvester plot for analysis of Atlantic blue ling populations at MAF = 0.05

Table 25: Cross entropy values for LEA structure data produced using the dataset for MAF = 0.05

	K=1	*K=2*	*K=3*	*K=4*	*K=5*	*K=6*	*K=7*	*K=8*	*K=9*	*K=10*
Min	0.680	0.687	0.694	0.701	0.704	0.713	0.719	0.720	0.729	0.735
Mean	0.685	0.692	0.698	0.706	0.711	0.717	0.722	0.728	0.732	0.738
Max	0.690	0.698	0.704	0.711	0.716	0.720	0.726	0.732	0.736	0.741

Table 26: FastStructure results for best fit of k calculated using the dataset for MAF = 0.05

	K
Model complexity that maximises marginal likelihood	1
Model components used to explain structure in data	4

The inference of one population within our Northern Atlantic samples continues with the *Structure*, *LEA*, and *FastStructure* analyses. Both *Structure* and *LEA* analyses failed to find any structure within the samples. This is also backed up by *structure harvester* which this time is more clearly exhibiting *K* = 1 as best fit (Fig. 30). We again find this in the cross-entropy values also (Table 25). *FastStructure* gives a range of between 1 and 4 for best fit of *K* (Table 26).

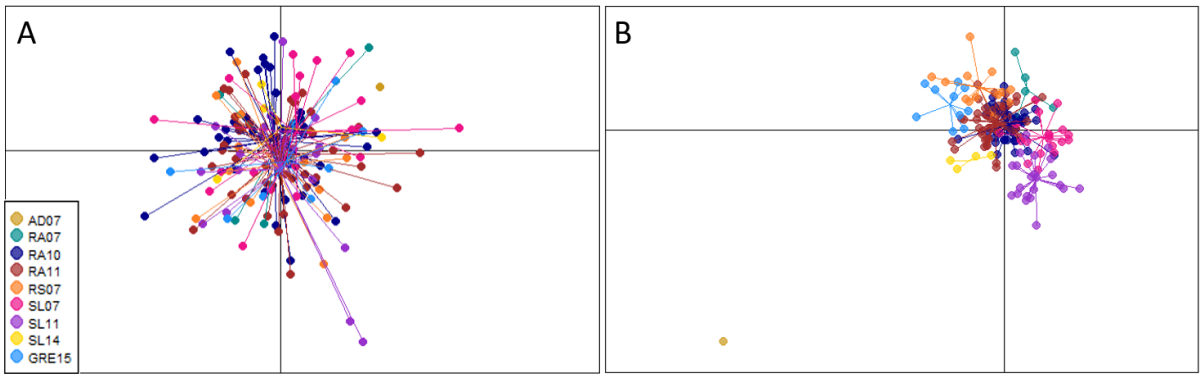

Figure 14: Analyses for the blue ling Atlantic populations using the filtered dataset with MAF = 0.025; a) principal component analysis and b) discriminant analysis of principal components

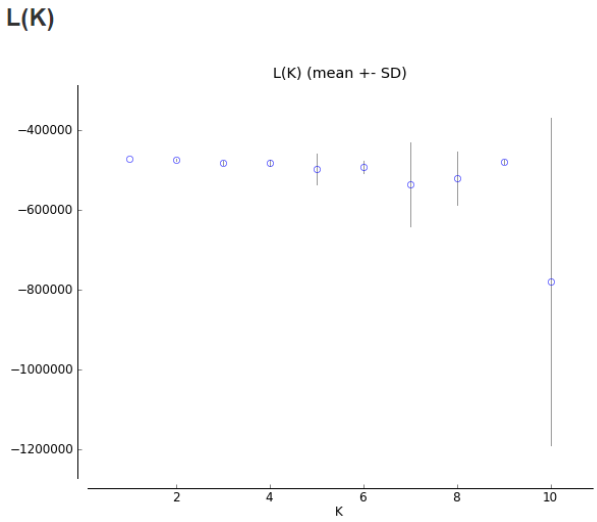

Figure 15: Structure harvester plot for analysis of Atlantic blue ling populations at MAF = 0.025

Table 27: Cross entropy values for LEA structure data produced using the dataset for MAF = 0.025

	K=1	*K=2*	*K=3*	*K=4*	*K=5*	*K=6*	*K=7*	*K=8*	*K=9*	*K=10*
Min	0.572	0.577	0.585	0.588	0.592	0.594	0.602	0.606	0.611	0.618
Mean	0.575	0.582	0.589	0.594	0.596	0.602	0.607	0.613	0.618	0.624
Max	0.579	0.585	0.594	0.597	0.601	0.605	0.609	0.618	0.623	0.628

Table 28: FastStructure results for best fit of k calculated using the dataset for MAF = 0.025

	K
Model complexity that maximises marginal likelihood	1
Model components used to explain structure in data	2

With the MAF 0.025 dataset we find similar patterns as that of the MAF 0.05 data. No structure is exhibited in *Structure* or *LEA* analyses. In *Structure Harvester* we find the flatline returns and it is more difficult to accurately place *K*. Upon closer inspection *Harvester* gives *K*

= 1, and we see this again with cross-entropy values. *FastStructure* is closer to this in its range, with values of 1 or 2 given as best fit.

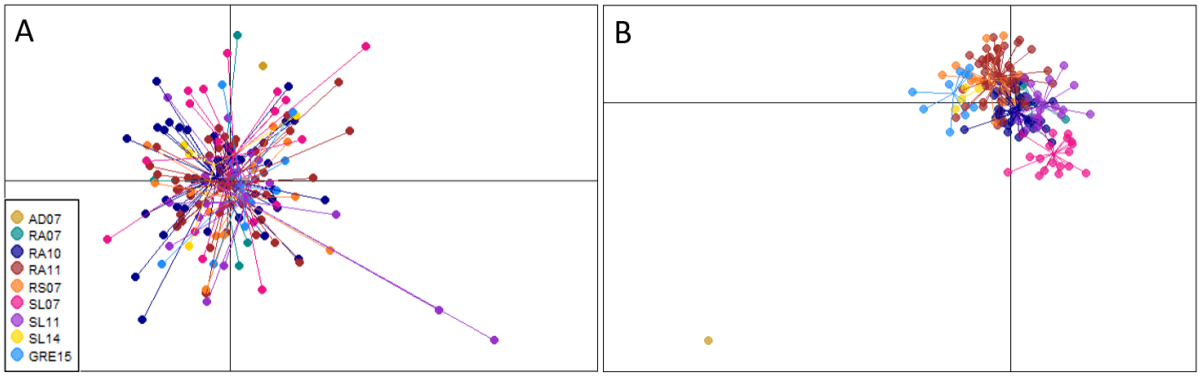

Figure 16: Analyses for the blue ling Atlantic populations using the filtered dataset with MAF = 0.01; a) principal component analysis and b) discriminant analysis of principal components

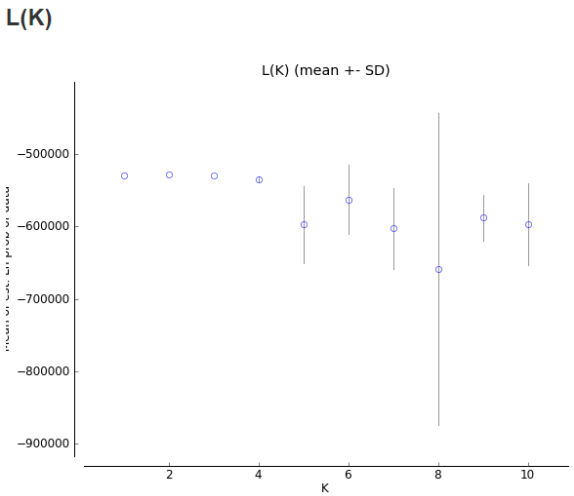

Figure 17: Structure harvester plot for analysis of Atlantic blue ling populations at MAF = 0.01

Table 29: Cross entropy values for LEA structure data produced using the dataset for MAF = 0.01

	K=1	*K=2*	*K=3*	*K=4*	*K=5*	*K=6*	*K=7*	*K=8*	*K=9*	*K=10*
Min	0.673	0.678	0.686	0.694	0.701	0.705	0.707	0.715	0.720	0.728
Mean	0.684	0.690	0.697	0.704	0.711	0.716	0.722	0.728	0.733	0.741
Max	0.695	0.702	0.708	0.715	0.722	0.729	0.732	0.741	0.742	0.748

Table 30: FastStructure results for best fit of k calculated using the dataset for MAF = 0.01

	K
Model complexity that maximises marginal likelihood	1
Model components used to explain structure in data	1

Similar results continue with the MAF 0.01 dataset. *Structure* and *LEA* again found a lack of structure indicative of one population, and the cross entropy and *harvester* values came out with *K* = 1 as best fit. Though a flatline for values is once again produced from *Harvester*, *FastStructure* chooses *K* = 1 as the best fit with no range.

Looking at the analyses across the filtering datasets, starting with PCA analyses of the Atlantic populations, we find no further structure which is not already seen in the initial analyses. There is one very clear cluster indicated in all three analyses. It appears that this data backs up that we have just one highly connected population of blue ling in the Northern Atlantic.

The DAPC plot produced from this data also exhibits one very clear cluster, but here we have one outlier from this group. This comes from a sample population containing just one individual. It may be possible that any differences found in this one individual would be amplified more so because of this. Had there been more samples taken from this site, we may have found this population to be less differentiated than it appears, or it could have allowed us to investigate further some fine-scale differentiation between individuals.

Again, this data is reinforcing the result of one population of blue ling in the Northern Atlantic. It seems likely that the one outlier is just an artefact of the DAPC analyses as it does not appear to show up anywhere else in our analyses. It also appears to be minor allele frequency

differences drawing this individual out from the cluster as we find increasing differentiation when relaxing the MAF filter.

6.4. Fjord population analysis in the blue ling

Moving onto the fjord populations next, we may expect to see more structure considering the DAPC plots of the broader datasets. With fjords scattered about quite evenly in the previous plots, it may appear that all could be individual populations with further analysis. It is important to look into this further to gain an understanding of the interactions exhibited between fjord individuals.

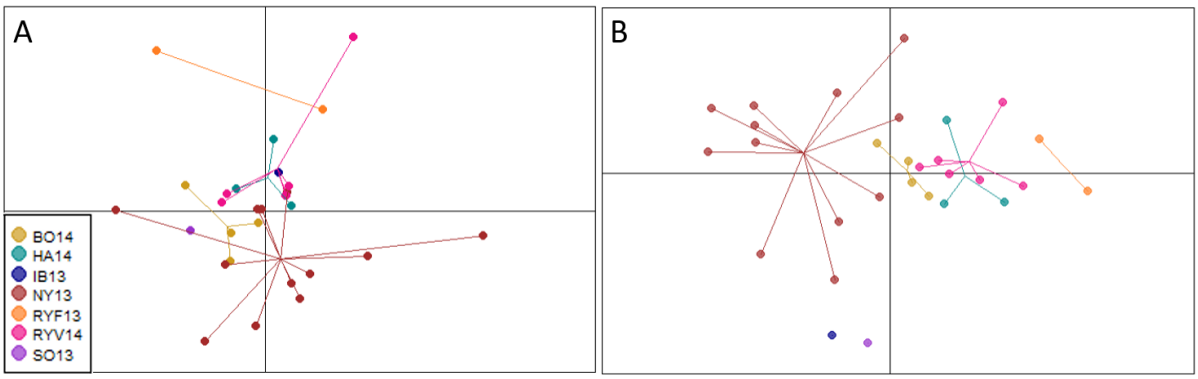

Figure 18: Analyses for the blue ling fjord populations using the filtered dataset with MAF = 0.05; a) principal component analysis and b) discriminant analysis of principal components

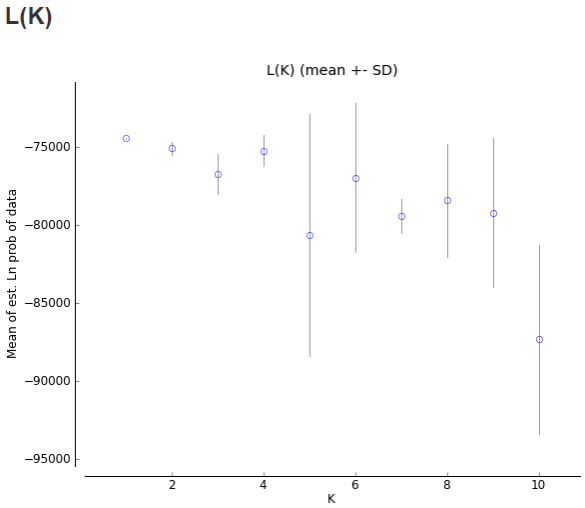

Figure 19: Structure harvester plot for analysis of fjord blue ling populations at MAF = 0.05

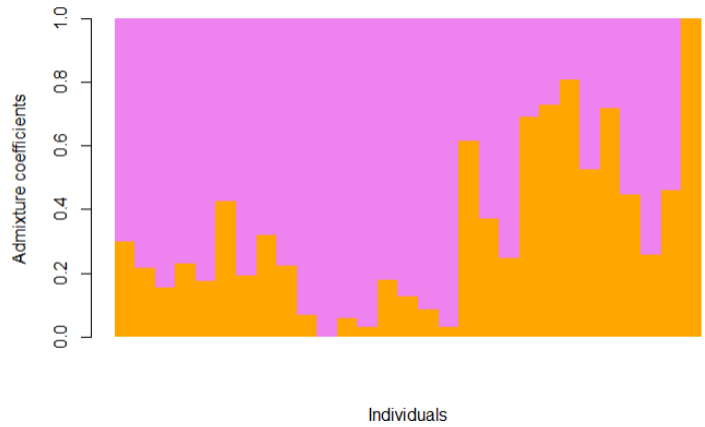

Figure 20: Structure plot produced in the LEA package in R for fjord blue ling populations at MAF = 0.05

Table 31: Cross entropy values for LEA structure data at MAF = 0.05

	K=1	*K=2*	*K=3*	*K=4*	*K=5*	*K=6*	*K=7*	*K=8*	*K=9*	*K=10*
Min	0.764	0.776	0.828	0.902	0.988	1.043	1.130	1.172	1.238	1.283
Mean	0.787	0.812	0.875	0.940	1.014	1.091	1.150	1.221	1.304	1.320
Max	0.813	0.831	0.906	0.965	1.054	1.117	1.189	1.260	1.343	1.405

Table 32: FastStructure results for best fit of k calculated using the dataset for MAF = 0.05

	K
Model complexity that maximises marginal likelihood	1
Model components used to explain structure in data	1

First, for the analyses using the MAF = 0.05 data, we see PCA and DAPC plots (Fig. 35) which give no further insight into structure than those previously produced in the broader analyses. Fjord populations appear to be slightly spread out from one another like in the previous plots, suggesting there may be some differentiation between them. The *Structure* analysis failed to find any structure, which is confirmed by *structure harvester* (Fig. 36) where *K* = 1 appears to be best fit. Despite no structure being detected with the *structure* analysis, we appear to see a bit of structure hinted at in the *LEA* plot (Fig. 37). To the right of the plot we see the orange bars take up more of the plot compared to the left which is dominated by the pink colour. These more orange individuals are of the NY13 population to the North of Norway. This may hint at greater differentiation of this population from the others, which is separated out from the rest by a greater distance.

Although we have these hints of structure, the cross entropy values to go with the *LEA* analysis indicate *K* = 1 as best fit (Table 31). This fits with the values given by *fastStructure*, which also gives *K* = 1 (Table 32).

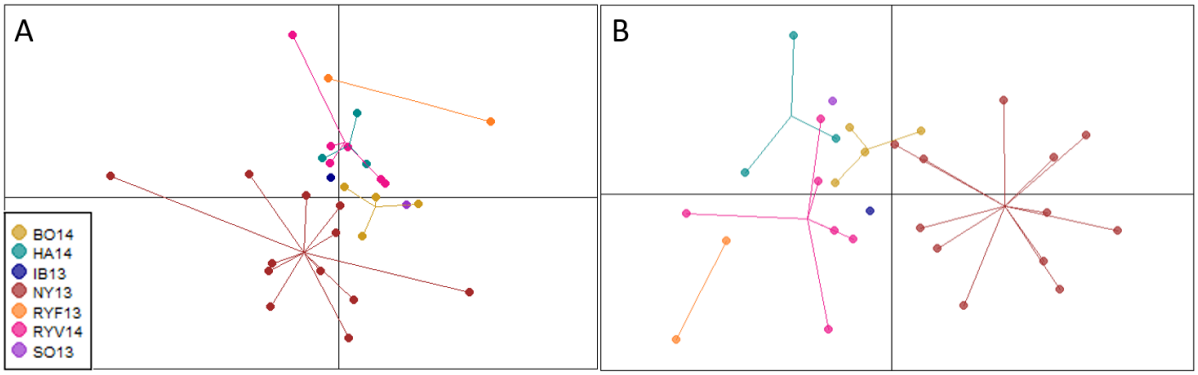

Figure 21: Analyses for the blue ling fjord populations using the filtered dataset with MAF = 0.025; a) principal component analysis and b) discriminant analysis of principal components

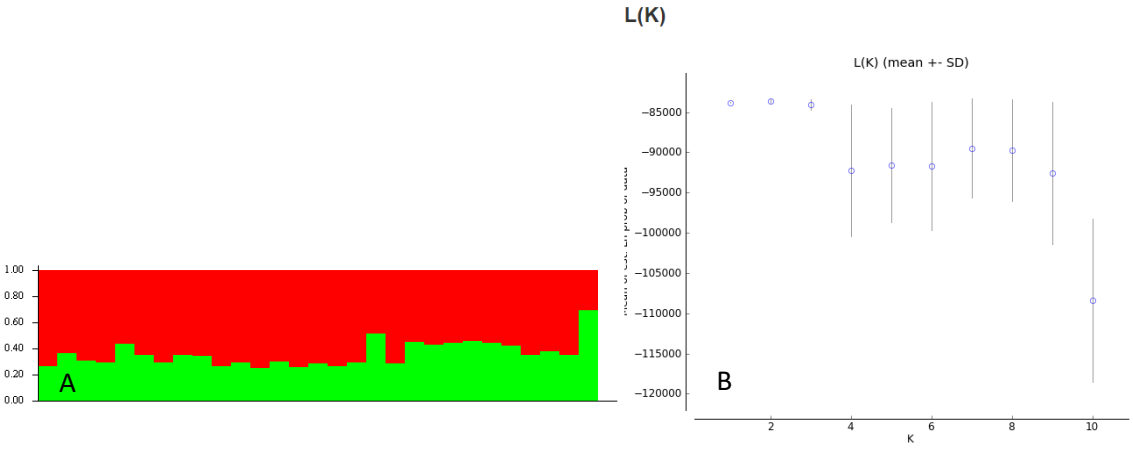

Figure 22: a) structure plot and b) structure harvester plot for analyses of fjord blue ling populations at MAF = 0.025

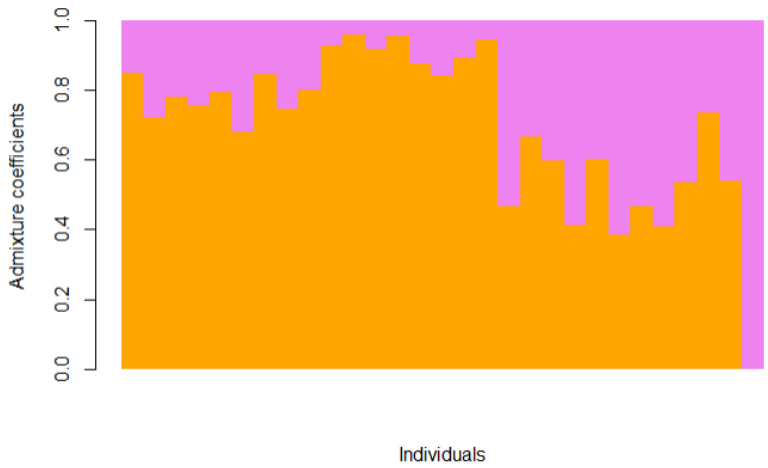

Figure 23: Structure plot produced in the LEA package in R for fjord blue ling populations at MAF = 0.025

Table 33: Cross entropy values for LEA structure data at MAF = 0.025

	K=1	*K=2*	*K=3*	*K=4*	*K=5*	*K=6*	*K=7*	*K=8*	*K=9*	*K=10*
Min	0.739	0.759	0.817	0.868	0.931	0.984	1.053	1.152	1.235	1.253
Mean	0.751	0.775	0.833	0.887	0.970	1.020	1.111	1.168	1.248	1.308
Max	0.766	0.802	0.852	0.897	1.024	1.064	1.137	1.180	1.272	1.346

Table 34: FastStructure results for best fit of k calculated using the dataset for MAF = 0.025

	K
Model complexity that maximises marginal likelihood	1
Model components used to explain structure in data	1

Similarly, with the MAF = 0.025 we have populations separated out in the PCA and DAPC plots (Fig. 38), although a little more so with this data. Here it looks a little clearer that the NY13 population is slightly more deviated from the rest. The *Structure* plot doesn't give any clear indication of structure (Fig. 39a), but to the right the green bars do appear to take up more of the plot compared to the more red coloured left, much like that which is seen in the *LEA* plot (Fig. 40). This is minimal and may not be noticed if it weren't for the *LEA* plot but could be backing up the data in suggesting that NY13 is differentiated a little from the other fjord populations.

Structure harvester is less clear with this analysis (Fig. 39b), and appears to find any value between 1 and 3 to be best fitted for *K*. This is not backed up by the *LEA* cross entropy values (Table 33) or the *fastStructure* values (Table 34) which both find *K* = 1.

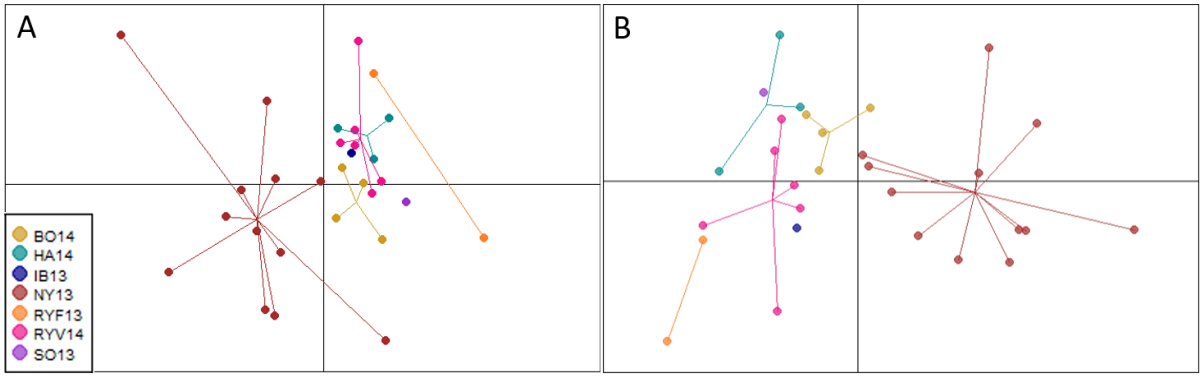

Figure 24: Analyses for the blue ling fjord populations using the filtered dataset with MAF = 0.01; a) principal component analysis and b) discriminant analysis of principal components

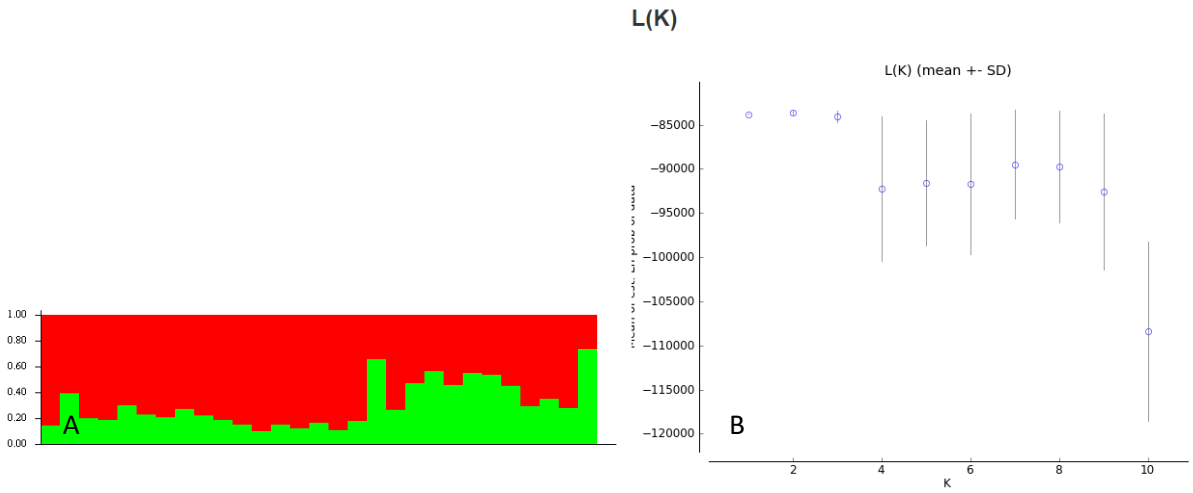

Figure 25: a) structure plot and b) structure harvester plot for analyses of fjord blue ling populations at MAF = 0.01

Figure 26: Structure plot produced in the LEA package in R for fjord blue ling populations at MAF = 0.01

Table 35: Cross entropy values for LEA structure data at MAF = 0.01

	K=1	*K=2*	*K=3*	*K=4*	*K=5*	*K=6*	*K=7*	*K=8*	*K=9*	*K=10*
Min	0.723	0.744	0.798	0.862	0.932	0.952	1.027	1.126	1.170	1.184
Mean	0.735	0.757	0.811	0.875	0.956	1.010	1.073	1.134	1.199	1.272
Max	0.759	0.781	0.840	0.887	0.984	1.064	1.135	1.146	1.246	1.305

Table 36: FastStructure results for best fit of k calculated using the dataset for MAF = 0.01

	K
Model complexity that maximises marginal likelihood	1
Model components used to explain structure in data	1

We find with the MAF = 0.01 data that the PCA and DAPC plots become more spread out again (Fig. 41). NY13 becomes even more clearly separate from the other populations. This continues with the *Structure* and *LEA* plots. In the *Structure* analysis (Fig. 42a) we now see a clearer deviation between the left and right of the plot, which shows the NY13 population to be slightly differentiated from the other fjords. Again, this is also seen in the *LEA* plot (Fig. 43).

With the *structure harvester* plot (Fig. 42b) we get the same results as the analysis of the MAF = 0.025 data, with the value for *K* being between 1 and 3. Again, these findings are lacking in the cross entropy (Table 35) and *fastStructure* (Table 36) results, both of which find *K* = 1.

Overall, with the PCA analyses we see there is differentiation across the fjords, with a possible main cluster exhibited by some populations tightly grouped. This corresponds with the locations of fjords these samples are taken from, with these fjords relatively quite close in proximity. It appears the great differentiation is seen within the NY13 population, and this may be causing it to be clustering away from the other fjords. Geographically this population

is greatly separated from the rest of the fjords, and so this may account for such differentiation.

The DAPC plots produced for this data show fjords more spaced out, but still it appears that the NY13 population is more separated from the rest. This could again infer some separation between this population and the rest of the fjords.

When it comes to the structure analyses, it appears that such methods are unable to pick up on these same structural differences that lead to patterns shown in the PCA and DAPC plots. Overall, there is a trend for *K* = 1, and thus there appears to be no separation of the different fjords. Despite this, with the *LEA* plots, and slightly shown in *Structure*, we can see the NY13 population drawing out from the rest of the fjords. Because of this, we cannot ignore the deviation exhibited by this population. Although there may be one fjord population overall, there is some differentiation in this NY13 population which could increase over time and potentially should be monitored for this. It is also possible that we are not seeing a full picture here as we do not have as many fjord individuals as we do those from the Atlantic.

6.5. Outlier loci in the common ling; population structure

Population analyses were repeated on the outlier dataset. Through outlier analyses 3 SNPs were identified in the common ling MAF = 0.05 dataset.

Table 37 presents the pairwise F_{st} values for the common ling outlier dataset. We see here that values are much higher than that of the neutral dataset, but this bias can be expected with such a small number of outliers.

Table 37: Pairwise F_{st} values for common ling populations using the outlier SNP dataset

	BB15	*BE08*	*BO14*	*HA14*	*IB13*	*NY13*	*RA08*	*RA14*	*RYF13*	*RYV14*	*SO14*	*TF05*
BB15		0.212	0.399	0.400	0.429	0.429	0.363	0.276	0.333	0.485	0.195	0.336
BE08	0.212		0.118	0.178	-0.134	0.104	0.263	0.047	-0.017	0.069	-0.043	0.115
BO14	0.399	0.118		-0.018	0.226	-0.032	0.169	-0.026	-0.130	0.070	0.114	-0.055
HA14	0.400	0.178	-0.018		0.212	0.039	0.072	-0.034	-0.140	0.180	0.135	-0.054
IB13	0.429	-0.134	0.226	0.212		0.273	0.018	0.037	0.143	0.347	-0.208	0.030
NY13	0.429	0.104	-0.032	0.039	0.273		0.280	0.007	-0.212	-0.006	0.142	-0.026
RA08	0.363	0.263	0.169	0.072	0.018	0.280		0.047	0.016	0.399	0.140	0.097
RA14	0.276	0.047	-0.026	-0.034	0.037	0.007	0.047		-0.117	0.107	0.013	-0.065
RYF13	0.333	-0.017	-0.130	-0.140	0.143	-0.212	0.016	-0.117		0.025	-0.004	-0.252
RYV14	0.485	0.069	0.070	0.180	0.347	-0.006	0.399	0.107	0.025		0.153	0.087
SO14	0.195	-0.043	0.114	0.135	-0.208	0.142	0.140	0.013	-0.004	0.153		0.085
TF05	0.336	0.115	-0.055	-0.054	0.030	-0.026	0.097	-0.065	-0.252	0.087	0.085	

In Figures 27a and 27b we see PCA and DAPC of the outliers only dataset. With only 3 SNPs in total we may not expect the analysis to present any structure, but we can see a vague similarity with our PCA analysis. It appears to show BE08 and SO14 outstretching from the main cluster, like in the original analysis, and we can also see RA08 deviating slightly. These same two populations (BE08 and SO14) also deviate from the main cluster in the DAPC. This differs from the original DAPC but has more similarity with the PCA.

The structure analyses of this data is as expected, with no structure found by the *Structure* or *LEA* analyses. *Structure harvester* (Fig. 28) gives a plot with a more certain value of $K = 1$ than the previous analyses.

It should also be noted that the PCA and DAPC analyses remove missing data, and we see this here with our plots showing less individuals than we have sampled. This is due to the extremely small number of SNPs, which are not represented in all individuals. This makes the analysis redundant.

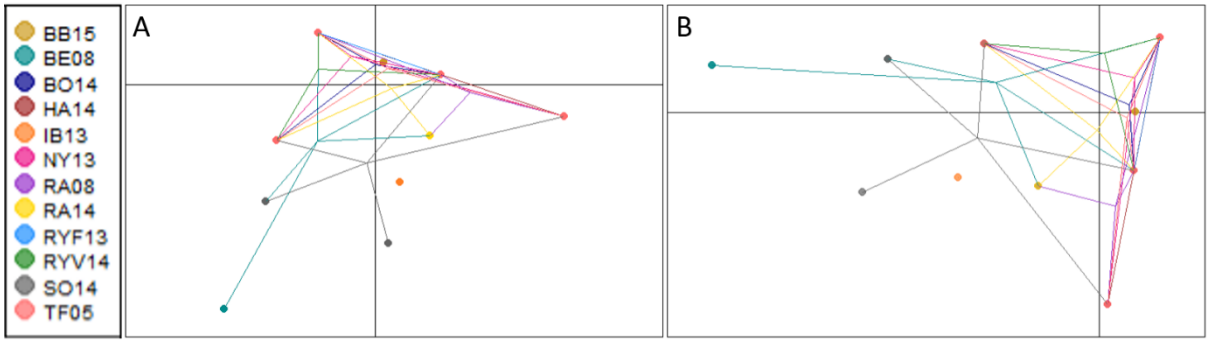

Figure 27: Analyses for the common ling populations with outlier SNPs alone; a) principal component analysis and b) discriminant analysis of principal components

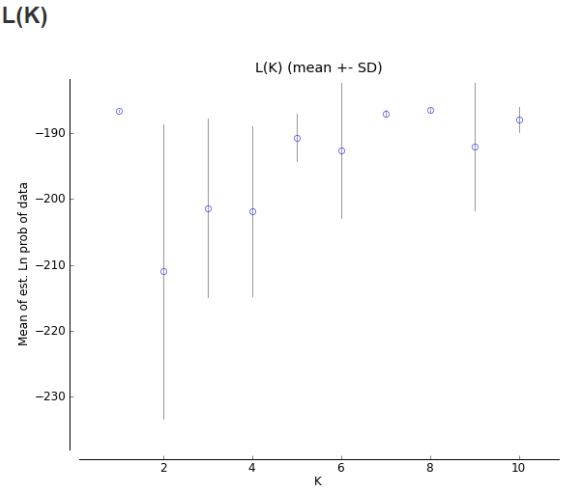

Figure 28: Structure harvester plot for analysis of common ling data with outliers alone

6.6. Outlier loci in the blue ling; population structure

With the blue ling we again used the MAF = 0.05 data, creating a dataset with the five outliers only. Again, the F_{st} values calculated are higher than those from the neutral dataset, which can be expected with so few SNPs to work with (Table 38). With the outlier data alone, we get a PCA plot which shows some similarities to that of the original data. Although the two main clusters join into one, the same two individuals deviate from the main cluster; HA14 and SL14. The same two are also seen in the DAPC, so again we produce a DAPC plot more similar to the PCA than to the original DAPC of the full dataset. The structure analyses again found no structure; the same with *harvester*.

Table 38: Pairwise F_{st} values for blue ling populations using the outlier SNP dataset

	AD07	*BO14*	*HA14*	*IB13*	*NY13*	*RA07*	*RA10*	*RA11*	*RS07*	*RYF13*	*RYV14*	*SL07*	*SL11*	*SL14*	*SO13*	*GRE15*
AD07		0.034	-0.732	0.000	0.399	0.273	0.322	0.297	0.588	0.600	0.280	0.446	0.092	-0.372	0.000	0.166
BO14	0.034		0.002	0.034	0.244	0.171	0.254	0.217	0.292	0.132	0.123	0.246	0.121	0.033	0.133	0.175
HA14	-0.732	0.002		-0.578	0.148	0.087	0.298	0.265	0.357	-0.105	0.072	0.337	0.124	-0.241	-0.700	0.172
IB13	0.000	0.034	-0.578		-0.057	0.429	0.419	0.370	0.564	0.600	0.138	0.520	0.243	-0.177	0.000	0.397
NY13	0.399	0.244	0.148	-0.057		0.247	0.288	0.270	0.284	0.203	0.096	0.313	0.201	0.194	-0.059	0.276
RA07	0.273	0.171	0.087	0.429	0.247		-0.037	-0.019	0.005	-0.135	-0.004	-0.024	-0.025	0.051	0.250	-0.097
RA10	0.322	0.254	0.298	0.419	0.288	-0.037		-0.005	0.021	-0.087	0.016	-0.004	0.007	0.185	0.190	-0.013
RA11	0.297	0.217	0.265	0.370	0.270	-0.019	-0.005		0.023	-0.076	0.018	0.000	0.013	0.145	0.165	0.004
RS07	0.588	0.292	0.357	0.564	0.284	0.005	0.021	0.023		-0.121	0.027	0.010	0.053	0.257	0.441	0.032
RYF13	0.600	0.132	-0.105	0.600	0.203	-0.135	-0.087	-0.076	-0.121		-0.091	-0.075	-0.041	-0.066	0.600	-0.148
RYV14	0.280	0.123	0.072	0.138	0.096	-0.004	0.016	0.018	0.027	-0.091		0.032	-0.002	0.071	-0.082	0.014
SL07	0.446	0.246	0.337	0.520	0.313	-0.024	-0.004	0.000	0.010	-0.075	0.032		0.016	0.227	0.345	-0.011
SL11	0.092	0.121	0.124	0.243	0.201	-0.025	0.007	0.013	0.053	-0.041	-0.002	0.016		0.075	0.045	-0.016
SL14	-0.372	0.033	-0.241	-0.177	0.194	0.051	0.185	0.145	0.257	-0.066	0.071	0.227	0.075		-0.285	0.106
SO13	0.000	0.133	-0.700	0.000	-0.059	0.250	0.190	0.165	0.441	0.600	-0.082	0.345	0.045	-0285		0.179
GRE15	0.166	0.175	0.172	0.397	0.276	-0.097	-0.013	0.004	0.032	-0.148	0.014	-0.011	-0.016	0.106	0.179	

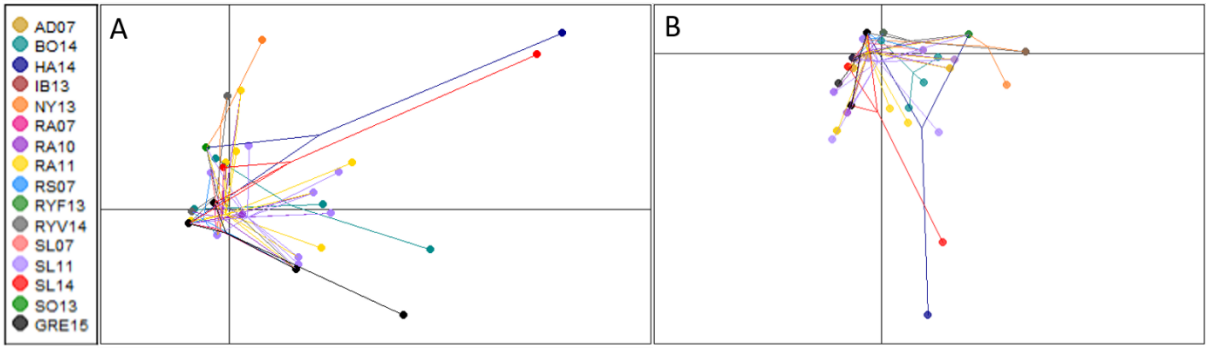

Figure 29: Analyses for the blue ling populations with outlier SNPs alone; a) principal component analysis and b) discriminant analysis of principal components

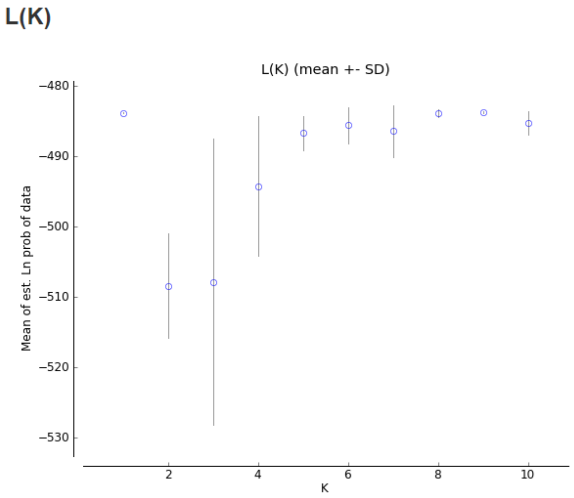

Figure 30: Structure harvester plot for analysis of blue ling data with outliers alone