Population genomics of two deep sea sharks: Centroselachus crepidater and Deania calcea

KEGGIN, THOMAS,JAMES

Population genomics of two deep sea sharks: *Centroselachus crepidater* and *Deania calcea*

Thomas J. Keggin

Master of Science by Research

Supervised by Prof. Rus Hoelzel and Dr Martyn Lucas

School of Biological and Biomedical Sciences

Durham University

2017

Abstract

This study investigated two superficially similar species of deep sea sharks, *Centroselachus crepidater* and *Deania calcea*, with the purpose of better understanding the drivers of differentiation and subsequent biodiversity in the deep sea. To achieve this four sample sites were used to coarsely represent the Atlantic and Pacific distribution of each species, with the genomes of each individual being sampled using a ddRADseq protocol. From these genome samples thousands of single nucleotide polymorphisms (SNPs) were identified and bioinformatically separated into neutral and outlier (under possible selective forces) marker sets which were then investigated for patterns of population genetic structure and adaptive function.

Despite comparable life history traits and overlapping sample sites, the two species showed differing patterns of differentiation across their range. In the neutral loci, *C. crepidater* showed significant differentiation between the Pacific and Atlantic, and homogeneity in the Atlantic. *D. calcea* showed effective panmixia across all sample sites. In the outlier loci, *C. crepidater* again showed Atlantic-Pacific differentiation, but also a split between the Rockall Trough and the Mid Atlantic Ridge. *D. calcea* also showed outlier differentiation between the Atlantic and Pacific, as well as significant divergence between two spatially similar sample sites in the Rockall Trough. Two possible proteins were identified linked to adaptive function, which were not insightful in themselves, but provide a starting point for future study.

These data combined with previous work on this deep sea taxon suggest that there is variation in patterns of differentiation between closely related species. However, common to all species so far studied are apparent high levels of vagility and dispersive capacity across global ranges. This provides further evidence that geographical barriers to dispersal are of lesser importance in determining genetic structure in deep sea basins, whilst isolation by distance and diversification along environmental clines likely play a more predominant role.

Word count: 299

Contents

List of Tables

List of Illustrations

Acknowledgements

There are too many people I would like to thank who have helped me learn and work through this project. Primarily, Rus for being willing to take me on in the first place and having faith that I was getting on with work, despite my sporadic attendance at the department. To Sean Twiss and Bob Baxter for their reassurance. Also, a huge thank you to Michelle Gaither, without her patient guidance I would have fallen flat on my face within the first couple of weeks and most likely missed the deadline for the sequencing run, and grossly misunderstood a great deal of the bioinformatics utilised in the project. Also, to Shelley Jones who was willing to unconditionally lend a hand in a suspicious feat of pure altruism. Further thanks for everyone in the lab for their productive discussions, advice, and reminders that it is not unusual for things not to work the first time. Especially to Menno de Jong for frank and insightful conversation and Daniel Moore for regular injections of fishy enthusiasm. My housemates, Mario Erandi Bonillas Monge, Monica Olguin Villa, and Sam Hibdige also need thanks for unfaltering maintenance of morale. Finally, a massive thank you to Natalie Allum, for her stoic proof-reading and encouragement throughout.

Chapter 1: Introduction

It was not 200 years ago that the deep sea was declared devoid of life (Forbes, 1844) and that Thomas Huxley famously argued that "probably all the great sea fisheries, are inexhaustible". It was not until even more recently that the idea of a slow-evolving, species-poor deep sea was debunked in light of the vast biodiversity being found at these depths (Hessler and Sanders, 1967; Danovaro *et al.*, 2014), possibly unsurprising considering the inaccessibility of the deep sea. That said, exploitation of the marine environment has been occurring for millennia, and our ability to heavily impact the oceans has only increased over time, especially with the relatively recent boom in the intensity and reach of the fishing industry (Van Andel, 1989; Jackson *et al.*, 2001; Marean *et al.*, 2007). This has resulted in population estimates of many exploited fish stocks lying around 10% that of pre-industrial levels, with many having being historically depleted and unable to recover (Myers and Worm, 2003).

Fishery Exploitation

The historical pattern of marine exploitation has not been uniform, with fisheries targeting the most easily accessible and in-demand species found close to major ports in calm and shallow-waters first, then moving through to those further away and harder to exploit (Myers and Worm, 2003). This has resulted in the overexploitation, and even depletion, of many shallow-water and pelagic fisheries (Pauly *et al.*, 2002). Subsequently, the attention of the fishing industry has recently turned to those species living deeper down, requiring greater fishing effort, but offering a virgin marine resource. The trend generally started during the 1960's, being pushed by a focus on various benthic and pelagic deep sea fish species; including orange roughy (*Hoplostethus atlanticus*), grenadiers (especially *Coryphaenoides rupestris*; Macrouridae), alfonsinos (Berycidae), pelagic armourhead (*Pseudopentaceros wheeleri*), and deep-water sharks (Morato *et al.*, 2006; Clark, 2009). It is notable that the trend is again not uniform, with Antarctica seeing the greatest rate of expansion into its deep waters at more than 100m each decade since 1960, attributed to area-specific collapse of shallow-water fisheries and the implementation of fishing restrictions (Morato

et al., 2006). The main concern many have voiced with this expansion into deeper waters is that deep sea fish species show life history traits that are overly vulnerable to conventional, intense fishery practices (Morato *et al.*, 2006). The deep sea offers a high pressure, low temperature, low energy environment in which a surprising diversity of species can be found (Grassle, 1989; Grassle and Maciolek, 1992). In order to cope with these conditions, deep sea organisms must have differing physiological and life history characteristics than their shallow-water relatives. Common amongst many deep sea fishes seem to be a delayed maturity, long life span and low productivity (Norse *et al.*, 2012). Koslow (1996) argues that there are two main adaptive strategies employed by deep-water fishes to deal with this environment; either fishes are dispersed and have slow metabolisms adapted for 'sit and wait' predation, or they have relatively faster metabolisms and partake in aggregating behaviours around geographical features for breeding and feeding, such as orange roughy (*Hoplostethus atlanticus*). The latter show a longer maturation period, thought to be the price paid for the higher metabolic costs (Koslow, 1996). Limiting growth rates is not only the low energy environment, but the high variability in resource availability. This variability has been hypothesised to be the driver for larger eggs, higher yolk reserves and lower fecundity in deep sea fish (Allain, 2001). Regardless, most of these species have longer life-spans, slow growth rates and are inherently vulnerable to rapid, high impact human exploitation pressures (Norse *et al.*, 2012), from which shallow-water species in higher energy environments, with higher growth and reproductive rates, may be able to recover (Garcia *et al.*, 2008).

Chondrichthyan Vulnerability

One group in particular that is especially vulnerable are members of Chondrichthyes, which are both actively targeted by some fisheries for their liver oils and are common bycatch (Dulvy *et al.*, 2008; Norse *et al.*, 2012). Particularly concerning are analyses by Garcia *et al.* (2008) indicating that only an average fishing mortality of 38-58% the magnitude required to drive a shallow-water chondrichthyan species to extinction was required to enact the same effect on deep-water species. Considering that even shallow-water chondrichthyan populations may be more vulnerable to the effects of fishing mortality than their teleost counterparts (Myers and Worm, 2005), this statistic is a major cause for concern. An example from the Southern Coast of Australia

shows a reduction in Harrisson's dogfish (*Centrophorus harrissoni*) survey trawl catch rate of more than 99% over a 20 year period of fishery exploitation, and an average global chondrichthyan decrease of more than 80% over that same period (Graham *et al.*, 2001). Aptly, the rapid population depletion of these vulnerable species has been likened more to mining than fishing (Norse *et al.*, 2012), far from a sustainable approach to managing environmental resources. This all said, it still remains difficult to make specific managerial arguments for species without species-specific information to inform these choices. It is therefore evident that this deficit between our impact on the marine biosphere and our knowledge of it needs to be closed if we are to contribute to the effective management of deep sea chondrichthyan species.

Evolutionary Processes

Initially, the great biodiversity of the oceans was at odds with its vast inherent openness, and it seems to have been initially assumed that the observed diversity must be the result of some form of unseen barriers between geographical regions or sheer distance, causing subsequent species diversity (Scheltema, 1968). The logical first step in investigating the oceans would be to compare patterns of variation with those known to exist in the terrestrial environment and applying that knowledge to this new environment.

However, terrestrial and marine systems fundamentally differ in ways that are likely to affect how diversity evolves. In terrestrial systems, the spatial distributions of species are dependent on their relationship with the surface substrate. Even airborne organisms cannot carry out their entire life history without interacting with this substrate. The purely pelagic life history traits of many marine organisms are not possible outside of this environment, with behaviours such as diel vertical migrations through the water column (Brierley, 2014) being absent in terrestrial systems away from a solid surface to support it. This gives much more profound three-dimensional spatial distributions to many marine organisms. This is despite obligately benthic species sharing many dynamics, such as elevation (depth), geographical barriers, environmental gradients, climatic variables (oceanographic processes) with species found on the terrestrial surface. For the majority of species not obligately associated with the benthos, the enhanced mobility and disassociation present in the marine environment may effectively annul many processes that are thought to be

important in the isolation and subsequent differentiation of organisms across a spatial scale (McClain and Hardy, 2010).

For example, a 2000m high mountain range in a terrestrial system offers a significant barrier to dispersal for most terrestrial organisms, but could be entirely permeable in the marine environment for pelagic species, or species with a pelagic life stage that can simply travel over the top. How these processes behave in a marine context is considered in the following.

Isolation by Distance

Following early, implied assumptions that the marine environment can be considered an open, barrier-less, three-dimensional space (Scheltema, 1968); a significant driver of differentiation could be a simple isolation by distance mechanism, whereby individuals of increasing geographic distance from one another become more highly differentiated as determined by reduced gene flow between them (Wright, 1943). This is not necessarily restricted to variation between populations; as even within large, geographically continuous populations division into genetic subgroups following a pattern of isolation by distance has been observed (Hardy and Vekemans, 1999). This dynamic may play quite a large role in determining observable genetic structure through species' ranges, especially considering large, highly vagile species such as fish, cephalopods, and cetaceans. In some cases, at least, this appears to hold true with patterns of genetic structuring following a pattern of isolation by distance being observed in several species (Pogson *et al.*, 2001; Planes and Fauvelot, 2002; Bradbury and Bentzen, 2007; Oscar *et al.*, 2009; Mendez *et al.*, 2010).

Historic Geological Events

The planet's oceans are not the summation of a large, featureless, three-dimensional space. The oceans have a volumetric limit and are made discontinuous by continental land masses and crust formation at tectonic interfaces. To marine species, these land-barriers offer a barrier that is impermeable to gene flow. This results in their significant contribution to the genetic structuring of the marine environment and the formation of oceanic diversity, with putative historical vicariance events and significant genetic structure between populations being attributed to changes in the nature of these barriers, i.e. the closure or opening of seaways between continental masses (Coates

et al., 1992). These barriers may interact with the isolation by distance dynamics, with a hypothetical species range being divided by landmasses, then subsequent gene flow occurring at restricted levels around this barrier as dictated by the distance required to circumvent it. Inextricably tied into this dynamic are the subsequent changes in patterns of oceanic circulation which in themselves can, through facilitation of dispersal and changes in environmental conditions (discussed below), influence the distribution and connectivity of species (e.g. formation of the Thermohaline Expressway; Strugnall *et al.*, 2008).

Oceanographic Influences

It is unlikely that physical barriers alone could be responsible for the huge wealth of variation found within large, continuous bodies of water. Early thought on the subject of marine differentiation between and within species seems to accept mechanics such as dispersal capacity and conditions that impact upon these at different mobile life stages played the largest role in determining differentiation (Scheltema, 1968). The prevalence of a passively dispersed life stage present in many marine organisms indicated that influences that impact on this life history stage may play a role in determining patterns of population connectivity and differentiation – especially for sessile species such as bivalves and corals whose only mobile life stage is planktonic. In short, the passive nature of this dispersal should result in dispersal distance and direction, and subsequent patterns of gene flow, being largely determined by oceanographic current systems (Becker *et al.*, 2007; Cowen and Sponaugle, 2009). This has led to a large amount of work attempting to predict population connectivity through the use of oceanographic models - sometimes referred to as "seascape genetics". Using the same framework as the established "landscape genetics", the idea is to pull population genetics and spatial ecological models together (Manel *et al.*, 2003; Galindo *et al.*, 2006) in order to help predict these patterns of differentiation based on oceanic models of water flow. It is important to note that models are often not solely based on larval dispersal and include life history traits, spawning, and other behavioural traits to better explain patterns of connectivity (Selkoe *et al.*, 2016). Further to this are the indirect effects and trophic interactions that current systems have on the distribution of organisms; for example, large, nutrient-rich coastal upwelling is

thought to be a major determining factor in the distribution of plankton and their predators (anchoveta) in the Peruvian anchovy fishery (Bertrand *et al.*, 2008).

Environmental Influences

Inextricably tied to these geographical barriers and oceanographic systems are the complex differences in environmental conditions across the marine space and their compounding impacts on marine variation.

The concept that conspecific individuals being acted upon by differing environmental pressures will differentially adapt to those specific conditions – driving genetic differentiation and even speciation – is not new (Rundle and Nosil, 2005). Importantly, this process of differentiation has been seen to occur in controlled systems whereby there are no barriers to dispersal (Rice and Hostert, 1993). Whilst some sessile species may be more reliant on oceanographic process for dispersal, this factor could be especially pertinent to marine species with high levels of vagility, widespread distributions, and insignificant barriers to dispersal. In fact, given the low likelihood that processes involving geographical barriers to gene flow play a significant role in the diversification in the open environment of an ocean basin, it could be quite likely that differential environmental pressures push divergent selection - resulting in differentiation in this space.

Ingram (2011) succinctly summarises the concept of speciation through divergence in the α- and β-niches of the environment, where the β-niche is comprised of macro-habitat differences and environmental gradients through space, and the α-niche consists of micro-habitat preference for different local resources, such as prey items. The question of which niche type of divergence is predominant in creating differentiation between individuals in a certain area can be applied to the problem of variation in the marine environment.

It could be that the marine environment offers a mosaic of differing environmental conditions and therefore habitats within a large, barrier-less space that does not impede gene flow through geographical obstacles. Even though individuals are free to move from one area to the next, the conditions differ such that non-specialised phenotypes are selected against as they move between areas - this would fit in with the β-niche concept of speciation. It could also be true that the marine environment is largely homogenous, but contains many opportunities for specialisation

to microhabitats or species interactions, allowing differentiating individuals to coexist in the same area– this time fitting in with the α-niche speciation concept. In truth, it would seem likely that the two processes work in tandem, with environmental gradients causing differentiation, but also fine-scale local α-niche variation providing opportunities to differentiate within the same region.

In order to address the above, it would be pertinent to know the ways in which this environment changes. To know which spatial vectors would provide the greatest amount of environmental change and habitat heterogeneity. If large enough environmental gradients exist in the oceans within a species' realised distribution, then a major driver of differentiation could be β-niche differentiation along those gradients. Headway has been made with one such gradient in depth down the water column. With increasing depth, there are vast changes in light, temperature, salinity, oxygen levels, and pressure. Coinciding with this gradient there is strong evidence of population differentiation through adaptation, as depth increases - including speciation events within lineages (Collins *et al.*, 2005; Ingram, 2011; White *et al.*, 2011; Drazen and Haedrich, 2012; Jennings *et al.*, 2013). This particular example provides sufficient evidence for a strong driver of differentiation over a relatively small spatial scale (<10km) within a comparatively huge continuous environment – potentially offering a powerful driver of the generation of the high diversity found in the oceans.

Relative Contributions of Drivers of Differentiation

Given that all of the drivers discussed above contribute to the generation of genetic diversity in the marine environment, the matter to be resolved is the relative importance of each process. Historically, there has been a heavy focus on larval dispersal as the major mechanism driving differentiation (Cowen *et al.*, 2007), which has shown great advances in our understanding of the dynamics of this particular, common life history trait (Selkoe *et al.*, 2016). However, even in species whose only mobile stage is larval, patterns of differentiation, at least at particular scales, can be predominantly determined by environmental gradients rather than patterns of larval dispersal. For example, Stuckas *et al.* (2017) found that the two mussel species, *Mytilus edulis* and *M. trossulus*, experience massive levels of interspecific gene flow in the Baltic Sea, but with differentiation occurring between the western and central regions. They found this to be over a

transition zone of high salinity change, and that larval dispersal could not be used to explain the barrier. So, despite a heavy focus on larval dispersal in predicting connectivity and differentiation in species with similar life history traits, in this particular instance the dominant driver for differentiation between individuals is a cline in environmental conditions.

However, species that do not utilise this planktonic trait, such as deep sea Chondrichthyans, remain underrepresented in the literature. Whilst improvement of genetic methods over the last 25 years have increasingly facilitated the study of this broad taxon (Dudgeon *et al.*, 2012), very few investigations into patterns of differentiation in their deep sea members have been carried out (Straube *et al.*, 2011; Veríssimo *et al.*, 2011; Cunha *et al.*, 2012; Catarino *et al.*, 2015; Gubili *et al.*, 2016). These studies seem to suggest extraordinarily low levels of differentiation across a global, geographic scale. That said, a possible bathymetric barrier to dispersal has been found in the Straits of Gibraltar, with species showing various levels of differentiation between the Mediterranean and Atlantic (Patarnello *et al.*, 2007; Catarino *et al.*, 2015). It has also been suggested that some sharks, including some deep sea species, show signs of female philopatry and male-mediated dispersal (Schultz *et al.*, 2008; Veríssimo *et al.*, 2012; Gubili *et al.*, 2016). Further to this, it has been argued that the existence of geographically uncorrelated clades distributed across various populations are an artefact of historical vicariance events (Catarino *et al.*, 2015). However, none of these studies into deep sea Chondrichthyans have utilised more recent Next Generation Sequencing (NGS) techniques, which provide much greater resolution for assessing fine-scale population structure (Jeffries *et al.*, 2016).

Given this lack of deep sea chondrichthyan data, it is also useful to consider patterns found in their deep sea teleost counterparts, which likely share common drivers of differentiation. For example, the genus *Coryphaenoides* shows differentiation across the Charlie-Gibbs Fracture Zone bisecting the Mid Atlantic Ridge, attributed to either the effect of separate gyre systems to the north and south of the Sub-Polar Front carrying larvae in different directions, or environmental differences across this region (White *et al.*, 2010; Ritchie *et al.*, 2013). This same genus shows an even greater level of differentiation down the environmental depth gradient, even to the point of lineage divergence (White *et al.*, 2011; Gaither *et al.*, 2016). In another teleost species, *Hoplostethus altlanticus*, investigations have concluded effective panmixia (White *et al.*, 2009), or

otherwise very low levels of differentiation following a pattern of isolation by distance (Varela, Ritchie and Smith, 2012, 2013), whilst another has shown fine-scale structure across discrete spawning-site associated topographic features (Carlsson *et al.*, 2011). Overall, these studies suggest that environmental gradients, isolation by distance, and behavioural traits are likely prominent drivers of differentiation in the deep sea for these mobile species. However, the contribution of larval dispersal remains unclear, with deep sea Chondrichythans internally brooding and mostly producing tens or less active, well-developed offspring (Simpfendorfer and Kyne, 2009), rather than the thousands of passive propagules produced in synchronised spawning events in *H. altlanticus* (Koslow *et al.*, 1995).

Methodological challenges

Sampling at these extreme depths is impractical and is mostly done through trawling. These trawls can be either dedicated research trawls, or samples can be taken opportunistically from the bycatch of commercial operations (Morato *et al.*, 2006; Norse *et al.*, 2012). Whilst deep trawling is an effective technique to acquire individuals, the depth change is usually fatal (Daley *et al.*, 2014) meaning that this methodology is inherently destructive. That said, non-destructive methods of collecting data regarding these species may also be possible. For shallower species that may survive the ascent and a subsequent descent, there has been some success in studies using tagging methodology developed for shallow-water elasmobranchs (Kohler and Turner, 2001). These have been used to support the development of marine reserves to aid the conservation of deep sea sharks (Daley *et al.*, 2014). Other non-destructive methods employed have been baited camera traps and remotely, or directly, controlled submersibles (Yau *et al.*, 2002; Roberts, 2002). With these direct imaging approaches, there is a great limitation on visual range, and the behaviours observed (Danovaro *et al.*, 2014). Although, cameras attached to trawl nets have successfully been used to discern differential behaviours associated with different types of gear, which then go on to affect estimates generated from catch data, e.g. overestimation of stock levels due to herding behaviour in flatfish (Bryan *et al.*, 2014).

Regardless, work on deceased samples can give us insight into their physiology, reproductive method and rate, age structure, abundance and realised niche, population trends, and

information on movement and trophic interactions using stable isotopes (Marques *et al.*, 2014). More directly relevant to patterns of differentiation, tissue samples from these individuals can be taken for use in genetic analyses and, aside from morphological studies, form the basis of resolving the relative importance of drivers of differentiation and speciation in the deep sea. This is becoming more accessible with significant progress having been made in the last century on population genetics and genomics, especially since the turn of the century with a significant increase in the capacity and affordability of genetic sequencing (Van Dijk *et al.*, 2014).

Genetic Analyses

Population genetics uses the patterns of heritable characteristics as biological markers to infer patterns of genetic differentiation between individuals. These markers started as morphological features, moved to allozymes, then specific targeted regions of mitochondrial and chromosomal DNA (Hedrick, 2011). However, since the development of Next Generation Sequencing (NGS), markers for population genetic applications are increasingly being sequenced through this relatively new technology (Tautz *et al.*, 2010; Narum *et al.*, 2013), which moves datasets from a few informative loci, to very high numbers of less informative loci – with an overall improvement in signal detection (Jeffries *et al.*, 2016). This relatively new technology also cuts out the necessary prior knowledge of the marker and the lengthy primer design steps required for previous marker types.

The method by which these markers are chosen or "discovered" remains under constant development and varies with desired application and available prior knowledge of the target genome (Davey *et al.*, 2011). Whilst it is possible, increasingly cheaper, and in some cases useful to sequence entire genomes (Wallberg *et al.*, 2014; Llorente *et al.*, 2015), for many applications it remains neither necessary, nor affordable. The current compromise is to sequence a subset of the genome (genome sampling) and apply established population genetics methods to this subset. This genomic sampling can be targeted, or non-targeted depending on prior knowledge of the genome. For targeted applications, whereby the genome is already well characterised, vast numbers of specific makers can be specifically chosen by developing bespoke microarrays for each application (Heller, 2002). However, if the genome remains unknown, then markers must be acquired blindly.

This is commonly done by fragmenting the genome using frequently cutting restriction enzymes, sequencing those fragments, and scanning them for Single Nucleotide Polymorphisms (SNPs). These techniques come under the umbrella of Restriction-site Associated DNA sequencing (RAD-seq) and follow the same basic fragmentation by restriction enzyme principle, but with various protocol modifications in order to adjust marker coverage, protocol efficiency, and target regions (Andrews *et al.*, 2016).

The first RADseq protocol to use an NGS platform was described by Baird *et al.* (2008) and uses one restriction enzyme to produce a variable number of fragments in multiple individuals, with the specific cut sites ensuring the same loci being represented in each individual. The addition of identifying barcodes to these fragments also allows multiplexing of large numbers of individuals into one Illumina sequencing lane to further reduce project costs. Whilst this basic technique allows scanning a subset of the genome for possible SNPs to use as markers for genotyping and as population genetic markers, only the reads starting from the restriction cut-site will likely start from the same site – as the fragments are randomly sheared and then size-selected to a rough window. A modification of this protocol has been put forward by Peterson *et al.* (2012), termed double digest RADseq (ddRADseq), whereby two restriction enzymes are used. Fragments here are selected by size and by those that are flanked by one of each restriction enzyme specific cut-site, eliminating the need of costly random shearing and enzymatic repair. This allows greater control over the number of fragments to be sequenced, and also gives a set position for the end of each fragment, allowing sequences from the end of fragments to be read in reverse – giving two reads per fragment and increasing the chance of SNP detection at each locus. This modified version also introduces the use of a secondary set of barcodes (termed "indices") to be ligated to the second cut-site, allowing for further multiplexing in a single sequencing lane.

Given a well annotated genome, these markers can then be aligned to the reference genome and their levels of function and linkage can be ascertained. Without a reference genome it is still possible to compare observed levels of heterozygosity (diversity) and F_{ST} to those that would be expected under a neutral model (Beaumont and Nichols, 1996; Antao *et al.*, 2008). This allows the separation of loci that fit the neutral model (those that are assumed to be neutral), and those that show a pattern of positive selection. Common population genetics statistics can then be applied to

the "neutral" set, such as F_{ST} estimates (Wright, 1951; Weir and Cockerham, 1984), individual population assignment probabilities (Pritchard *et al.*, 2000; Raj, Stephens and Pritchard, 2014), and ordination techniques to reveal patterns of neutral genetic clustering (Jombart, 2008; Jombart, *et al.*, 2010). The outlier markers, on the other hand, can be separately analysed using similar methodology (Andrews *et al.*, 2016). These markers may be either directly under selective forces, e.g. the SNP occurs in a coding region of a gene and causes a functional change in a protein, or could be linked to a gene under selection. In the former case, depending on the level of conservation of a gene, it is possible to match the locus in which the outlier SNP is found to known homologues and to an extent infer adaptive function. The patterns of differentiation in these outlier sets can also be directly compared to the patterns seen in the corresponding neutral sets. This comparison can highlight situations where patterns of gene flow and patterns of selection differ, e.g. adaptive differentiation with gene flow.

Aims and Hypotheses

The key objective of this study is to address the paucity of data concerning the predominant drivers of differentiation in the deep sea for species with mobile, viviparous traits, coupled with low reproductive output and slow rates of growth. To better understand these drivers, the aim is to apply relatively new, high resolution NGS technologies to deep sea chondrichthyans across their global range. This should give insight into the importance of reproductive mode in determining these patterns, as well as indicating common drivers of these patterns across taxa in the deep sea. Furthermore, this information will be useful in informing sustainable management and exploitation of these species, as greater understanding of geographical connectivity and regions important for critical life stages may be identified and implemented into conservation strategies.

The work described here aims to contribute a portion of this information, utilising a ddRADseq protocol (Peterson *et al.*, 2012) to resolve patterns of population structure in two superficially similar species of deep sea sharks, *Centroselachus crepidater* and *Deania calcea*, for which reference genomes are currently not available. Whilst one similar study has been carried out for *C. crepidater* using microsatellites and a fragment of the mitochondrial control region (Cunha *et al.*, 2012), no genetic data for *D. calcea* has been collected to date. The aim is to investigate both

patterns of neutral differentiation between samples taken across their global range, and patterns of differentiation in outlier loci that may be under the influence of selective pressures. These outlier loci can also potentially be used to uncover the basis of any adaptive differences between geographic locations.

Therefore, two hypotheses are to be tested:

1. Neutral variation will show effective panmixia across all sampled sites.
2. Patterns of differentiation found in outlier loci will differ from that of neutral loci, on the assumption that differentiation is caused not by barriers to dispersal and gene flow, but differences in environmental pressures between sample sites.

Chapter 2: Population genomics of the longnose velvet dogfish (*Centroselachus crepidater*) in the North Atlantic and South Pacific

Introduction

C. crepidater is a benthopelagic species with a patchy, global known distribution found over a depth range of 270-2080m (mostly deeper than 500m) along continental and insular slopes, as well as seamounts (Last and Stevens, 1994; Stevens, 2003a; Kyne and Simpfendorfer, 2007; Moore *et al.*, 2013). Body length does not usually exceed 100cm, and males and females are estimated to live up to 30 and 50 years, respectively. This sex-based life expectancy is reflected in age at maturation of roughly 9-15 years in males and 20 years in females (Stevens, 2003a; Irvine, Stevens and Laurenson, 2006), which is combined with an ovoviviparous habit, delivering a small litter size of 4-9 (mean of 6) after an unknown gestation time. Maximum embryo length and shortest neonate length were found to be around 30cm, with pups being born well developed. Furthermore, they have a non-continuous, aseasonal reproductive cycle of unknown periodicity, although a 1-2 year cycle is thought to be likely (Daley *et al.*, 2002; Irvine *et al.*, 2006).

Diet

Their diet indicates a generalist feeding strategy, preying upon a wide range of smaller organisms and detritus from the benthic zone to the mesopelagic – mostly consisting of cephalopods, other fish (mostly teleost), crustaceans and polychaetes (Mauchline and Gordon, 1983; Daley *et al.*, 2002; Pethybridge *et al.*, 2011). Interestingly, dietary variation between different geographic locations has been reported, with individuals at different sites focussing on different prey species. This is pronounced between sample sites in the waters of New Zealand and Australia, and off the coast of the British Isles, with individuals predating primarily on mesopelagic

fishes and various benthic micronekton, respectively (Mauchline and Gordon, 1983; Dunn *et al.*, 2010, 2013). That said, the study by Mauchline and Gordon (1983) of the Rockall Trough (British Isles) was carried out roughly 30 years prior to those in the Pacific. It is unknown as to whether large-scale disturbances to community assemblages in the deep sea since then may have affected prey preferences in each ocean (Anderson and John, 1999; Jennings *et al.*, 1999; Ruhl, 2004), as seen in Australian waters where a shift towards cephalopods and away from teleost fish was attributed to the decline of the teleost species (Caddy and Rodhouse, 1998; Pethybridge *et al.*, 2011). There was finer scale prey preference differentiation found between the Australia/New Zealand sample sites (Dunn *et al.*, 2013). Additionally, there was ontogenetic variation in dietary choice, with younger individuals having a greater preference for polychaetes and crustaceans, and no evidence for predation upon other fish (Pethybridge *et al.*, 2011). This ontogenetic variation could be explained by younger, and therefore smaller, individuals being unable to hunt for large, mobile species, resulting in reliance upon more sedentary benthic prey. Taken together, this suggests a flexible feeding strategy allowing for acceptable fitness across a range of predation strategies.

Migration

Migration and movement of this species is understudied, with no explicit investigations into geographical nor diel vertical migrations, and only a few inferences of behaviour from studies into other aspects of their biology. For example, it was noted in a biodiversity report that all *C. crepidater* individuals sampled from the Anton Dohrn seamount (British Isles) were large females in the latter stages of pregnancy, despite the trawl apparatus allowing for the catch of smaller individuals (Neat *et al.*, 2008). Additionally, Moore *et al.* (2013) and Clarke *et al.* (2001) collected individuals from the Rockall Trough which consisted of a female-biased sex ratio, with few of those from Moore *et al.* (2013) bearing pups. It seems likely then that this species has pupping and perhaps nursery locations to which females will travel and aggregate at synchronous times (Moore *et al.*, 2013). This sex bias was also recorded by Daley *et al.* (2002) in Southern Australian waters, with a higher proportion of females than males, although there was variation in this ratio. Unfortunately, whether this sex-ratio bias is due to differential sex-dependent behaviours, or

skewed survivorship of each sex is not known. Further, trawl data from the closely related *Centrophorus squamosus* in the North Atlantic shows habitat partitioning based on maturity stage and sex – which falls in line with the postulations above (Clarke *et al.*, 2001; Veríssimo *et al.*, 2012).

Whilst it remains unstudied for this species, it is known for individuals to undergo diel vertical migrations in other elasmobranch species, usually following prey up and down the water column (Papastamatiou and Lowe, 2012). Given their known feeding preference for mesopelagic fish and squid, this is also plausible for *C. crepidater*.

Predicting population genetic structure

Patterns of genetic structure are influenced by the life history of the organism and its interaction with the environment. From the known depth range of *C. crepidater* and their association with the mesobenthic zone, it is possible that they can minimise predation risk and energetic costs in this habitat. If so, this habitat dependency could result in isolated populations forming along restricted continental slopes, with gene flow only occurring among similar habitats. Another way the depth barrier could inhibit gene flow is through a shallow limit, if two populations were, for example, separated by a region of shallow-water, migrating individuals would have to ascend to perhaps inviable depths in order to migrate to the neighbouring population, promoting isolation. This could be the case for another closely related species, *Centroscymnus coelolepsis*, which at 11 microsatellite markers showed very little genetic differentiation throughout its global range, except for a unique, isolated population in the Mediterranean. It was hypothesised that this was due to the shallow depth of the Strait of Gibraltar which likely functions as a historic barrier to dispersal (Catarino *et al.*, 2015).

Another important factor may be that their ovoviviparous reproductive mode and birth of well-developed neonates (Daley *et al.*, 2002) means that there is no planktonic egg, nor larval, stage. These propagules can be dispersed by oceanic currents, as seen in other marine organisms including teleost fishes (Hickford and Schiel, 2003), although the relative importance of larval dispersal in determining population structure remains under contention, and is often species-specific (see Chapter 1; Munroe *et al.*, 2012; Derycke *et al.*, 2013; D'Aloia *et al.*, 2015). Also

worth considering is female parturition behaviour, as female philopatry and utilisation of nursery areas is well described in a variety of other elasmobranch species (Dudgeon *et al.*, 2012). The aforementioned female-skewed sex-ratios found in some trawl sampling regimes, as well as the high number of pregnant individuals at the same reproductive developmental stage could imply that pregnant females aggregate and possibly take part in group parturition. This requires coordinated, spatially dependent behaviour, and if nursery grounds are inherited by subsequent female offspring, geographic patterns of differentiation could emerge through this reproductive philopatry – regardless of the presence or absence of barriers to dispersal. However, whilst an earlier mtDNA study for *C. crepidater* found two distinct clades, they were evenly distributed amongst sample sites (Cunha *et al.*, 2012), showing no indication of spatial structure through the female lineage.

Diet could also promote isolation between populations as sufficient specialisation could limit migration between locations by diet-specialised individuals. However, this idea seems unlikely, given their generalist diet, ontogenetic dietary changes, and continued persistence through relatively long time periods during which prey assemblages have changed.

High vagility has been assumed for benthopelagic sharks in general (Musick *et al.*, 2004). Veríssimo et al (2012) proposed that low levels of population structure were likely for *Centrophorus squamosus*, given segregation based on sex and maturity stage, requiring individuals to travel between these segregated areas as they develop and in order to reproduce. For *C. crepidater*, a previous study by Cunha *et al.* (2012) used both mitochondrial and microsatellite markers to investigate genetic structure across the species range. They describe two divergent mitochondrial DNA clades, which they attributed to oceanic cooling and the closure of the Tethys Sea 15mya. Further to this, they found significant levels of genetic structure between the Atlantic and the Pacific in microsatellite markers, suggesting present-day barriers to gene flow – attributed to prey movements dictated by the Agulhas Current Retroflection (in which the westerly Agulhas Current retroflects back into the Indian Ocean off the southern tip of Africa).

These results imply that within oceanic basins there exist very few barriers to dispersal, and that the species is highly vagile and routes of migration are not limited to the continental slope along which it is distributed. The cause of the differentiation speculated by Cunha *et al.* (2012) to be an indirect oceanic effect remains speculative. Other possible drivers could include a pattern of

isolation by distance, a direct oceanographic process, an unforeseen geographic barrier, or differential environmental pressures between oceans that select against interocean migrants.

With the above in mind, the following aims to test the hypothesis that this non-larval, potentially highly vagile species displays limited, or no neutral differentiation across its global range, showing effective panmixia. This pattern, or lack of, would imply that geographic and oceanographic barriers to dispersal play at most a minor role in producing variation across a geographic scale. Instead life history traits would imply that any observed differentiation is likely due to differing environmental pressures across the species range and subsequent differential selective pressures. This leads to testing the second hypothesis, that any patterns of genetic differentiation are present in regions of the genome that are under selective forces, and that these will differ from patterns in neutral loci. Overall, the expectation is that a process of adaptive differentiation with gene flow between sample sites will be uncovered.

Methods

Sampling

Samples for this species were taken from four different sites, three in the Atlantic and one in the Pacific (Figure 2.1). All of the samples were collected by benthic trawl at varying depths, the details of which are summarised in Table 2.1.

Table 2.1: Location and depth of C. crepidater *sample sites.*

Sample Name	**Latitude**	**Longitude**	**Depth (m)**	**Sample Size**
Hebrides North	57.27	-9.50	1032-1500	14
Hebrides South	55.97	-9.41	1000	17
Mid Atlantic Ridge	42.54	-28.97	883-1100	5
Chatham Rise	-42.47	-179.41	844-1035	5

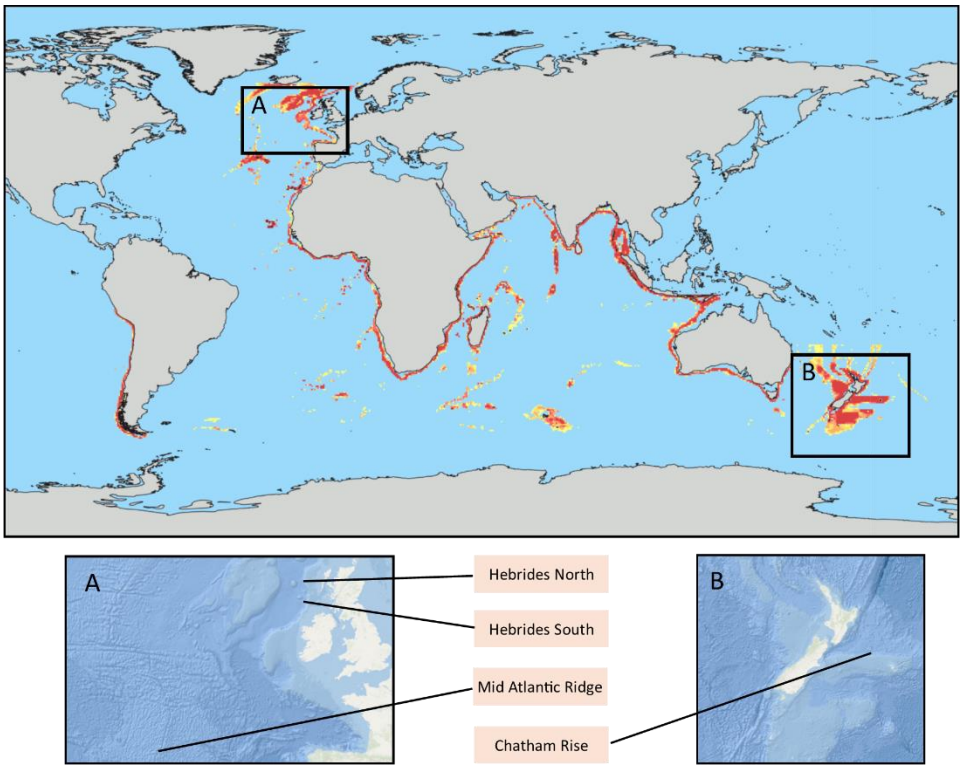

Figure 2.1: Sample sites and known distribution for Centroselachus crepidater.

Tissue samples of the caught individuals were then preserved in either 20% DMSO saturated with NaCL or in ethanol and sent to Durham University for storage.

Extraction

Both the samples from the Mid Atlantic Ridge and Chatham Rise were extracted by the author using the Omega bio-tek's E.Z.N.A. Tissue DNA Kit following the manufacturer's protocol. All but one of the Chatham Rise individuals yielded both poor quality genomic DNA (determined by gel electrophoresis) and low DNA yield (determined by Qubit Fluorometric Quantitation). The low DNA yields (<500ng) were mitigated against by carrying out repeated extractions of the same individuals and then pooling them into a single elute using GE Healthcare Life Sciences' Sera-Mag SpeedBeads, again following the manufacturer's protocol. Whilst this brought the amount of genomic DNA extracted above the required threshold of the library preparation protocol (Peterson *et al.*, 2012), it did not improve on the highly-degraded state of the samples. This same approach was taken for seven of the samples from the Mid Atlantic Ridge site which also gave a poor yield, although these samples did not appear to be nearly as degraded as those from Chatham Rise. The remainder of the sampled individuals from the Hebrides sites were taken from the lab archive, these

had all been previously extracted using a phenol-chloroform protocol and were all of immediately usable quality and quantity.

Identification of Mid Atlantic Ridge samples

Due to some ambiguity in the labelling of the Mid Atlantic Ridge samples species identity was confirmed using the COI gene. PCR amplification used the forward Fish2 and reverse Fish1 primers (Ward *et al.*, 2005). Reaction conditions comprised of 10.3µl H_2O, 4µl 5x GoTaq Flexi buffer, 2.4µl MgCl, 0.4µl dNTPs, 0.4µl F2 primer, 0.4µl R1 primer, 0.1µl of Promega's GoTaq G2 Flexi Polymerase, and 2µl sample genomic DNA extract. The thermal cycle for the PCR was then 5mins at 95°C, 35 cycles of; 30s at 95°C, 30s at 56°C, and 45s at 72°C, followed by 5mins at 72°C. These fragments were then sequenced and those sequences ran through Blastn (Agarwala *et al.*, 2016) for identity verification.

Library Preparation and Sequencing

The ddRADseq protocol developed by Peterson *et al.* (2012), based on original work by Baird *et al.* (2008), was followed with the following details. The restriction enzymes chosen were Sph1 and MluCI, which have a 6bp and a 4bp recognition site, respectively. These were chosen on the prior success of these enzymes with other vertebrate species, including fish and chameleons. There was no reference genome on which to base a more precise analysis. Sera-mag Speedbeads were used for DNA clean-up. A total of 75 individuals were pooled and multiplexed using a combination of 16 barcodes (first adaptor) and 5 indices (sequences incorporated in the second adaptor). During preparation, this gave 5 sub-pools for each index, 4 pools containing 16 individuals and the remaining pool containing 11 individuals. To account for this, the concentration of the remaining pool was altered to provide equal representation of all individuals in the sequencing run. A target fragment size of 375bps was chosen and size selection was done using Sage Science's Pippen Prep according to the manufacturer's protocol, the success of which was then checked using Agilent's 2200 TapeStation System, followed by concentration quality control

using qPCR with Bio-Rad's MyiQ™ Single-Color Real-Time PCR Detection System. The sequencing itself was then carried out using Illumina's HiSeq 2500 on a single lane.

Bioinformatic marker discovery

Stacks

The Stacks pipeline (Catchen *et al.*, 2011, 2013) was used to process raw reads and identify Single Nucleotide Polymorphisms (SNPs) for usage as markers for later analyses. Settings specified throughout the pipeline were as follows. During the process_radtags.pl program reads were trimmed down to 110bp in order to remove the poor-quality ends. The default setting to remove reads with a phred score of 10 or less was used, as well as removal of reads whose barcodes and restriction sites were not intact. This was followed by the three-in-one denovomap.pl program, which sorts reads into putative loci and identifies likely SNPs, using only the forwards reads. Reads were assigned to a stack only if there was a minimum of 3 identical reads. These stacks were then grouped into putative loci given a maximum of 2 mismatches between them within an individual, as well as only 2 mismatches permitted between loci when building the catalogue. The option to break up highly repetitive RAD-tags was also used. Individuals were then assigned manually to populations (sample sites) and run through the program Populations, in order to produce basic population genetics statistics and the data in common formats for subsequent analyses. Here, only loci with a minimum stack depth of 8 were retained as well as loci that were not present in at least 75% of individuals in each population. Prior to this step, individuals missing more than 80% of loci identified during denovo-map.pl were also removed. Of the 57,694 loci identified with a minimum stack depth of 8, an ultimate 3004 were retained after the loci filtering steps above.

Separation of neutral and outlier markers

Once these loci had been constructed and SNPs identified, loci were separated into two sets by the program LOSITAN (Antao *et al.*, 2008); neutral loci and outlier loci that are potentially under positive selection. In this instance, 95% confidence intervals were used to separate out

neutral and outlier loci, along with neutral F_{ST}, a forced mean F_{ST}, a 0.05 false discovery rate, and the infinite alleles mutational model selected. F_{ST} was approximated using 50,000 simulations. Markers with a *p*-value greater than 0.975 were thenceforth treated as outlier loci, and those with *p*-values between 0.975 and 0.025 treated as neutral loci.

Population structure

The multi-purpose program, Arlequin (Excoffier and Lischer, 2010), was then used to calculate population pairwise F_{ST} estimates using the Weir and Cockerham estimation (Weir and Cockerham,f84), using 50,000 permutations and a significance level of 0.05 (Bonferroni corrected). The R package, Adegenet (Jombart, 2008; Jombart and Ahmed, 2011), was used to carry out and plot a discriminant analysis of principle components (DAPC) for both the neutral and outlier sets of markers (Jombart *et al.* 2010). In this instance, two plots were produced. The first was a scatterplot of the two greatest discriminant functions of a DAPC using 17 principle components, with individuals assigned to their sample site groupings. The second was an assignment plot comparison between prior sample-site groupings and posterior group assignment based on discriminant functions. The DAPC analyses used to produce the assignment plots were the same as those used for the scatterplots.

Identifying genes under selection

It was also of interest to know if the SNPs identified by the Stacks pipeline were associated directly with specific genes, giving insight into how individuals from each sample group may be adapting to differential selective pressures. Two approaches were taken to identify SNPs that were potentially under positive selective pressure; the first used the aforementioned outlier-detection program LOSITAN (Antao *et al.*, 2008), using the same settings, but with a more stringent marker cut-off with a *p*-value of 1 being required. The second approach was using the Bayesian program, BayeScan (Foll and Gaggiotti, 2008), with a burn-in of 50,000, thinning interval of 10, 100,000 total iterations, and 20 pilot runs of length 5000.

Once SNPs that were likely under selection had been identified, the full sequence for that locus was then run through Blastx to align the nucleotide sequence to that of known or hypothetical

protein sequences. Since there is at the time of writing no reference genome available for *C. crepidater*, it was thought that aligning the sequence data to known nucleotide sequences might be misleading, despite the possibility of a marker being linked to a functional gene. Therefore, sequence matches were only searched for within functional genes using the Blastx nucleotide to protein sequence alignment tool (Agarwala *et al.*, 2016). Matches to queries were only then considered meaningful if the match had an E-value of less than 0.01.

Results

Separation of neutral and outlier markers

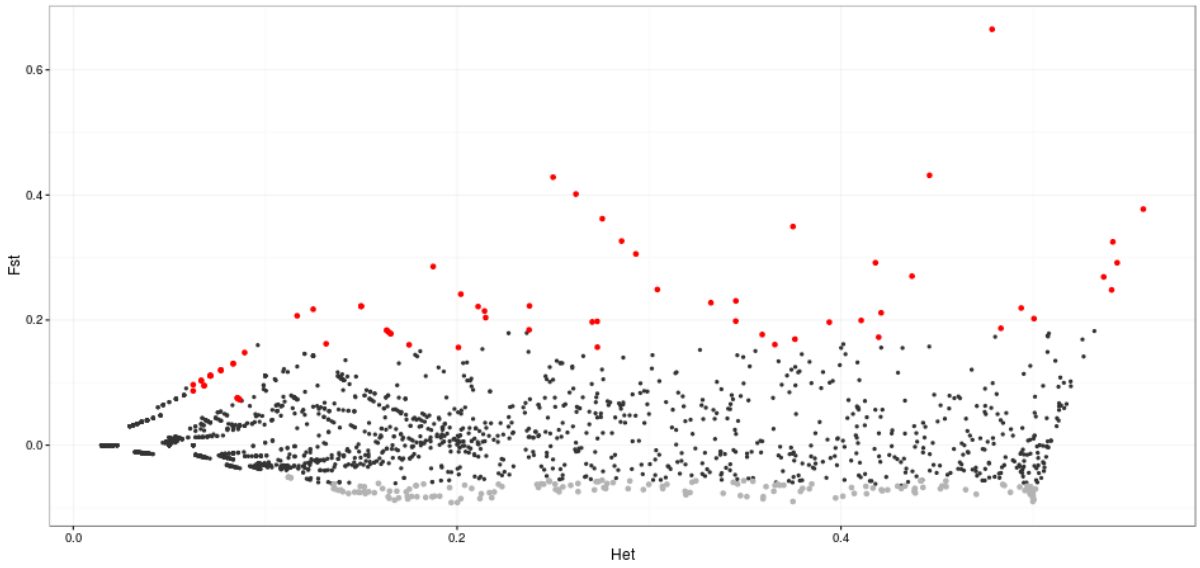

Figure 2.2: Identification of markers potentially under selection by plotting F_{ST} *against diversity (heterozygosity) using LOSITAN. Red markers are putatively under positive selection, black are neutral, and grey under balancing selective forces.*

Of 3004 SNP loci, 74 were identified by LOSITAN (Figure 2.2) as outliers and potentially being under positive selection, 2209 had values as expected with neutrality assumed and 721 potentially under balancing selection. Those of interest are the positive outliers which are putatively under the influence of positive selective forces and form the "outlier" group. Similarly, markers with neutrally expected values are used to form the "neutral" group, and those outliers potentially under balancing selection were discarded.

Pairwise F_{ST} Comparisons

Table 2.2: Pairwise F_{ST} *Comparisons. Below the greyed diagonal divide are values calculated from the neutral set of markers, whilst those above are calculated from the outlier set. Red indicates* $P<0.05$

	Hebrides North	Hebrides South	Mid Atlantic Ridge	Chatham Rise
Hebrides North		0.02589	0.14289	0.29721
Hebrides South	-0.00243		0.12610	0.27748
Mid Atlantic Ridge	0.00001	-0.00274		0.46227
Chatham Rise	0.01522	0.02056	0.00181	

At neutral loci the pairwise F_{ST} comparisons (Table 2.2) between the populations showed little differentiation. This is especially true among the Atlantic sample sites. Some comparisons are elevated between the Atlantic sample sites and the Pacific sample site, with statistical significance found between the Atlantic Hebrides South and Pacific Chatham populations. The outlier loci show a stronger pattern of differentiation, including some comparisons within the North Atlantic, but again strongest for comparisons between oceans,

Discriminant Analysis of Principle Components

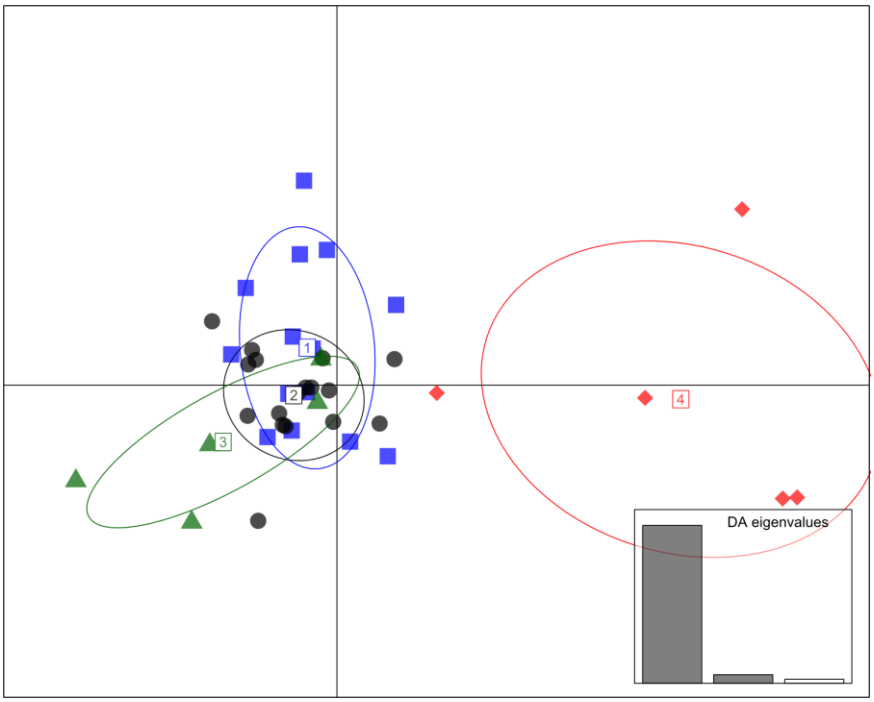

Figure 2.3: DAPC scatterplot of neutral markers of the two greatest discriminant functions. Blue squares (1) represent Hebrides North, black circles (2) Hebrides South, green triangles (3) Mid Atlantic Ridge, and red diamonds (4) Chatham Rise. Eigenvalues are 44.1, 2.42, and 1.1, respectively.

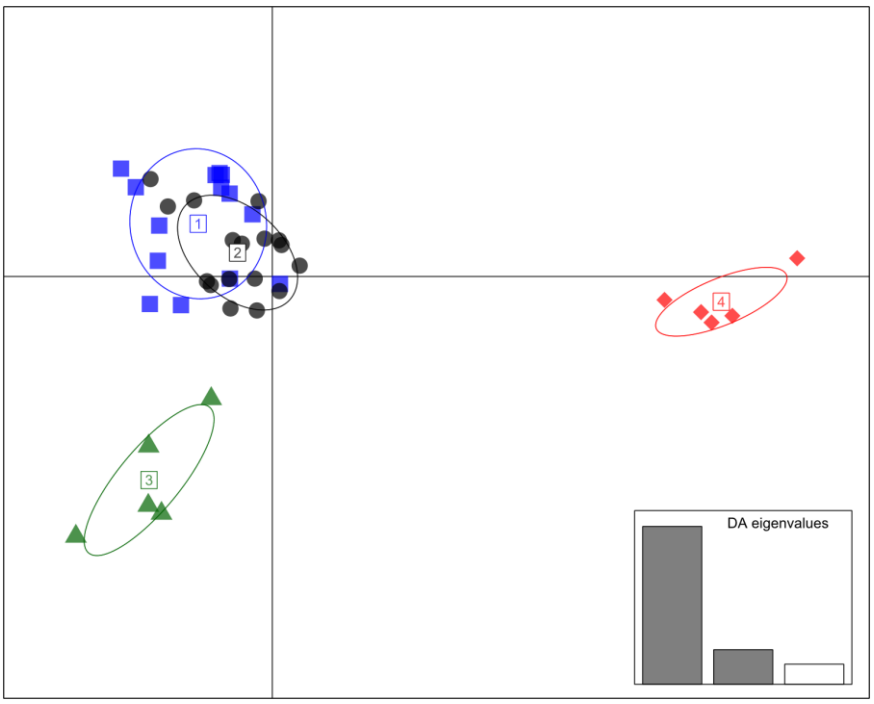

Figure 2.4: DAPC scatterplot of outlier markers of the two greatest discriminant functions. Blue squares (1) represent Hebrides North, black circles (2) Hebrides South, green triangles (3) Mid Atlantic Ridge, and red diamonds (4) Chatham Rise. Eigenvalues are 193.5, 42.44, and 24.71, respectively.

For the neutral loci (Figure 2.3) most differentiation is again seen between the Atlantic and Pacific samples. Some possible differentiation between the Mid-Atlantic Ridge and the Hebrides sites is poorly resolved due to small sample size at the Mid-Atlantic Ridge. Also notable is that the spread of the individuals of Chatham Rise is much greater than that of any of the Atlantic populations. Most of the variation was contained within the first discriminant function, suggesting that variation between groups was mostly defined along a single dimension, along the x-axis and the apparent divide between the Atlantic and the Pacific. The outlier markers (Figure 2.4) show a distinct pattern of differentiation. Hebrides North and Hebrides South are tightly clustered and greatly overlapped, with the individuals from the Mid Atlantic Ridge site being wholly separated from the cluster with no overlap of individuals. A similar, but greater, divide is also seen between the Chatham Rise individuals and the two Atlantic groups. Also, there is now relatively tight clustering of the Chatham Rise samples.

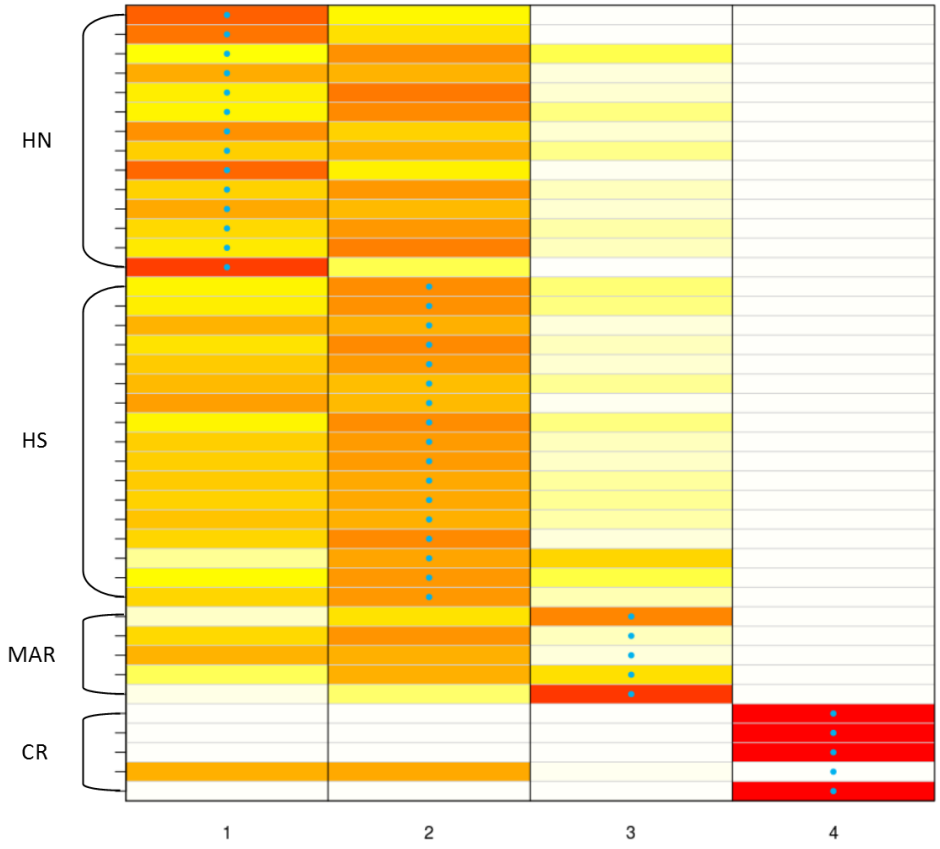

Figure 2.5: DAPC assignment plot for neutral markers. Blue dots represent prior group assignment based on sample site, whilst the heated cells represent group membership probabilities of each individual based on retained discriminant functions, with white representing 0 and red representing 1. HN = Hebrides North, HS = Hebrides South, MAR = Mid Atlantic Ridge, CR = Chatham Rise.

The assignment plots (Figures 2.5 and 2.6) reinforce the patterns seen in the scatter plots, showing how for neutral markers (Figure 2.5) the Atlantic and Pacific can be distinguished, but just a couple of samples show differentiation for the Mid-Atlantic Ridge. For the outlier set of markers (Figure 2.6), the two Hebrides populations appear to be highly homogenous, with nearly all individuals showing some likelihood of assignment to both Hebrides sample groups and more pronounced than for the neutral loci (Figure 2.5). For the Mid Atlantic Ridge site, almost all individuals showed total assignment to their prior grouping. The Chatham Rise individuals appear to be totally distinct and show complete assignment to their prior, sample-site grouping.

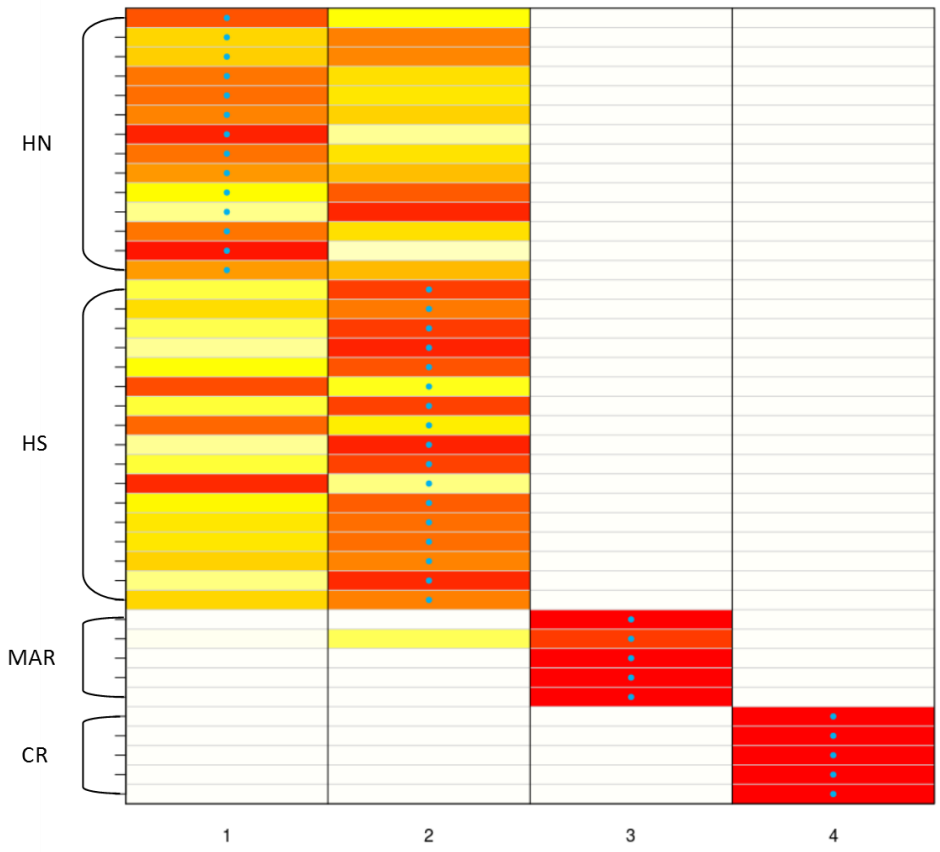

Figure 2.6: DAPC assignment plot for outlier markers. Blue dots represent prior group assignment based on sample site, whilst the heated cells represent group membership probabilities of each individual based on retained discriminant functions, with white representing 0 and red representing 1. HN = Hebrides North, HS = Hebrides South, MAR = Mid Atlantic Ridge, CR = Chatham Rise.

Outlier function

Of the two methodologies used to identify signals of selection, only LOSITAN identified likely outliers. Of the 20 loci identified as candidates under selection, three of the corresponding sequences matched an existing record in the NCBI database. Of those three, one matched too poorly to be considered, the second partially aligned to a zinc protein finger, but the SNP was found outside of the matched region and assuming the same reading frame, was functionally silent. The third locus, however, aligned sufficiently well to the conserved HARBI1 protein in a variety of species (channel catfish, coelacanth, African clawed frog). Assuming the same reading frame in *C.*

crepidater, the SNP causes an arginine to glycine substitution. There were three heterozygous individuals in the Hebrides North group carrying the glycine allele, whilst the rest were homozygous for the arginine allele.

Discussion

Neutral structure

For neutral variation, these RAD data were consistent with the Atlantic-Pacific divide described for this species by Cunha *et al.* (2012) based on microsatellite DNA markers. Cunha *et al.* (2012) also calculated that there was likely twice the number of migrants travelling from the Atlantic to the Pacific than in the opposite direction. Interestingly, of the five individuals representing the Chatham Rise population, one assigned more strongly to the Atlantic grouping than to the Pacific (Figure 2.5). This could be viewed as evidence of a contemporary Atlantic to Pacific migrant. That said, when looking at the same individuals through the lens of the outlier marker group, that same "migrant" is firmly assigned to its sample site prior. This does not necessarily discount the "migrant" from having relatively recently originated in the opposing oceanic basin, however, if the two sample sites represent very different environments with distinctly differing selection pressures, the lineage represented by this indication of admixture may have adaptively conformed.

Another finding by Cunha *et al.* (2012) was two deeply divergent mitochondrial clades present in both oceans, which was absent from their microsatellite DNA markers. They attributed this to a historical vicariance event during the Miocene over the stochastic retention of deep lineages. However, vicariance should be reflected in high resolution nuclear genetic data as well (as it reflects differentiation due to drift), including in the SNP markers used here. The data found here then supports this mitochondrial divergence through the latter process, with incomplete lineage sorting. A similar pattern of mitochondrial divergence was reported among populations of the African wild dog (*Lycaon pictus*; Girman *et al.*, 1993).

Focussing solely on the Atlantic individuals, all three sample sites appear to be highly homogenous, with no significant differentiation indicated from pairwise F_{ST} comparisons from

neutral loci (Table 2.2). Whereas there does appear to be some form of barrier between the ocean basins, there appears to be none present between Atlantic locations. This may indicate a highly migratory metapopulation, whereby particular regions are used by different demographic groups, dependent on age, sex, maturity and reproductive stage (Clarke *et al.*, 2001; Daley *et al.*, 2002; Neat *et al.*, 2008; Veríssimo *et al.*, 2012; Moore *et al.*, 2013). However, the ordination data (from the DAPC plot in Figure 2.3) suggest that the Mid Atlantic Ridge individuals are slightly more divergent from the two Hebridean sites than the Hebrides sites are from one another, with perhaps the greatest difference being between Hebrides North and the Mid Atlantic Ridge. Looking at the assignment plot (Figure 2.5), of the five Mid Atlantic Ridge individuals, two are entirely assigned to the Hebrides, whilst the remaining three show the highest likelihood of being reassigned to their prior sample group. Also, evident here is the Hebrides South individual that seems to show more affinity to the Mid Atlantic Ridge than its sister Hebrides group.

The reasoning behind this subtle, neutral, Atlantic patterning is cryptic, and perhaps would be clearer with a greater representation of the Mid Atlantic Ridge. As it stands, however, the slight differentiation of this sample site may simply be due to a relatively simple isolation by distance effect, with migration between this site and the Hebrides being not impossible, but more difficult than the ca. 100km between the two Hebridean sites. Also possible is that the North Atlantic Current, where it transects the Mid Atlantic Ridge at the Charlie-Gibbs Fracture Zone acts as an slight barrier to dispersal for this species, as has been found for the roundnose grenadier (*Coryphaenoides rupestris;* White *et al.*, 2010). This pattern supports Cunha *et al.* (2012), who initially revealed three distinct clusters in their microsatellite markers; Atlantic, Pacific, and Great Meteor Bank (south of the Azores), with the Great Meteor Bank cluster ultimately discarded due to low sample size and quality. If this pattern is in fact due to an oceanographic barrier to dispersal (North Atlantic Current), and not isolation by distance, the comparison with a species with passive larval dispersal (roundnose grenadier) is interesting. Given the similar patterns across this barrier, it implies that the difference in life history traits may not be important in determining patterns of dispersal and gene flow. Interestingly, a similar sampling regime has been used to investigate the closely related and similar Portuguese dogfish (*Centroscymnus coelolepis*; Catarino *et al.*, 2015) using putatively neutral microsatellite markers, this study found no structure between the

Hebridean and Mid Atlantic Ridge sample sites. Either there is weak structure and this species follows a pattern similar pattern to that seen in *C. crepidater,* or the two species with almost identical life history traits have different dispersal capacities.

Outlier structure

The structure uncovered in the outlier marker set is stronger than for neutral markers. However, there is the possibility that the outlier markers have abnormally high "outlier" F_{ST} values due to non-selective forces, i.e genetic drift (Beaumont and Nichols, 1996; Nielsen, 2005; Vitti *et al.*, 2013). In that case the outlier loci should express the same, yet exaggerated pattern of genetic structure as the neutral loci – as they are presumably being influenced by the same stochastic factors associated with isolation and genetic drift. Selection may instead generate a different pattern of structure than seen for the neutral markers. For this dataset, differentiation between ocean basin populations becomes more pronounced, but the main distinction between the neutral and outlier data is for the comparisons with the Mid-Atlantic Ridge sample. Whilst there was seemingly only marginal differentiation between the Hebridean sample sites and the Mid Atlantic Ridge in the neutral markers, the outlier markers show a much more defined split whilst the Hebridean sites remained clustered. The DAPC assignment plot shows almost as strong self-assignment of the Mid Atlantic Ridge individuals as those from Chatham Rise, with only one individual showing any level of similarity with the Hebrides. The DAPC scatterplot (Figure 2.4) shows this clearly, with the Hebrides sites still tightly clustered with one another, and the Mid Atlantic Ridge individuals forming a completely separate cluster. The sample size is small, but the effect is pronounced, and arguably suggests selection generating the differential pattern.

The interpretation of these data from this study would seem to suggest that, whilst gene flow is not necessarily inhibited between regions in the Atlantic, the Ocean offers at least two differing environments to which individuals must differentially adapt – making region-specific selection pressures a driving force behind differentiation. The one Pacific individual assigned to the Hebrides through its neutral markers seems to suggest that migration between ocean basins is possible, but infrequent, echoing the findings of Cunha *et al.* (2012) from their microsatellite markers. However, this individual assigns strongly to Chatham Rise for outlier loci. This may

simply reflect strong drift among the outlier loci, though there is the chance that this migrant lineage adapted to regional conditions in the Pacific, where environmental conditions are distinct (Priede *et al.*, 2013).

Outlier Function

Although genome sampling methods such as RADseq only provide data on a small proportion of the genome, it is possible that markers linked to or even within relevant functional loci can be identified. Some point out the limitations of this approach (Hoban *et al.*, 2011), but even a limited number of microsatellite DNA loci can sometimes identify outliers associated with functional adaptation to habitat (White *et al.*, 2009). With a reference genome, these markers could be mapped, and potential genes identified that may be responsible for the differential adaptive response. Unfortunately, no reference genome is yet available for *C. crepidater*, and so this undertaking is limited to identifying markers that are directly translated into known homologous proteins.

Interestingly, this limited approach identified one outlier marker to likely be embedded in the HARBI1 (Harbinger transposase derived 1) gene, with the polymorphism discovered being responsible for an arginine to glycine substitution. As the name suggests, the protein is thought to be derived from the highly-conserved *Harbinger* transposases found in many taxa. However, the exact function of HARBI1 proteins remains unresolved, but given that conserved motifs within the *Harbinger* transposases are also present in HARBI1 proteins, it is expected that HARBI1 has nuclease activity (Kapitonov and Jurka, 2004). It has also been hypothesised that, since transposons derived from the encoding transposon of the protein most similar to HARBI1 (*Harbinger3_DR*) are characterised as being highly specific for a 17bp target site, HARBI1 is a nuclease with high target site specificity (Kapitonov and Jurka, 2004). The amino acid substitution from arginine to glycine - from a large, polar R-group to an R-group comprised of a single hydrogen – would make it plausible that this substitution would have some sort of effect on protein structure and therefore function. The nature of this structural and functional change and how that impacts the overall fitness of the carrier individuals is, however, not known. In this sample set, only three individuals carried the glycine variant, all of whom represented Hebrides North and were heterozygous at that

locus. Therefore, this site in particular may not be relevant to a functional difference between the Hebrides and Mid-Atlantic Ridge populations, though this or other SNPs may be linked to relevant loci.

Overview

The results laid out here point towards a mixed set of dynamics. Whilst there is a clear break between the Atlantic and Pacific, there is also limited structuring within the Atlantic. This could point towards a process of isolation by distance, barriers to dispersal, or both. Within the Atlantic, it is possible that dispersal between the Hebrides and the Mid Atlantic Ridge site is in part restricted by the North Atlantic Current, in a similar fashion to patterns found in a comparable teleost species. This suggests that the presence of passive larval dispersal plays a minor role in determining patterns of differentiation across a species range. The differing patterns between neutral and outlier suggest adaptive differences, implying that differential environmental pressures between locations are driving differentiation, despite evidence of gene flow. This is perhaps not surprising given the sample site locations as the Hebrides sites and the Mid Atlantic Ridge are found on either side of the Charlie-Gibbs Fracture Zone/Sub-Polar Front that bisects the Mid Atlantic Ridge between 52° and 53°N (Hebrides and Mid Atlantic Ridge sites at 55-57°N and 42°~N, respectively). Either side of this formation have been shown to offer different environmental conditions (Priede *et al.*, 2013) and in themselves form a barrier to gene flow for some, but not all, deep sea species (White *et al.*, 2009; White *et al.*, 2010; White *et al.*, 2011; White *et al.*, 2011; Teixeira *et al.*, 2012; Knutsen *et al.*, 2012; Ritchie *et al.*, 2013). The exact mechanism of the barrier is not known, and even congeneric species with the same distribution across the front have shown distinct patterns of differentiation, possibly associated with habitat depth (White *et al.*, 2010; White *et al.*, 2011). The exact nature of this barrier, and the relative contributions of levels of gene flow and adaptive pressures to differentiation across it have been unclear, possibly because the structuring is weak and can only be identified through more recent methodologies (Jeffries *et al.*, 2016). Here, however, there is strong evidence of differentiation through environmental pressures, and possibly the restriction of gene flow across the current system or sheer distance. The exact cause of the division between Atlantic and Pacific sample sites remains unresolved.

Finally, regarding fishery exploitation and bycatch, the neutral split between ocean basins should encourage distinct management policy for each region, whilst the distinct outlier groupings suggest spatially distinct adaptive phenotypes containing variation worth conserving, despite varying levels of gene flow.

Future studies

Perhaps needless to say, a more inclusive sampling of regional populations and larger sample sizes would be useful to better reflect the wider patterns of genetic diversity. Whilst the results of this study provide preliminary evidence of possible isolation by distance or oceanographic barriers to dispersal, and strong evidence for differential adaptive responses within and between oceans, these patterns are not fully explored. The species is widely distributed throughout the Indian Ocean and eastern Pacific, but the current data, from both this study and that of Cunha *et al.* (2012), are only representative of the South-West Pacific and Northern Atlantic. Cunha *et al.* (2012) ascribed the Atlantic-Pacific divide to the indirect influence of the Agulhas Current Retroflection (Cape of Good Hope), whilst another possible avenue of dispersal would be through the Drake Passage (South America). The extent of an isolation by distance effect would be shown by greater representation of intermediate sample sites between the North Atlantic and South-West Pacific. If this dynamic is a key driver of differentiation, then a continuum should be apparent across the species range, with breaks in this continuum revealing barriers to dispersal, such as the suggested role of the Agulhas Current Retroflection.

Currently, there are only three resolved elasmobranch genomes; elephant shark, whale shark, and the little skate, none of which represent the diverse Order Squaliformes to which many representatives of the deep sea belong. A reference genome of this species, or of a close relative would also greatly support the investigation of signals for selection and place greater confidence in assumptions of neutrality. It would also allow for greater control over marker discovery, patterns of linkage, and allow more accurate taxonomic resolution of these species. In terms of population genetics, it could allow discernment of the drivers that lead to possibly differing patterns of differentiation and environmental tolerance between many of these superficially similar sharks, in

turn helping elucidate the drivers of differentiation and the contribution of life history traits on these processes.

Potential depth and regional preference based on ontogenetic, maturity, and reproductive stage are underlying factors that would most likely determine population genetic patterns. Understanding these dynamic, seemingly long-distance behavioural life history traits would complement genetic data greatly. This could be investigated through depth-targeted and date-controlled sampling regimes, and if the technology becomes available, tagging programmes to follow the migratory behaviour of individuals through their life cycle would present a much clearer, contemporary picture of the seemingly complex migratory behaviours of this species. Again, these data would be directly comparable to superficially similar, closely related species to help discern the extent to which behavioural patterns influence patterns of differentiation in the deep sea.

Conclusion

Overall, the neutral loci found in *C. crepidater* show Atlantic homogeneity, but marginal differentiation between the Atlantic and Pacific. In the outlier loci, however, the two ocean basins are clearly differentiated, as well as a signal for the differentiation between the Hebridean sites and the Mid Atlantic Ridge. The causal drivers of this pattern of differentiation remains unclear, which could be a result of isolation by distance effects, differential environmental pressures, a barrier to dispersal across the North Atlantic Current and between oceans, or a combination of these working in concert.

Chapter 3: Population genomics of the birdbeak dogfish (*Deania calcea*) in the North Atlantic and South Pacific

Introduction

Deania calcea remains largely understudied, especially in respect to migratory behaviour. The species has a maximum length of around 120cm, and a life expectancy and age at maturation dichotomy between the sexes; with life expectancy being 24-32 and 31-36 years, and age at maturity being 10.5-15.5 and 17.5 – 21.5 years for males and females, respectively (Irvine *et al.*, 2012). Furthermore, the sex difference also includes a size difference, males are at a smaller 54-109cm, and females at 56-117cm (Clarke *et al.*, 2002). Their depth range is relatively restricted, with individuals mostly found between 660m and 1100m, but records also exist of individuals caught between a depth range of 70m to 1450m (Stevens, 2003b). This species is ovoviviparous, with a mean litter size of 6 and a maximum litter size of 10 being observed. The reproductive cycle is aseasonal, with a known rest period after pregnancy – leaving reproductive cycles thought to be slow, estimated to be probably between 1 to 3 years, or even up to 4 years between pregnancies (Clark and King, 1989; Irvine *et al.*, 2012).

Diet

D. calcea has a relatively restricted diet, focussing mainly on cephalopods and teleost prey species. There is a preference for mesopelagic teleost fishes, a lesser preference for crustaceans and cephalopods, and in general targets fewer prey species than similar relatives. There are also relatively few unknown species recorded in their diet, indicating a specialised predation strategy (Pethybridge *et al.*, 2011). This focus on mesopelagic fishes perhaps indicates a relatively weak association with the benthic zone. The overall consensus for this species is that the bulk of their diet is made up of teleost fishes (especially Myctophids), but that they are also capable of predating

on squid and crustaceans should the need arise (Blaber and Bulman, 1987; Ebert *et al.*, 1991; Yano, 1991; Saldanha *et al.,* 1995). Most authors report a high proportion of those caught to have empty stomachs, which could be interpreted as infrequent feeding and therefore employing a lower energy life strategy, however, it was suggested by Yano (1991) that this could be due to the washing out of the stomach by water as they were being hauled. *D. calcea* does not appear to have much variation in diet between sample sites, despite the above referenced studies sampling from a wide range of locations (Japan, Australia/New Zealand, Namibia, British Isles), further suggesting a rigidity in diet of this species. Ontogenetic differences in prey specialisation is currently unknown.

Migration

No direct evidence exists to explain migration patterns of *D. calcea*. That said, there is evidence of differential habitat usage by different demographics at different times, which would require individuals to migrate from one location to another as they develop, or engage in mating behaviour. It was noted by Clarke *et al.* (2002) that in their survey of *D. calcea* in the Rockall Trough and the slopes of Porcupine Bank in the British Isles smaller and younger individuals were not represented, also missing were gravid females, with the majority of females caught being immature. Additionally, they noted that it was not until greater depths that the male-dominant sex-ratio found at shallower depths was equalised. They compared these observations to the survey by Machado and Figueiredo (2000) off the coast of Portugal, where an entirely different set of demographics was represented, including the "missing" younger individuals. This, they suggested, was evidence of migration along continental slopes between each area. They further built this case by citing the progressive increase in abundance along Chatham Rise, New Zealand (Wetherbee, 2000), again suggesting that this could be indicative of a pattern of migration. Another report by Clark and King (1989) found an association of the smallest individuals with large females down to a limit of 800m also off the coast of New Zealand, this was followed by a progression in these individuals from West to East around North Island – a pattern claimed by Clarke et al (2002) to show extensive migrations associated with reproductive behaviour.

Predicting population genetic structure

Many of the arguments for postulating structure for *C. crepidater* (see Chapter 2) can be applied to the possible structure present in *D. calcea*. Both species share a similar depth range and are associated with the mesobenthic zone. It is possible that *D. calcea* is adapted to minimise energetic costs and benefit from a reduced predation risk in this habitat. This dependency may then result in isolated populations forming along these continental slope regions, and gene flow being restricted along habitat corridors. A comparison with the Portuguese dogfish is possible (*Centroscymnus coelolepsis*; Catarino *et al.*, 2015), in that a shallow limit may provide a bathymetric barrier to gene flow. However, with individuals having been sampled from less than 100m, this may not be significant.

Furthermore, this species is ovoviviparous, giving birth to large, well-developed offspring (Irvine *et al.*, 2012). This eliminates the passive dispersal stage attributed to planktonic larvae and eggs, which has been heavily studied in determining patterns of differentiation in species (Hickford and Schiel, 2003). If this species shows any kind of philopatric behaviour, with no passive dispersal, geographically distinct populations may be highly isolated. Additionally, the possible natal philopatry in females is also worth considering as this behaviour has been described in various other elasmobranch species (Dudgeon *et al.*, 2012). However, the medium size of this species may indicate a greater adult dispersal capacity, reducing the extent of geographical differentiation. This possibly enhanced dispersal capacity is reinforced by the seeming partitioning of habitats by developmental stage, which could facilitate gene flow. As with *C. crepidater* (see Chapter 2), if individuals move through a series of spatially distinct geographic regions as they develop, they would by necessity have the capacity to migrate across the distances between these sites. This behaviour would in itself result in geographically separate areas containing individuals at different developmental stages of the same population – effectively homogenising the genetic variation between those regions. The geographic extent of this behaviour could be vast, and has the potential to effectively genetically homogenise large portions of their range through these developmentally determined migrations. Conversely, it is also possible for individuals from different natally philopatric groups to mix in feeding grounds, but separate to different spawning

grounds – allowing for physical mixing of individuals, but with limited gene flow, e.g. as in Atlantic bluefin tuna (*Thunnus thynnus*; Block *et al.*, 2005; Carlsson *et al.*, 2007).

D. calcea is more specialised in diet than other comparable deep sea sharks, with a preference for teleost fishes (Pethybridge *et al.*, 2011), which may limit responses to environmental conditions through changes in prey abundance and distribution. That said, their preference for teleost fish seems quite generalist in itself, with a range of fishes being preyed upon, and their preferred teleost prey, of family Myctophidae, being both diverse and broadly distributed (Catul *et al.*, 2010; Pethybridge *et al.*, 2011), making prey-based restriction and subsequent isolation less likely. Further to this, the preference for mesopelagic fishes may also lessen the association of *D. calcea* to the benthos – resulting in a more pelagic lifestyle. This greater independence from the continental shelf benthic zone may give a better suitability for open-water movement and facilitate longer distance migrations across oceanic basins.

Unfortunately, no prior genetic studies have been carried out on this species. However, there is the possibility that *D. calcea* undertakes obligatory migration with developmental progression, has a greater dissociation with the benthos, and a preference for globally prolific and well distributed prey (Pethybridge *et al.*, 2011). This could give this species an extremely high level of vagility across its global distribution, and overall support the hypothesis (see Chapter 1) that genetic structure will be little influenced by geography. Instead, it would seem more likely that that any patterns of genetic structure are likely to be dictated by differential environmental conditions and be prevalent in areas of the genome under these selective forces, as described by the second hypothesis (see Chapter 1).

Methods

Sampling

Samples were taken from four different sites all by benthic trawl at various depths along the continental shelf. Three sites were representative of the Atlantic; Hebrides North, Hebrides South, and Bay of Biscay; and Chatham Rise representing the Pacific – details of which are

summarised in Table 3.1. Tissue samples of the caught individuals were then preserved in either 20% DMSO saturated with NaCl or in ethanol and sent to Durham University for storage.

Table 3.1: Location and depth of D. calcea *sample sites.*

Sample Name	Latitude	Longitude	Depth (m)	Sample Size
Hebrides North	57.16-58.86	-9.50	525-1032	15
Hebrides South	55.97	-9.41	500-750	17
Bay of Biscay	42.54	-28.97	1150	21
Chatham Rise	-42.47	-179.41	646-979	6

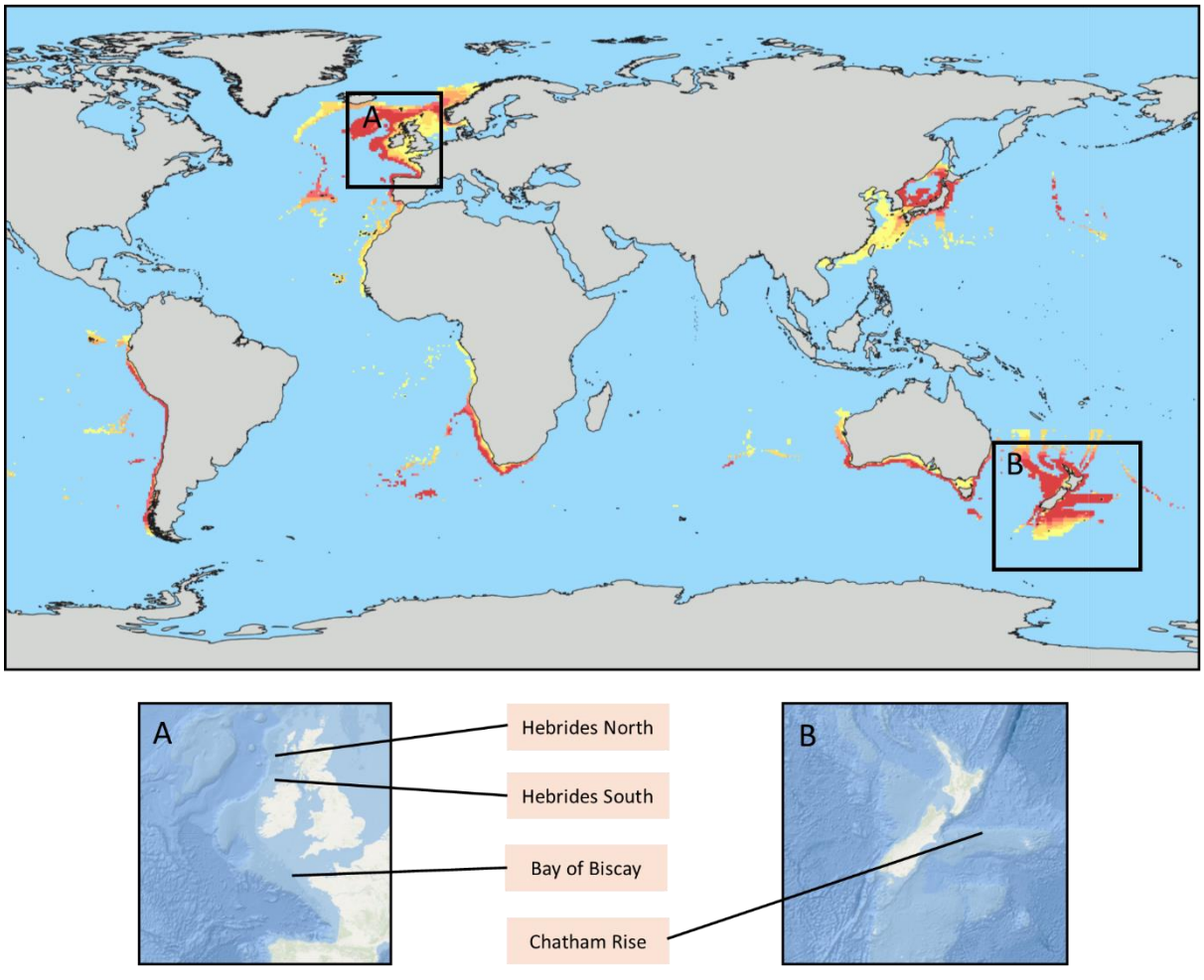

Figure 3.1: Sample sites and known distribution for Deania calcea.

Extraction

Only the samples from the Chatham Rise were extracted by the author, and were done following the same protocols as described in Chapter 2. This includes the pooling of multiple

extractions as the extraction quality was low. The rest of the samples were taken from the lab archive and were of immediately usable quantity and quality. There was some ambiguity in the labelling of four archived samples, whose identity was confirmed using the COI gene. These samples were barcoded using the same primers as described in Chapter 2.

Library Preparation and Sequencing

The details of the library preparation and sequencing are as described in Chapter 2.

Bioinformatic marker discovery

The bioinformatic handling of the raw reads using the Stacks pipeline was as described in Chapter 2, but for one read handling difference. In Chapter 2 only the forward reads were used, whereas for *D. calcea*, both forward and reverse reads were used. Two possible ways were hypothesised to do this. These were referred to as concatenating and pasting.

Concatenating takes the forward read file and appends the reverse reads to the end of the file, effectively doubling the number of lines and therefore loci. In practice, this gives approximately double the number of identified SNPs, but as the forward and reverse reads are being treated as separate loci and are obviously linked, this method was considered inviable for meaningful analyses.

Pasting, as the name suggests, takes the reverse read and pastes it onto the end of the forward read, making one continuous sequence, to be processed as one locus. However, the reverse read must first be modified. Since paired-end sequencing-by-synthesis only produces reads in a 5' to 3' direction, with the paired reads of each fragment representing the complimentary strand of one another. This also means that, in this context, the reads outputted from the Illumina platform begin with the flanking restriction sites and finish towards the centre of the fragment. Furthermore, important to note, is that there is an inherent reduction in read quality towards the end of the read in sequencing-by-synthesis methods (Fuller *et al.*, 2009). Here, it would make more biological sense to reverse the compliment of the reverse read, and then paste this onto the end of the forward read, essentially giving one continuous representation of the initial fragment, albeit with a large deletion

in the centre. However, given that; the end of each read is of lower quality (mitigated against through trimming to 110bp in this case), that there may by multiple putative SNPs present in each read, and that SNPs are discovered by a directional scan of the read from start to finish – retaining only the first SNP (to mitigate against marker linkage); it is advantageous to have the reverse read pasted on in the original restriction-site-first order. The idea is that if multiple SNPs are present in the reverse read, the higher quality ones (sequenced first) will be retained over the poorer ones sequenced later. With this method, it is important to remember to reverse the reverse read's compliment after marker discovery, should there be a need to use the pasted sequence for further applications.

The second approach, pasting, was carried out using an adapted script originally sourced from Claudius Kerth at a genomics workshop in Sheffield. The final filtering steps used in the populations program at the end of the Stacks pipeline removed individuals with more than 80% missing loci, along with any loci that were not present in at least 70% of individuals in each population. Of the 103,842 loci identified with a minimum stack depth of 8, a final 4185 were retained after the loci filtering steps above.

Analyses

The settings used for separating outlier and neutral markers using LOSITAN (Antao *et al.*, 2008) were the same as those described in Chapter 2, as well as the settings used to calculate pairwise F_{ST} comparisons (Bonferroni corrected) and DAPC based analyses. Investigating adaptation also followed the same methodology as was done for *C. crepidater* (Chapter 2) in that LOSITAN was again used, but with a more stringent P value threshold of 1, as well as usage of the program BayeScan (Foll and Gaggiotti, 2008). These putatively adaptive loci were also queried using Blastx, but an additional bioinformatic step was required to separate out the compliments of the reverse reads from the pasted reads used in denovo_map.pl.

Results

Separation of neutral and outlier markers

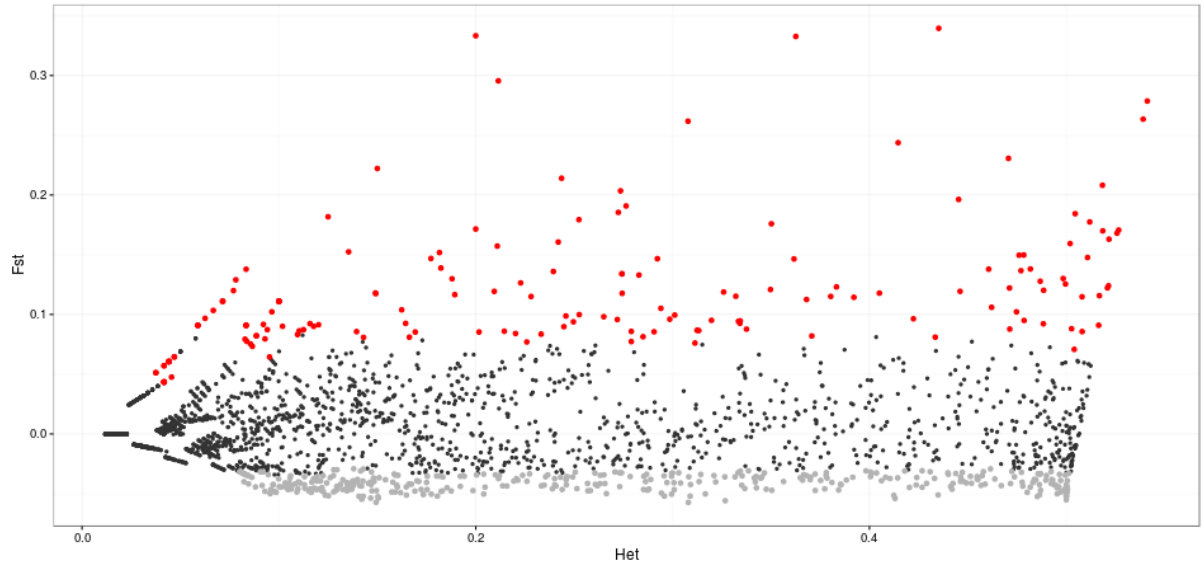

Figure 3.2: Identification of markers potentially under selection by plotting F_{ST} *against diversity (heterozygosity) using LOSITAN. Red markers are putatively under positive selection, black are neutral, and grey under balancing selective forces.*

From a total of 4185 markers, 3773 fit the expected neutral distribution and were assigned to the "neutral" marker group to be analysed as a single dataset. 172 markers were high outliers of the expected neutral distribution, and therefore putatively under positive selective forces. These were assigned to the "outlier" group. The remaining 238 were low outliers and potentially under balancing selection and not considered for analysis (Figure 3.2).

Pairwise F_{ST} Comparisons

Table 3.1: Pairwise F_{ST} *Comparisons. Below the greyed diagonal divide are values calculated from the neutral set of markers, whilst those above are calculated from the outlier set. Yellow and red indicate P<0.05 and P<0.001, respectively.*

	Hebrides North	Hebrides South	Bay of Biscay	Chatham Rise
Hebrides North		0.0407	0.0152	0.2492
Hebrides South	0.0071		0.0024	0.1937
Bay of Biscay	-0.0006	-0.0011		0.1864
Chatham Rise	0.0002	-0.0136	-0.0030	

Neutral loci here show complete homogeneity across all four sample sites. The outlier loci, on the other hand, show highly significant differentiation between ocean basins. Also significant in the outlier loci is the comparison between the two Hebrides site, with no significant difference between the Hebrides and the Bay of Biscay (Table 3.2).

Discriminant Analysis of Principle Components

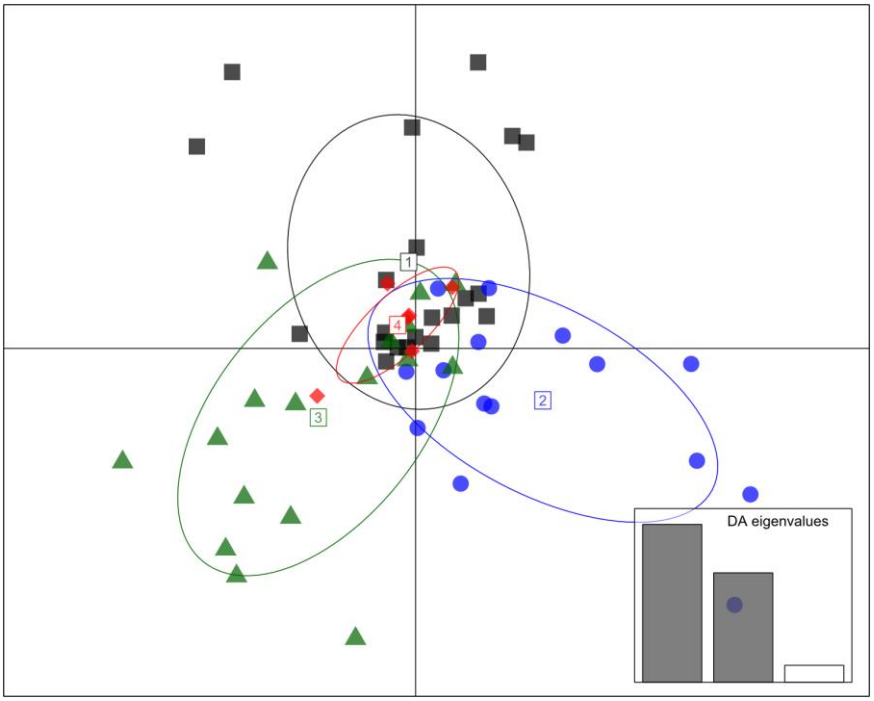

Figure 3.3: DAPC scatterplot of neutral markers of the two greatest discriminant functions. Black squares (1) represent Bay of Biscay, blue circles (2) Hebrides North, green triangles (3) Hebrides South, and red diamonds (4) Chatham Rise. Eigenvalues are 14.9, 10.3, and 1.6, respectively.

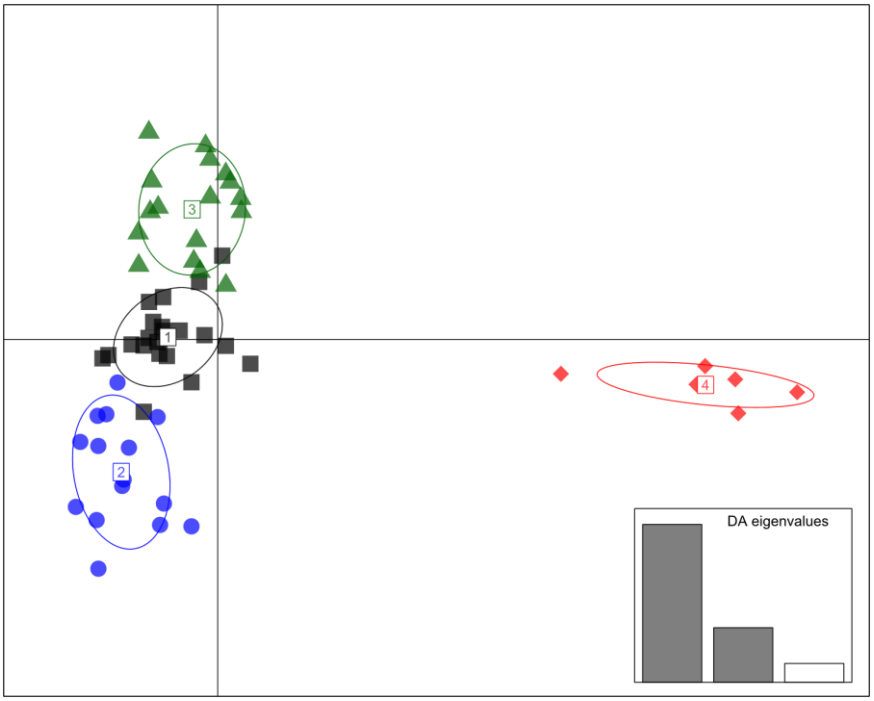

Figure 3.4: DAPC plot of outlier markers of the two greatest discriminant functions. Black squares (1) represent Bay of Biscay, blue circles (2) Hebrides North, green triangles (3) Hebrides South, and red diamonds (4) Chatham Rise. Eigenvalues are 310.0, 107.3, and 37.06, respectively.

The neutral DAPC scatterplot (Figure 3.3) shows an essentially homogeneous single cluster of all the sample sites, with the individuals representing Chatham Rise nested in the centre. For outlier markers, the DAPC scatterplot (Figure 3.4) shows distinct clusters for each sample site. The greatest difference is across the x-axis and shows a clear differentiation between Chatham Rise and the Atlantic clusters. Within the Atlantic, each sample site is clearly differentiated, with the Bay of Biscay cluster being effectively flanked by each Hebrides site. There is limited overlap between the Bay of Biscay and each Hebrides site.

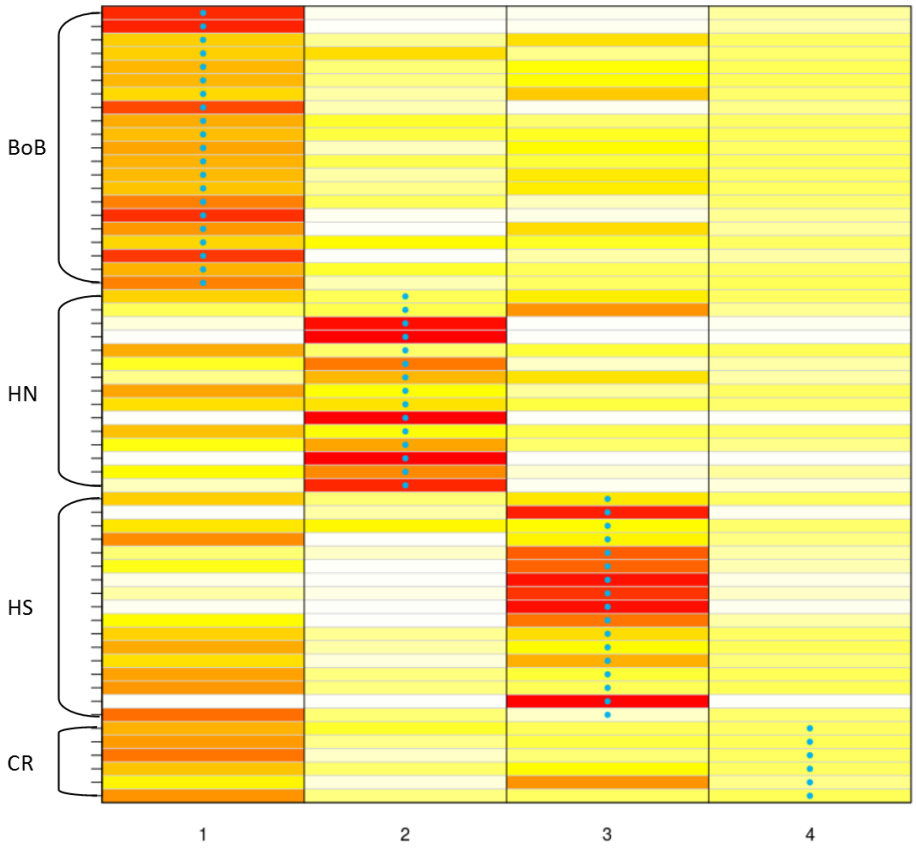

Figure 3.5: DAPC assignment plot for neutral markers. Blue dots represent prior group assignment based on sample site, whilst the heated cells represent group membership probabilities of each individual based on retained discriminant functions, with white representing 0 and red representing 1. HN = Hebrides North, HS = Hebrides South, BoB = Bay of Biscay, CR = Chatham Rise.

The assignment plot (Figure 3.5) for the neutral markers reflect the patterning shown in the DAPC scatter plot (Figure 3.3). The Chatham Rise individuals all show a higher assignment to other groups, with all but one showing the greatest assignment to the Bay of Biscay. This assignment is of comparable strength to the self-assignment of some individuals sampled from the Bay of Biscay. The Hebridean groups show the highest levels of self-assignment, but these can be sub-divided into individuals with high levels of self-assignment and those that show equal likelihood assignment to other groups, mostly the Bay of Biscay. There is limited cross-assignment between Hebridean sites.

The outlier loci (Figure 3.6) show strong self-assignment for all groupings, with limited cross-assignment between the Bay of Biscay and each Hebridean sample. There is no cross-assignment between Hebridean sites. The Pacific sample site also remains entirely distinct from all other sample sites.

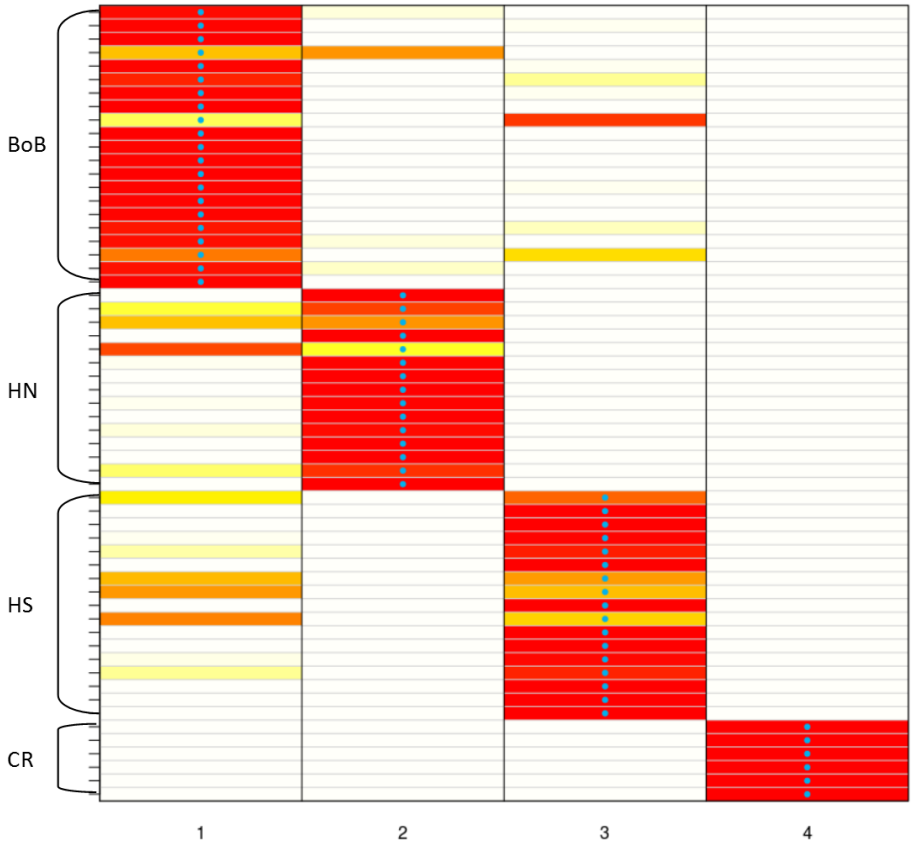

Figure 3.6: DAPC assignment plot for outlier markers. Blue dots represent prior group assignment based on sample site, whilst the heated cells represent group membership probabilities of each individual based on retained discriminant functions, with white representing 0 and red representing 1. HN = Hebrides North, HS = Hebrides South, BoB = Bay of Biscay, CR = Chatham Rise.

Outlier function

Two methods were used to identify strong outliers that are likely under selective forces. The first method, using BayeScan, identified no loci. However, the approach using LOSITAN with a high P value threshold resulted in 17 loci for investigation. Each locus consisted of the pasted version of the forward and reverse reads, which were separated and individually queried using Blastx (Agarwala *et al.*, 2016). Of those 17 queried, 4 loci queried no significant similarity was found, 12 loci very poorly matched a variety of proteins and were discarded. The final locus whose query scored a result, did so well and with both forward and reverse reads to the same predicted

protein for West Indian Ocean Coelacanth (*Latimeria chalumnae*). As of yet, the protein is uncharacterised and is labelled as "LOC106705466". The forward also sufficiently matched with a predicted "RUN domain-containing protein 3b" in the Chinese alligater (*Alligator sinensis*), and the reverse to "Hemicentin-1" in the Atlantic herring (*Clupea harengus*).

The SNP is present in the forward read and, assuming the same reading frame as in LOC106705466, is responsible for an arginine to leucine amino acid substitution, with the R-group of arginine being positively charged and that of leucine being non-polar. In this dataset, the arginine variant was most common, with the leucine encoding allele being present in three individuals from the Bay of Biscay sample site, all heterozygous at this locus.

Discussion

Neutral structure

The overall signal in the neutral loci was that of homogeneity, with no significant differentiation between populations found in any of the analyses. This suggests extremely high levels of gene flow, even between ocean basins. Whilst high levels of gene flow within basins has been observed in closely related species (Straube *et al.*, 2011; Veríssimo *et al.*, 2011; Catarino *et al.*, 2015; Gubili *et al.*, 2016), the study by Cunha *et al.* (2012) into *C. crepidater* found significant putatively neutral differentiation between the Atlantic and Pacific. Given the sheer distance between these sites, the level of gene flow indicated here is surprising and must be facilitated by substantially unrestricted migration across the range, supporting the notion that this species is indeed highly vagile, even between basins. This also provides an instance of a species with small litters of well-developed pups instead of a dispersive planktonic stage not showing limited gene flow. This connectivity also supports the observations by previous researchers in finding different demographic, ontogenetic, and reproductive groups at different sample sites, and their speculation that this may be due to large migratory movements of individuals to different areas at different life and reproductive stages (Clark and King, 1989; Machado and Figueiredo, 2000; Wetherbee, 2000; Clarke *et al.* 2002).

These results also provide evidence that despite this species having a more restricted diet than close relatives (Pethybridge *et al.*, 2011), they appear not to be restricted by prey specialisation, either indicating that prey abundance has little impact on patterns of connectivity, or that the specialisation on globally abundant taxa (Catul *et al.*, 2010) negates this prey abundance dependence.

Outlier

Again, it is important to note here, as in the discussion of Chapter 2, that caution must be taken in interpreting these outlier loci, as whilst a marker may have high outlying F_{ST} values that are indicative of loci under selective forces, there is no guarantee that is the case. As through stochastic processes, some loci may experience greater levels of drift to result in high values of F_{ST} despite not being acted upon by selection. If this were the case, patterns of structure in the neutral and outlier sets would be highly similar, but with these being exaggerated in the outlier set.

However, the outlier markers show a different pattern than that seen for neutral markers, with a clear divide between the Atlantic and Pacific basins. Within the Atlantic, the only significant pairwise comparison (Table 3.2) was between the two Hebridean sites. These observations are reinforced by the DAPC scatterplot (Figure 3.4), with the largest distance between groups being that between the Atlantic sample sites and the Chatham Rise. More interestingly, is that the three Atlantic sites themselves form three partially overlapping clusters in a rough continuum, with the Bay of Biscay site flanked by the two Hebrides sites. The assignment plot (Figure 3.6) further shows similar information, with Chatham Rise being completely self-assigning and distinct from all other sample sites, and with the majority of Atlantic individuals also being assigned to their prior group designation. The Atlantic individuals that only partially self-assigned showed assignment from the Bay of Biscay to one of the Hebrides sites, or from one of the Hebrides sites to the Bay of Biscay. Individuals that were reassigned in the Atlantic showed a likelihood of group assignment to only two sites, and no individuals from the Hebrides showed any likelihood of assignment to the opposing Hebridean group. The cause of this differentiation between Hebridean sites is unclear, and so far unique in the literature. One exception is neutral differentiation found along the Porcupine Ridge in orange roughy (*Hoplostethus atlanticus*), attributed to the differential

usage of discrete spawning areas (Carlsson *et al.*, 2011). However, a similar dynamic is likely not present in *D. calcea*, as signal for this differentiation would also be present in the neutral loci.

Between ocean basins, despite high levels of apparent gene flow, the outlier loci showed a very clear pattern of differentiation. This is not surprising given the putative differences in environmental conditions between them (Reid, 1961; Rex and Etter, 2010), and likely represents loci that are under differential selective forces.

Overview

The most prominent signal in this dataset is the marked difference in patterns of differentiation in the neutral and outlier loci, with effective panmixia across sample sites in the neutral loci and a clear pattern of differentiation between all sample sites in the outlier loci. This may imply significant differences in selective pressures between sites connected by high levels of apparent gene flow. Interestingly, the pattern of differentiation shown in the outlier markers (Figure 3.4) is what would be expected for neutral loci under a hierarchical islands model, or even a hierarchical stepping-stone model for neutral differentiation (Jombart *et al.*, 2010). This could be interpreted as a hierarchical environmental islands, or environmental stepping stone pattern, determined by environmental distance rather than levels of gene flow.

Most interesting, was the significant differentiation in outlier loci between the Hebrides, and the intermediate position of the Bay of Biscay**.** One interpretation that may fit these data is that Hebrides North and Hebrides South again offer two distinct suites of selective pressures over a fine-scale, causing differential adaptation between these sites regardless of histories of gene flow between groups. The Bay of Biscay then, marginally overlapping with both Hebridean sites, could sit in an adaptive middle-ground, whereby some selective pressures are shared with one group, whilst other selective pressures are shared with the other. Plausibility for this hypothesis is given when considering the environmental heterogeneity of the Bay of Biscay (Koutsikopoulos and Le Cann, 1996), which may support a variety of adaptive traits in the individuals sampled from the area that partially overlap with both Hebrides sites. Fine-scale population structure has also been found using neutral loci in orange roughy (*Hoplostethus atlanticus*) along the geographically proximal Porcupine Ridge, attributed to discrete spawning area usage (Carlsson *et al.*, 2011). It is

possible that the heterogeneous environment resulting from the complex topography of the Rockall Trough, along with the intersection of the North Atlantic Current and the Slope Current (Neat *et al.*, 2008), is enough to create sufficiently different environmental conditions across a fine spatial scale, to which different groups must adapt. Overall, these results and their interpretation here suggest that environmental differences and behaviours across varying spatial scales are much greater drivers of differentiation in this species than limits to dispersal and geographic barriers to gene flow.

Ultimately, the nature of these selective pressures and how they differ are unclear, and from these data, the only indication to functional difference is in the single marker that successfully aligned with the LOC106705466 protein. Whilst this protein is undefined and whose function is as-of-yet unknown, the alternative match, and perhaps similar gene, Hemicentin-1, is better studied. The hemicentin protein family is closely related to the fibulin protein family and are thought act as an extracellular adhesive agent and be widely conserved in vertebrates (Xu *et al.*, 2013). This locus has also been linked to macular degeneration in humans (Schultz *et al.*, 2003; Seitsonen *et al.*, 2006), suggesting an integral role in organism survival. However, even if LOC106705466 shares a similar function, it remains unclear as to how variation in this protein would affect the overall adaptive traits of an individual, and would likely be only a part of a group of interacting genes.

Future studies

Whilst the first step in determining levels of gene flow across this species' range has been established, it is evident that this species shows cryptic patterns of genetic structure and adaptation, especially within the Hebrides region. It is also evident that these patterns of structure are unlikely to result from restrictions to dispersal capacity as there seems to be high levels of gene flow even between ocean basins. Structure is then likely due to differing environmental conditions between sites, and perhaps within-species behavioural differences that are driving differentiation, even between closely spaced sites. This implies that geographical spatial distribution is not the only significant variable determining genetic structure in this species, and future study should reflect that. Investigating the life-long behaviours of *D. calcea* should be a promising start, as the species is clearly highly vagile, and previous observations have implied complex patterns of migration and

spatial usage in different developmental and demographic groups (Clark and King, 1989; Machado and Figueiredo, 2000; Wetherbee, 2000; Clarke *et al.* 2002). In this effort, if a tagging regime could be implemented to follow these life-long migratory behaviours of individuals, it would give much more insight into how different spatial areas are used and an idea of what environmental factors might limit movement, restricting groups into their well-defined adaptive groupings. Another variable that may have added to the complexity of the patterns uncovered in this study may be the depth gradient, which was uncontrolled in the sampling scheme utilised here, and which provides a gradient of environmental change across a relatively short distance (Levin *et al.*, 2001; Rex and Etter, 2010). Finally, the fine-scale outlier structuring between the Hebrides sites demonstrates that genetic structuring can occur at a very fine geographic resolution, despite a high capacity for long-distance gene flow. This should be accounted for in further investigation in the use of fine-scale sampling, as Hebridean-like divides may be present in different instances throughout *D. calcea*'s vast range.

Conclusion

Overall, these data show effective panmixia across the four representative sample sites in the neutral loci, but a clear pattern of differentiation in the outlier loci. This is likely a due to differential environmental pressures, or behaviours driving differentiation between these groups, despite high levels of gene flow. Especially remarkable is the differentiation in the outlier loci between the two Hebridean sites which are only separated geographically by roughly 100km. The causation of this pattern of differentiation remains unknown.

Chapter 4: Synthesis

This study has uncovered broad patterns of genetic structure and possible signals of selection in two closely-related deep sea elasmobranch species, *Centroselachus crepidater* and *Deania calcea*. The purpose of this final chapter is to synthesise the results of each species presented in chapters 2 and 3, address the hypotheses put forward in chapter 1, and to place these into the context of the current state of the field.

Neutral Patterns of Differentiation

The first hypothesis stated in chapter 1 was that these two species would show effective panmixia across their global distribution. For *D. calcea*, between the sites sampled here, this appears to be true. No pairwise population comparisons showed any significant differentiation between sites (Table 3.2), and little to no discernible pattern could be found in the ordination analysis (Figures 3.3 and 3.5). A similar result was found for *C. crepidater*, with no significant pairwise F_{ST} comparisons being found between sample sites (Table 2.2). However, the latter result was not as clear cut, and differentiation between the Atlantic groups and the Pacific was apparent in the DAPC analysis (Figures 2.3 and 2.5), possibly found to be insignificant by the F_{ST} methodology through the poor representation of the Pacific site and one individual that showed a high likelihood assignment to the Atlantic. The occurrence of a false negative result in the pairwise comparisons is supported the previous work of Cunha *et al.* (2012) who found significant inter-ocean differentiation in *C. crepidater* using microsatellite markers. The barrier hypothesised by the authors was the indirect effects of the Agulhas Current Retroflection off the south coast of Africa on prey movements. The comparison of the two species in this study can be used to further this hypothesis, as both sharks are thought to have largely overlapping, but differing preference in prey species (Pethybridge *et al.*, 2011). Notable is that *C. crepidater* is thought to be a more generalist predator than *D. calcea*, whose diet is restricted to a subset of that of *C. crepidater* (Pethybridge *et al.*, 2011). Given this more generalist strategy, movement and dispersal of *C. crepidater* could be thought to be less limited to the distributions of prey species than *D. calcea*, but the results

presented here clearly indicate much greater levels of gene flow in *D. calcea* than *C. crepidater*. This would imply that the indirect impacts of prey movements may not be the main drivers of the apparent inter-basin differentiation found in *C. crepidater*. However, *D. calcea* was found to have a greater preference for mesopelagic teleost prey species (especially myctophids), which is consistent across their global range (Blaber and Bulman, 1987; Ebert *et al.*, 1991; Yano, 1991; Saldanha *et al.*, 1995; Pethybridge *et al.*, 2011), whereas geographic variation in diet in *C. crepidater* has been reported (Mauchline and Gordon, 1983; Dunn *et al.*, 2010, 2013). The result may be that geographically determined prey preference in *C. crepidater* limits the dispersal of individuals of this species, whilst the consistent preference of *D. calcea* for broadly distributed prey overcomes this limitation. Regardless, the two species share similar life history traits and presumably similar dispersal capacities, and with no direct investigation into migratory behaviours of either species, the cause of these different inter-basin patterns of differentiation remains cryptic. The only other interocean comparison of a closely related species currently available is of the Portuguese dogfish (*Centrscymnus coelolepsis*), which through 11 putatively neutral microsatellite markers shows no significant differentiation between sample sites in the North Atlantic and sites of the south coasts of South Africa and Australia (Catarino *et al.*, 2015), matching the pattern seen in *D. calcea* in this study.

Within basins, however, both species showed comparable levels of homogeneity, and add to a limited, but growing body of similar within-basin patterns found in other closely related deep sea elasmobranchs. Homogeneity in the North Atlantic was found in the Portuguese dogfish (*Centroscymnus coelolepsis*; Veríssimo *et al.*, 2011; Catarino *et al.*, 2015) and velvet belly lanternshark (*Etmopterus spinax*; Gubili *et al.*, 2016). Within the Pacific, however, *E. spinax* showed weak, but significant, differentiation between the East and West extents of the basin, attributed to a recent cessation to gene flow, or an isolation by distance effect (Straube *et al.*, 2011). Information regarding levels of gene flow between basins now remains limited to *C. coelolepsis*, *D. calcea*, and *C. crepidater*, with only the latter showing any signal for significant differentiation between basins (Cunha *et al.*, 2012; Catarino *et al.*, 2015). Overall, despite limited interspecies variation, the results obtained here combined with information currently available in the literature all indicate a pattern of very little neutral differentiation across the global range of these species,

suggesting extremely high levels of gene flow across considerable geographic distance. The only known exception to this apparently shared pattern of differentiation is the break of homogeneity between individuals in the Mediterranean and the rest of the global range in *C. coelolepsis* and *E. spinax* (Catarino *et al.*, 2015; Gubili *et al.*, 2016) – a barrier shared by many other marine taxa (Patarnello *et al.*, 2007). However, this apparent consensus of little differentiation in deep sea sharks remains inconclusive, as they are still grossly underrepresented in the literature and the much greater wealth of knowledge of their shallow-water counterparts indicates large variation in patterns of genetic differentiation between species (Dudgeon *et al.*, 2012).

Outlier Patterns of Differentiation

The second hypothesis from chapter 1 stated that patterns of differentiation in outlier loci, i.e. those potentially under selective forces, will differ from any patterns found in neutral loci. The basis of this hypothesis is that patterns of differentiation across a geographical scale will not be determined by dispersal capacity, nor barriers to gene flow, but by differential environmental pressures to which individuals must adapt.

This appears to be true for both species. In *C. crepidater* the partial differentiation between the Pacific site and the Atlantic is much more clearly defined in the outlier loci. Whilst only the population F_{ST} comparisons between the Hebrides and the Pacific were statistically significant (Table 2.2), the discriminant analysis of principle components (DAPC) shows both the clear out-grouping of the Pacific as well as outlier-unique differentiation between the Hebrides and the Mid Atlantic Ridge (Figures 2.4 and 2.6). The comparison between the Mid Atlantic Ridge and the Pacific shows an elevated F_{ST} value (F_{ST}=0.46, p=0.05). Whilst the p-value is at the border of the arbitrary significance cut-off commonly used, the population pairwise comparison itself is extremely high and may indicate population structure between the two samples. However, this high p-value may be due to the inherent overestimation bias that can occur when using this estimator with low sample sizes for both sites in ddRADseq data (n=5; Willing *et al.*, 2012; Gautier *et al.*, 2013; Andrews *et al.*, 2016). Regardless, the clear differentiation between the Mid Atlantic Ridge and the Hebrides in the DAPC shows differentiation despite high levels of gene flow, suggesting that differential selective pressures are acting on the individuals at each site. Between basins the

enhanced signal for differentiation may be also due to selective differences between basins, but as the pattern is similar to that found in the neutral loci, it is also possible that this signal is merely the patterning of loci experiencing greater levels of drift.

D. calcea, on the other hand, shows a much more dramatic difference in patterns of differentiation in neutral and outlier loci. Where in the neutral loci effective panmixia was found, the outlier loci uncovered strong signals of differentiation. The greatest of which was between the Atlantic and the Pacific sites which gave consistently high and significant pairwise F_{ST} comparisons and categorical self-assignment of Pacific individuals to their outgroup. More interesting was the structuring of the Atlantic sites, which shows significant differentiation between the two Hebrides sites, but overlap of each Hebrides site with the Bay of Biscay. This suggests non-neutral selective forces driving differentiation between these sites despite extremely high levels of homogeneity across the global species distribution.

The two species share all sample sites but the Mid Atlantic Ridge and the Bay of Biscay. The two sharks share similar levels of outlier differentiation between the Atlantic and Pacific, which is perhaps not surprising considering the environmental differences between the two ocean basins (Reid, 1961; Rex and Etter, 2010). Unexpected was the fine-scale comparison between the two Hebrides sites found in *D. calcea* which is entirely absent in *C. crepidater*, despite almost identical sampling sites. Given the lack of differentiation between these sites in *C. crepidater*, and the similar morphology and life history traits of the two species, unless as-of-yet unknown differences in habitat utilisation are responsible, it is unlikely for the differentiation seen in *D calcea* between the Hebrides sites to be caused by abiotic factors, as they would likely be shared between the two species. This directly implies that the cause of the differentiation in *D. calcea* is biotic. The data here is insufficient to resolve the type of biotic factor responsible, but likely candidates include differential prey preferences between species, parasite pressures, and predation. Indirect environmental influences may also impact the shark species differently through abundance differences in interacting species.

Both shark species also showed indirect evidence of differential spatial usage for different ontogenetic stages (Clark and King, 1989; Machado and Figueiredo, 2000; Wetherbee, 2000; Clarke *et al.*, 2002; Daley *et al.*, 2002; Neat *et al.*, 2008; Moore *et al.*, 2013). If each species uses

this same area for different developmental stages, then selection pressures on each one may be significantly different, and drive differentiation across this spatial scale differently between them. This could be through prey preference, which is known to be different by developmental stage in at least *C. crepidater* (Pethybridge *et al.*, 2011), or parasite load, which has also been shown to differ with ontogenetic stage in the lesser spotted dogfish (*Scyliorhinus canicula*; Silva *et al.*, 2017). Also likely are differences in environmental tolerances between the species, with one being more stenotypic than the other, resulting in differing molecular responses to the same environment. Unfortunately, outlier loci of this kind have not been investigated in other deep sea sharks and further study is required to resolve this difference in fine-scale differentiation.

Conservation

The life history traits of these two species make them vulnerable to exploitation, but the high levels of gene flow and the underlying migration may make these species more resilient to local depletion or extinction events. That said, the likely usage of a variety of geographic regions through the life cycle that may facilitate homogeneity across large spatial scales would have the opposite effect. Local exploitation would continuously remove key ontogenetic groups and break the reproductive cycle, potentially causing the collapse of wide-spread, migratory populations. So whilst the implementation of marine reserves may be an effective conservation tool for some shallow-water species of shark (Baum *et al.*, 2003; Bond *et al.*, 2012; Daley *et al.*, 2014), their effective implementation may be more difficult as local protection may only protect one demographic group of migratory deep sea species.

That said, this apparent high level of migration has not prevented structuring in outlier loci, revealing likely differential signals of selection. Despite high levels of gene flow between sites, there is between-site variation that should be conserved. These possibilities would make the aforementioned implementation of marine reserves a complex task, as it would be likely that a set of reserves would have to be chosen that would maximise the protection of vital life stages, as well as hotspots of variation. This is made yet more difficult if these adaptively valuable groups are highly mobile in themselves, and perhaps the marine reserve approach would not be appropriate for the conservation of these species. Regardless, more information on species-specific behaviour

needs to be collected before effective management can be carried out. The fine-scale outlier structure in *D. calcea* shows that complex patterns of differentiation are likely across other parts of its range, which may also be true for similar sharks. This highlights the need for greater sampling effort to identify valuable reservoirs of adaptive variation in these widely-dispersed elasmobranchs.

More immediately applicable than marine reserves would be restrictions of fisheries impacting these species. Fortunately, this is already being considered by entities such as the European Union, who have proposed a no-catch policy on all deep sea sharks for the next two year term, 2017-2018 (European Union, 2016). The results of this study should be used to support policies such as these and to encourage other political entities to follow suit.

Future Studies

The fine-scale structure found in the Hebrides in *D. calcea* highlights the need for more extensive sampling strategies to be utilised in future genetic studies such as this. This should include investigation into possible structure across the environmental depth gradient, given the high likelihood that environmental conditions play a large role in determining structure in these species and the known patterns associated with depth in other taxa (Rex and Etter, 2010). A greater geographic sampling resolution would also allow for the representation of intermediate individuals that would show isolation by distance effects and resolve the nature of differentiation between the Mid Atlantic Ridge and the Hebrides, as well as the nature of the relationship between the Bay of Biscay and the Hebrides. Future genomic studies of this type would also greatly benefit from the availability of a reference genome of at least one of these deep sea squaloid sharks, allowing greater control over marker discovery, gene identification, and linkage disequilibrium.

The most effective route of study, however, would be to elucidate the specific behaviours of these, and closely-related, deep sea sharks. The data shown here indicates that the two species, despite their similarities, utilise different regions differently and display different levels of gene flow across geographic distance. The patterns of migration and reproductive behaviour are likely factors that determine these differences. Understanding the range of these behaviours in these fish would contribute significantly to the understanding of how these species have evolved in this understudied environment, and would prove informative in conservation and fishery management.

The most appropriate approach would be the use of tagging techniques to follow the movement of individuals, collecting data on diel patterns of movement, and movement throughout individual life-cycles. However, further development of tagging methodologies is needed in order to successfully carry this out in these deep sea species (Daley *et al.*, 2014).

Next Generation Sequencing Data and Population Genomics

The choice of methodology for this study is based on the reported greater ease of *de novo* marker discovery (Davey *et al.*, 2011) and apparent robustness in uncovering patterns of population differentiation (Cariou *et al.*, 2013; Jeffries *et al.*, 2016). However, despite the advantage of the use of a couple of orders of magnitude greater numbers of markers to use to estimate levels of differentiation, there are valid concerns over the application of population genetics methods to these data. The root of a large aspect of this are the higher levels of sequencing error inherent in NGS technologies compared to precursory Sanger sequencing. In this case, Illumina platform error rates are around 1% (Nielsen *et al.*, 2011; Andrews and Luikart, 2014), compared to down to 0.001% in Sanger sequencing (Shendure and Ji, 2008). This increased error rate calls into question the accuracy of genotype and SNP calling, and the downstream inferences based on these data (Nielsen *et al.*, 2011). To mitigate against this, sequencing depth and required coverage for bioinformatic discovery of markers are important in determining between sequencing error and real variation in the samples. This depends on each dataset. Whilst a higher bioinformatic locus coverage requirement for SNP identification increases the confidence in that locus, as this bioinformatic cut off approaches the actual sequencing coverage variation will be lost due to allelic drop out (Mastretta-Yanes *et al.*, 2015; Huateng and Knowles, 2016). This problem of allelic drop out is thought to be exacerbated in RADseq datasets if polymorphisms occur in the restriction site itself, resulting in sequencing failure of that allele (Gautier *et al.*, 2013). Downstream, this can result in overestimation of F_{ST} and underestimation of genomic diversity (Andrews *et al.*, 2016). Further mitigation is carried out with efforts to remove poor quality reads entirely (Nielsen *et al.*, 2011), as well as trimming of the poor-quality ends of reads, as done in the Stacks pipeline here (Catchen *et al.*, 2011).

However, despite these drawbacks, empirical studies based on simulated genotypes consistently show a surprising level of accuracy in determining levels of population differentiation (Willing *et al.*, 2012, Cariou *et al.*, 2013) given careful consideration to bioinformatic filtering steps (Davey *et al.*, 2013; Huateng and Knowles, 2016). Furthermore, RADseq methods have been successful in uncovering patterns of population differentiation, with greater efficacy, that validate previous estimates using established population genetic markers, such as microsatellites and mitochondrial sequences (Hohenlohe *et al.*, 2013; Cruaud *et al.*, 2014; Bradbury *et al.*, 2015; Andrews *et al.*, 2016; Jeffries *et al.*, 2016). The validation of the interocean differentiation found by Cunha *et al.* (2012) in *C. crepidater* in this study contributes to this trend, despite substantially fewer RADseq sampled individuals (160 individuals at 7 microsatellites; 41 individuals at 2209 SNPs). Similarly, Jeffries *et al.* (2016) report resolving finer-scale population structure in RADseq derived SNPs using only 17.6% of samples compared to microsatellites (848 individuals at 13 microsatellites; 149 individuals at 13,189 SNPs).

Overall, RADseq is a relatively new tool for elucidating patterns of population differentiation (Baird *et al.*, 2008) for which methods for identifying and overcoming inherent biases and errors are still being actively developed (Andrews *et al.*, 2016). As such, inferences based on analyses of these datasets deserve due scepticism. However, so far, the ability to discover and use thousands of genomic markers are establishing this methodology as a powerful way of sampling the genome for population genetic purposes (Andrews *et al.*, 2016). Out of this is a tentative assertion that the greater information contained within these many genomic markers allows for accurate estimation of population differentiation using smaller sample sizes (Willing *et al.*, 2012), *i.e.* the larger number of markers compensates for a smaller size. This applies here, as whilst the sample sizes for several of the populations in this study are small, perhaps leading to small inherent bias, the information derived from this dataset is in part validated against more accepted methodology (Cunha *et al.*, 2012), and arguably remains informative in elucidating population differentiation in these species.

Bioinformatic Improvements

New genetic technologies have allowed the production of unprecedented amounts of data (Van Dijk *et al.*, 2014). However, now that these large datasets are being produced, the challenge still remains as to how to efficiently process them under the duress of both time- and cost-constraints. More efficient software is being written, such as fastSTRUCTURE (Raj *et al.*, 2014), but commonly used coalescent-based programs such as IMa (Hey, 2007, 2010; Sethuraman and Hey, 2016) and Migrate (Beerli and Felsenstein, 1999, 2001) still struggle to process large numbers of low-information markers, such as the SNPs generated here using a RADseq protocol, as opposed to widely used microsatellites and mitochondrial loci. Running times have also been partly alleviated through parallelisation of these software and the running of these parallelised versions on High Performance Computing Clusters, which allow for a program that originally processed the dataset through a single thread, to process the same dataset essentially *n* times faster, with *n* being the number of extra threads, be that parallel threads through newer parallel processors, or multiple processors running in parallel, or both. This additional computational power runs pre-existing programs more efficiently and allows use of the full capacity of more powerful computing architectures.

Conclusion

In conclusion, this study has shed light on the overall patterns of population structure and adaptation in two deep sea shark species. It shows that whilst these species do not show strong limitations to gene flow across their global range, differentiation is still present. Whilst the exact causal factors of these patterns of structure and adaptation remain unknown, it is likely that environmental differences and behavioural strategies are responsible. It also indicates that further study of environmental differences in the deep sea and potentially complex behavioural patterns is required if we are to fully understand the drivers of differentiation, and conserve biodiversity, in this expansive environment.

Appendix A: Sample Data

Table A.1: Sample data for Centroselachus crepidater. *Samples were collated from different sources with varying levels of associated data. For sex, "1" and "2" denote the sexes; it is unknown as to which is male and which is female.*

ID	Collection location			Depth (m)	Date Extracted	Length (cm)	Weight (g)	Sex
	Population	Latitude	Longitude					
23 C1	Hebrides	56.7658	-9.3482	1500	09/02/2007			
23 C7	Hebrides	57.1665	-9.3955	1032	09/02/2007			
23 D3	Hebrides	57.1665	-9.3955	1032	09/02/2007			
23 D6	Hebrides	57.1665	-9.3955	1032	09/02/2007			
23 D8	Hebrides	57.1665	-9.3955	1032	09/02/2007			
23 D9	Hebrides	57.1665	-9.3955	1032	09/02/2007			
23 F10	Hebrides	57.1665	-9.3955	1032	12/02/2007			
23 F4	Hebrides	57.1665	-9.3955	1032	12/02/2007			
23 F5	Hebrides	57.1665	-9.3955	1032	12/02/2007			
23 F7	Hebrides	57.1665	-9.3955	1032	12/02/2007			
23 G10	Hebrides	57.1665	-9.3955	1032	12/02/2007			
23 G3	Hebrides	57.1665	-9.3955	1032	12/02/2007			
23 G4	Hebrides	57.1665	-9.3955	1032	12/02/2007			
23 G6	Hebrides	57.1665	-9.3955	1032	12/02/2007			
28 A10	Hebrides	55.9712	-9.4138	1000	25/06/2007	52	614	1
28 A5	Hebrides	55.9712	-9.4138	1000	25/06/2007	63	1156	1
28 A6	Hebrides	55.9712	-9.4138	1000	25/06/2007	79	2356	2
28 A7	Hebrides	55.9712	-9.4138	1000	25/06/2007	62	1056	1
28 A9	Hebrides	55.9712	-9.4138	1000	25/06/2007	54	706	2
28 B1	Hebrides	55.9712	-9.4138	1000	25/06/2007	46	396	2
28 B2	Hebrides	55.9712	-9.4138	1000	25/06/2007	52	590	2
28 B3	Hebrides	55.9712	-9.4138	1000	25/06/2007	59	1064	1
28 B5	Hebrides	55.9712	-9.4138	1000	25/06/2007	62	1048	1
28 B7	Hebrides	55.9712	-9.4138	1000	25/06/2007	64	1170	1
28 B9	Hebrides	55.9712	-9.4138	1000	25/06/2007	62	1060	1
28 C2	Hebrides	55.9712	-9.4138	1000	25/06/2007	53	696	1
28 C3	Hebrides	55.9712	-9.4138	1000	25/06/2007	48	550	1
28 C5	Hebrides	55.9712	-9.4138	1000	25/06/2007	61	1048	1
28 C6	Hebrides	55.9712	-9.4138	1000	25/06/2007	82	2862	2
28 C8	Hebrides	55.9712	-9.4138	1000	25/06/2007	82	3428	2
28 C9	Hebrides	55.9712	-9.4138	1000	25/06/2007	26	74	1
67	Mid Atlantic Ridge	42.4973	-28.9908	1100	06/11/2015			
68	Mid Atlantic Ridge	42.4317	-29.1220	883	06/11/2015			
69	Mid Atlantic Ridge	42.4317	-29.1220	883	06/11/2015			
74	Mid Atlantic Ridge	42.4317	-29.1220	883	06/11/2015			
78	Mid Atlantic Ridge	42.4317	-29.1220	883	06/11/2015			
TAN1212	New Zealand	-42.4737	-179.4127	977	19/10/2012	58.6	830	F
TAN1217	New Zealand	-42.4237	-178.5609	1035	23/10/2012	52.2	520	F
TAN1242	New Zealand	-42.4237	-178.5609	1035	08/11/2012	51.4	580	F
TAN1248	New Zealand	-42.4616	-178.5121	844	08/11/2012	78.1	1980	F
TAN1261	New Zealand	-42.4616	-178.5121	844	08/11/2012	89.3	3610	F

Table A.2: Sample data for Deania calcea. *Samples were collated from different sources with varying levels of associated data.*

ID	Collection location			Depth (m)	Date Extracted	Length (cm)	Weight (g)	Sex
	Population	Latitude	Longitude					
12I9	Bay of Biscay	47.5633	-8.2247	1150.9	30/04/1997	110	6400	F
12J1	Bay of Biscay	47.5633	-8.2247	1150.9	30/04/1997	82	2200	M
12J10	Bay of Biscay	47.5633	-8.2247	1150.9	30/04/1997	82	2300	M
12J3	Bay of Biscay	47.5633	-8.2247	1150.9	30/04/1997	86	2800	F
12J4	Bay of Biscay	47.5633	-8.2247	1150.9	30/04/1997	95	3600	F
12J5	Bay of Biscay	47.5633	-8.2247	1150.9	30/04/1997	105	4600	F
12J8	Bay of Biscay	47.5633	-8.2247	1150.9	30/04/1997	81	2200	M
13A1	Bay of Biscay	47.5633	-8.2247	1150.9	30/04/1997	82	3100	F
13A10	Bay of Biscay	47.5633	-8.2247	1150.9	21/05/1997	92	3000	F
13A2	Bay of Biscay	47.5633	-8.2247	1150.9	30/04/1997	96	3900	F
13A4	Bay of Biscay	47.5633	-8.2247	1150.9	30/04/1997	92	3300	F
13A5	Bay of Biscay	47.5633	-8.2247	1150.9	30/04/1997	86	2200	M
13A9	Bay of Biscay	47.5633	-8.2247	1150.9	30/04/1997	88	2900	F
13B7	Bay of Biscay	47.5633	-8.2247	1150.9	21/05/1997	84	2200	M
14C1	Bay of Biscay	47.5633	-8.2247	1150.9	04/07/1997	83	2200	M
14C10	Bay of Biscay	47.5633	-8.2247	1150.9	04/07/1997	79	1800	M
14C5	Bay of Biscay	47.5633	-8.2247	1150.9	04/07/1997	81	2000	M
14C7	Bay of Biscay	47.5633	-8.2247	1150.9	04/07/1997	84	2400	M
14C8	Bay of Biscay	47.5633	-8.2247	1150.9	04/07/1997	87	2400	M
14D1	Bay of Biscay	47.5633	-8.2247	1150.9	04/07/1997	95	3400	F
14E6	Bay of Biscay	47.5375	-8.2096	1338.6	04/07/1997	82	2200	M
23C3	Hebrides	57.1665	-9.3955	1032	09/02/2007			
23C4	Hebrides	57.1665	-9.3955	1032	09/02/2007			
23C6	Hebrides	57.1665	-9.3955	1032	09/02/2007			
23C8	Hebrides	57.1665	-9.3955	1032	09/02/2007			
23D4	Hebrides	57.1665	-9.3955	1032	09/02/2007			
23E10	Hebrides	57.6593	-9.7290	1000	12/02/2007			
23E8	Hebrides	57.6593	-9.7290	1000	12/02/2007			
23F8	Hebrides	57.1665	-9.3955	1032	12/02/2007			
23G9	Hebrides	57.1665	-9.3955	1032	12/02/2007			
23J3	Hebrides	58.4328	-9.4922	1000	14/02/2007			
24E7	Hebrides	58.8612	-7.9318	1000	21/02/2007			
25B10	Hebrides	57.0878	-9.2990	525	06/03/2007			
25C2	Hebrides	57.0878	-9.2990	525	06/03/2007			
28G10	Hebrides	55.2353	-10.0393	500	27/06/2007			
28G8	Hebrides	56.8365	-9.1787	1030	27/06/2007			
28G9	Hebrides	55.2353	-10.0393	500	27/06/2007			
28H1	Hebrides	55.2353	-10.0393	500	27/06/2007			
28H2	Hebrides	55.2353	-10.0393	500	27/06/2007			
28J1	Hebrides	55.2928	-10.0613	750	27/06/2007			
28J10	Hebrides	55.2928	-10.0613	750	27/06/2007			
28J3	Hebrides	55.2928	-10.0613	750	27/06/2007			
28J5	Hebrides	55.2928	-10.0613	750	27/06/2007			

28J7	Hebrides	55.2928	-10.0613	750	27/06/2007			
28J8	Hebrides	55.2928	-10.0613	750	27/06/2007			
29A1	Hebrides	55.2928	-10.0613	750	27/06/2007			
29A10	Hebrides	55.2928	-10.0613	750	27/06/2007			
29A2	Hebrides	55.2928	-10.0613	750	27/06/2007			
29A3	Hebrides	55.2928	-10.0613	750	27/06/2007			
29A4	Hebrides	55.2928	-10.0613	750	27/06/2007			
29A5	Hebrides	55.2928	-10.0613	750	27/06/2007			
29A7	Hebrides	55.2928	-10.0613	750	27/06/2007			
29A8	Hebrides	55.2928	-10.0613	750	27/06/2007			
TAN1296	New Zealand	-42.4991	178.2136	646	17/01/2013	91.1	3270	F
TAN1298	New Zealand	-42.4991	178.2136	646	17/01/2013	84.9	2130	M
TAN1307	New Zealand	-42.4991	178.2136	646	17/01/2013	79.8	1900	M
TAN1308	New Zealand	-42.4991	178.2136	646	17/01/2013	97.3	3480	F
TAN1311	New Zealand	-42.4991	178.2136	646	17/01/2013	80	1950	F
TAN1312	New Zealand	-42.3908	176.1972	979	17/01/2013	90.8	2790	M

Bibliography

Agarwala, R. *et al.*, 2016. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 44(D1), pp.D7–D19.

Allain, V., 2001. Reproductive strategies of three deep-water benthopelagic fishes from the northeast Atlantic Ocean. *Fisheries Research*, 51(2–3), pp.165–176.

van Andel, T., 1989. Late Pleistocene Sea Levels and the Human Exploitation of the Shore and Shelf of Southern South Africa. *Journal of Field Archaeology*, 16(2), pp.135–155.

Anderson, P. & John, F., 1999. Community reorganization in the Gulf of Alaska following ocean climate regime shift. , 189, pp.117–123.

Andrews, K. & Luikart, G., 2014. Recent novel approaches for population genomics data analysis. *Molecular Ecology*, 23, pp.1661-1667.

Andrews, K. *et al.*, 2016. Harnessing the power of RADseq for ecological and evolutionary genomics - Supplement. *Nature Reviews Genetics*, 17(January), pp.81–92.

Antao, T. *et al.*, 2008. LOSITAN: A workbench to detect molecular adaptation based on a F st -outlier method. *BMC Bioinformatics*, 9(1), pp.1–5.

Baird, N. *et al.*, 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLoS ONE, 3(10), e3376.

Baum, J. *et al.*, 2003. Collapse and conservation of shark populations in the Northwest Atlantic. *Science*, 299(5605), pp.389–392.

Beaumont, M. & Nichols, R., 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, 263, pp.1619–1626.

Becker, B. *et al.*, 2007. Complex larval connectivity patterns among marine invertebrate populations. *Proceedings of the National Academy of Sciences of the United States of America*, 104(9), pp.3267–3272.

Beerli, P. & Felsenstein, J., 1999. Maximum-likelihood estimation of migration rates and effective population numbers. *Genetics*, 152(1990), pp.763–773.

Beerli, P. & Felsenstein, J., 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*, 98(8), pp.4563–8.

Bertrand, A. *et al.*, 2008. Schooling behaviour and environmental forcing in relation to anchoveta distribution: An analysis across multiple spatial scales. *Progress in Oceanography*, 79(2–4), pp.264–277.

Blaber, S. & Bulman, C., 1987. Diets of fishes of the upper continental slope of eastern Tasmania: content, calorific values, dietary overlap and trophic relationships. *Marine Biology*, (95), pp.345–356.

Block, B. *et al.*, 2005. Electronic tagging and population structure of Atlantic bluefin tuna. *Nature*, 434(7037), pp.1121-1127.

Bond, M. *et al.*, 2012. Reef sharks exhibit site-fidelity and higher relative abundance in marine reserves on the Mesoamerican Barrier reef. *PLoS ONE*, 7(3), e32983.

Bradbury, I. & Bentzen, P., 2007. Non-linear genetic isolatio by distance: implications for dispersal estimations in anadromous and marine fish populations. *Marine Ecology Progress Series*, 340, pp.245–257.

Bradbury, I. *et al.*, 2015. Transatlantic secondary contact in Atlantic Salmon, comparing microsatellites, a single nucleotide polymorphism array and restriction-site associated DNA sequencing for the resolution of complex spatial structure. *Molecular Ecology*, 24, pp.5130-

5144.

Brierley, A., 2014. Diel vertical migration. *Current Biology*, 24(22), pp.1074–1076.

Bryan, D. *et al.*, 2014. Quantitative video analysis of flatfish herding behavior and impact on effective area swept of a survey trawl. *Fisheries Research*, 154, pp.120–126.

Caddy, J. & Rodhouse, P., 1998. Cephalopod and groundfish landings: evidence for ecological change in global fisheries? *Reviews in Fish Biology and Fisheries*, 8(4), p.431.

Cariou, M. *et al.*, 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecology and Evolution*, 3(4), pp.846-852.

Carlsson J. *et al.*, 2007. Genetic Identity of YOY Bluefin Tuna from the Eastern and Western Atlantic Spawning Areas. *Journal of Heredity*, 98(1), pp.23-28.

Carlsson, J. *et al.*, 2011. Fine-scale population structure in a deep sea teleost (orange roughy, *Hoplostethus atlanticus*). *Deep Sea Research Part I: Oceanographic Research Papers*, 58(6), pp.627–636.

Catarino, D. *et al.*, 2015. The Pillars of Hercules as a bathymetric barrier to gene-flow promoting isolation in a global deep sea shark (*Centroscymnus coelolepis*). *Molecular Ecology*, 24, pp.6061–6079.

Catchen, J. *et al.*, 2013. Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22, pp.3124–3140.

Catchen, J. *et al.*, 2011. Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes,Genomes,Genetics*, 1(3), pp.171–182.

Catul, V. *et al.*, 2010. A review on mesopelagic fishes belonging to family Myctophidae. *Reviews in Fish Biology and Fisheries*, 21(3), pp.339–354.

Clark, M. & King, K., 1989. *Deepwater fish resources off the North Island, New Zealand: results of a trawl survey, May 1985–June 1986.*

Clark, M., 2009. Deep-sea seamount fisheries: a review of global status and future prospects. *Latin American Journal of Aquatic Research*, 37(3), pp.501-512.

Clarke, C. *et al.*, 2015. Global mitochondrial DNA phylogeography and population structure of the silky shark, Carcharhinus falciformis. *Marine Biology*, 162(5), pp.945–955.

Clarke, M. *et al.*, 2001. Aspects of reproduction of the deep water sharks Centroscymnus coelolepis and Centrophorus squamosus from west of Ireland and Scotland. *Cambridge University Press*, 81(6), pp.1019–1029.

Clarke, M. *et al.*, 2002. Catch, discarding, age estimation, growth and maturity of the squalid shark Deania calceus west and north of Ireland. *Fisheries Research*, 56(2), pp.139–153.

Coates, A. *et al.*, 1992. Closure of the Isthmus of Panama: The near-shore marine record of Costa Rica and western Panama. *Geological Society of America Bulletin*, 104(July), pp.814–828.

Collins, M. a *et al.*, 2005. Trends in body size across an environmental gradient: a differential response in scavenging and non-scavenging demersal deep sea fish. *Proceedings of the Royal Society B: Biological Sciences*, 272(1576), pp.2051–7.

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

Cowen, R. & Sponaugle, S., 2009. Larval Dispersal and Marine Population Connectivity. *Annu. Rev. Mar. Sci*, 1, pp.443–66.

Cruaud, A. *et al.*, 2014. Empirical Assessment of RAD Sequencing for interspecific Phylogeny. *Molecular Biology and Evolution*, 31(5), pp.1272-1274.

Cunha, R. *et al.*, 2012. Ancient Divergence in the Trans-Oceanic Deep sea Shark *Centroscymnus crepidater*. *PLoS ONE*, 7(11), e49196.

D'Aloia, C. *et al.*, 2015. Patterns, causes, and consequences of marine larval dispersal. *Proceedings of the National Academy of Sciences*, p.201513754.

Daley, R. *et al.*, 2014. Can marine reserves conserve vulnerable sharks in the deep sea? A case study of *Centrophorus zeehaani* (Centrophoridae), examined with acoustic telemetry. *Deep sea Research Part II: Topical Studies in Oceanography*, 115, pp.127–136.

Daley, R. *et al.*, 2002. Catch analysis and productivity of the deepwater dogfish resource in southern Australia. *CSIRO Marine Research, Fisheries Research and De- velopment Corporation (FRDC), and NSW Fisheries, Australia, FRDC Project No. 1998/108.*, (January).

Danovaro, R. *et al.*, 2014. Challenging the paradigms of deep sea ecology. *Trends in Ecology and Evolution*, 29(8), pp.465–475.

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

Davey, J. *et al.*, 2012. Special features of RAD Sequencing data: implications for genotyping. *Molecular Ecology*, 22, pp.3151-3164.

Derycke, S. *et al.*, 2013. Dispersal and gene flow in free-living marine nematodes. *Frontiers in Zoology*, 10(1), p.1.

Van Dijk, E. *et al.*, 2014. Ten years of next-generation sequencing technology. *Trends in Genetics*, 30(9), pp.418–426.

Drazen, J. & Haedrich, R., 2012. A continuum of life histories in deep sea demersal fishes. *Deep Sea Research Part I: Oceanographic Research Papers*, 61, pp.34–42.

Dudgeon, C. *et al.*, 2012. A review of the application of molecular genetics for fisheries management and conservation of sharks and rays. *Journal of Fish Biology*, 80(7), pp.1789–1843.

Dudgeon, C. *et al.*, 2012. A review of the application of molecular genetics for fisheries management and conservation of sharks and rays. *Journal of Fish Biology*, 80(5), pp.1789–1843.

Dulvy, N. *et al.*, 2008. You can swim but you can't hide: the global status and conservation of oceanic pelagic sharks and rays. *Aquaric Conservation: Marine and freshwater Ecosystems*, 18(March), pp.459–482.

Dunn, M. *et al.*, 2010. The diet of deepwater sharks and the benefits of using DNA identification of prey. *Deep Sea Research Part I: Oceanographic Research Papers*, 57(7), pp.923–930.

Dunn, M. *et al.*, 2013. Trophic Interactions and Distribution of Some Squaliforme Sharks , Including New Diet Descriptions for *Deania calcea* and *Squalus acanthias*. *PLoS ONE*, 8(3), e59938.

Ebert, D. *et al.*,1991. A preliminary investigation of the feeding ecology of squaloid sharks off the west coast of southern Africa. *South African Journal of Marine Science*, 10(November), pp.71–81.

European Union, 2016. *Proposal for a COUNCIL REGULATION fixing for 2017 and 2018 the fishing opportunities for Union fishing vessels for certaindeep sea fish stocks*,

Excoffier, L. & Lischer, H., 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10, pp.564–567.

Foll, M. & 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.

Forbes, E., 1844. Report on the Mollusca and Radiata of the Aegean sea, and on their distribution, considered as bearing on geology. *Report of the British Association for the Advancement of Science*, pp.130–193.

Fuller, C. *et al.*, 2009. The challenges of sequencing by synthesis. *Nature Biotechnology*, 27(11), pp.1013–1023.

Gaither, M. *et al.*, 2016. 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.

Galindo, H. *et al.*, 2006. Seascape Genetics: A Coupled Oceanographic-Genetic Model Predicts Population Structure of Caribbean Corals. *Current Biology*, 16(16), pp.1622–1626.

Gautier, M. *et al.*, 2013. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology*, 22, pp.3165-3178.

Garcia, V. *et al.*, 2008. The importance of habitat and life history to extinction risk in sharks, skates, rays and chimaeras. *Proceedings of the Royal Society B: Biological Sciences*, 275, pp.83–89.

Girman, D. *et al.*, 1993. Molecular Genetic and Morphological Analyses of the African Wild Dog (*Lycaon pictus*). *Journal of Heredity*, 84(6), pp.450–459.

Graham, K. *et al.*,2001. Changes in relative abundance of sharks and rays on Australian South East Fishery trawl grounds after twenty years of fishing. *Marine and Freshwater Research*, 52, pp.549–61.

Grassle, J., 1989. Species diversity in deep sea communities. *Tree*, 4(1), pp.12–15.

Grassle, J. & Maciolek, N., 1992. Deep Sea Species Richness - Regional and Local Diversity Estimates From Quantitative Bottom Samples. *American Naturalist*, 139(2), pp.313–341.

Grytnes, J. & Vetaas, O., 2002. Species richness and altitude: a comparison between null models and interpolated plant species richness along the Himalayan altitudinal gradient, Nepal. *The American naturalist*, 159(3), pp.294–304.

Gubili, C. *et al.*, 2016. Connectivity in the deep: Phylogeography of the velvet belly lanternshark. *Deep Sea Research Part I: Oceanographic Research Papers*, 115, pp.233–239.

Hardy, O. & Vekemans, X., 1999. Isolation by distance in a continuous population: Reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity*, 83(2), pp.145–154.

Hedrick, P., 2011. *Genetics of Populations* 4th Editio., Jones & Bartlett Learning.

Heller, M., 2002. DNA Microarray Technology: Devices, Systems, and Applications. *Annual review of biomedical engineering*, 4(1), pp.129–153.

Hessler, R. & Sanders, H., 1967. Faunal diversity in the deep sea. *Deep Sea Research*, 14(1), pp.65–78.

Hey, J., 2007. IM - a computer program for the analysis of genetic data from closely related populations and species. *Molecular Biology and Evolution*, (2004), pp.1–16.

Hey, J., 2010. Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, 27(4), pp.905–920.

Hickford, M. & Schiel, D., 2003. Comparative dispersal of larvae from demersal versus pelagic spawning fishes. *Marine Ecology Progress Series*, 252, pp.255–271.

Hoban, S. *et al.*, 2011. Computer simulations: tools for population and evolutionary genetics. *Nat Rev Genet*, 13(2), pp.110–122.

Hohenlohe, P. *et al.*, 2013. Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular Ecology*, 22, pp.3002-3013.

Huateng, H. & Knowles, L., 2016. Unforeseen Consequences of Excluding Missing Data from Next-Generation Sequences: Simulation Study of RAD Sequences. *Systematic Biology*, 65(3), pp.657-365.

Ingram, T., 2011. Speciation along a depth gradient in a marine adaptive radiation. *Proceedings of the Royal Society B: Biological Sciences*, 278(1705), pp.613–618.

Irvine, S. *et al.*, 2012. Biological vulnerability of two exploited sharks of the genus *Deania* (Centrophoridae). *Journal of Fish Biology*, 80(5), pp.1181–1206.

Irvine, S. *et al.*, 2006. Surface bands on deepwater squalid dorsal-fin spines : an alternative method for ageing *Centroselachus crepidater*. *Canadian Journal of Fisheries and Aquatic Science*, 627, pp.617–627.

Jackson, J. *et al.*, 2001. Historical overfishing and the recent collapse of coastal ecosystems. *Science (New York, N.Y.)*, 293(5530), pp.629–37.

Jeffries, D. *et al.*, 2016. Comparing RADseq and microsatellites to infer complex phylogeographic patterns, an empirical perspective in the Crucian carp, *Carassius carassius*, L. *Molecular Ecology*, pp.2997–3018.

Jennings, R. *et al.*, 2013. Population Differentiation and Species Formation in the Deep Sea: The Potential Role of Environmental Gradients and Depth. *PLoS ONE*, 8(10), e77594.

Jennings, S. *et al.*, 1999. Structural change in an exploited fish community: A consequence of differential fishing effects on species with contrasting life histories. *Journal of Animal Ecology*, 68(3), pp.617–627.

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

Jombart, T. & Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), pp.3070–3071.

Jombart, T. *et al.*, 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11(1), pp.1–15.

Jombart, T. *et al.*, 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, 11(1), pp.1–15.

Kapitonov, V. & Jurka, J., 2004. Harbinger Transposons and an Ancient HARBI1 Gene Derived from a Transposase. *DNA and Cell Biology*, 23(5), pp.311–324.

Knutsen, H. *et al.*, 2012. Population genetic structure in a deepwater fish *Coryphaenoides rupestris*: Patterns and processes. *Marine Ecology Progress Series*, 460(2008), pp.233–246.

Kohler, N. & Turner, P., 2001. Shark tagging: A review of conventional methods and studies. *Environmental Biology of Fishes*, 60(1–3), pp.191–223.

Koslow, J., 1996. Energetic and life-history patterns of deep sea benthic, benthopelagic and seamount-associated fish. *Journal of Fish Biology*, 49(sa), pp.54–74.

Koslow, J. *et al.*, 1995. Fecundity and its variability in orange roughy: effects of population density, condition, egg size, and senescence. *Journal of Fish Biology*, 47(6), pp.1063–1080.

Koutsikopoulos, C. & Le Cann, B., 1996. Physical processes and hydrological structures related to the Bay of Biscay anchovy. *Scientia Marina*, 60(2), pp.9–19.

Kyne, P. & Simpfendorfer, C., 2007. A Collation and Summarization of Available Data on Deepwater Chondrichtyans: Biodiversity, Life History and Fisheries. *IUCN Report*, (February), pp.1–137.

Last, P. & Stevens, J., 1994. *Sharks and rays of Australia*, 1st Edition, Cambridge, Massachusetts: Havard University Press.

Levin, L. *et al.*, 2001. Environmental Influences on Regional Deep Sea Species Diversity. *Annual Review of Ecology and Systematics*, 32, pp.51–93.

Llorente, M. *et al.*, 2015. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science*, 350(6262), pp.820–822.

Machado, P. & Figueiredo, I., 2000. A technique for ageing the birdbeak dogfish (*Deania calcea* Lowe, 1839) from dorsal spines. *Fisheries Research*, 45(1), pp.93–98.

Manel, S. *et al.*, 2003. Landscape genetics: Combining landscape ecology and population genetics. *Trends in Ecology and Evolution*, 18(4), pp.189–197.

Marean, C. *et al.*, 2007. Early human use of marine resources and pigment in South Africa during the Middle Pleistocene. *Nature*, 449(7164), pp.905–908.

Marques, T. *et al.*, 2014. The Use of Stable Isotopes Analysis in Wildlife Studies. In *Applied Ecology and Human Dimensions in Biological Conservation*. pp. 159–174.

Mastretta-Yanes, A. *et al.*, 2015. Restriction site-associated DNA sequencing, genotyping error estimation and *de novo* assembly optimization for population genetic inference. *Molecular Ecology Resources*, 15, pp.28-41.

Mauchline, J. & Gordon, J., 1983. Diets of the sharks and chimaeroids of Rockall Trough, northeaster Atlentic Ocean. *Marine Biology*, 75, pp.269–278.

McClain, C. & Hardy, S., 2010. The dynamics of biogeographic ranges in the deep sea. *Proceedings of the Royal Society B: Biological Sciences*, 277(1700), pp.3533–3546.

Mendez, M. *et al.*, 2010. Isolation by environmental distance in mobile marine species: Molecular ecology of franciscana dolphins at their southern range. *Molecular Ecology*, 19, pp.2212–2228.

Moore, D. *et al.*, 2013. Population biology and ageing of the deep water sharks Galeus melastomus, Centroselachus crepidater and Apristurus aphyodes from the Rockall Trough, north-east Atlantic. *Journal of the Marine Biological Association of the United Kingdom*, 93(7), pp.1941–1950.

Morato, T. *et al.*, 2006. Fishing down the deep. *Fish and Fisheries*, 7(1), pp.24–34.

Munroe, D. *et al.*, 2012. The role of larval dispersal in metapopulation gene flow: Local population dynamics matter. *Journal of Marine Research*, 70(2), pp.441–467.

Musick, J. *et al.*, 2004. Historical zoogeography of the Selachii. *Biology of sharks and their relatives*, pp.33–78.

Myers, R. & Worm, B., 2003. Rapid worldwide depletion of predatory fish communities. *Nature*, 423(6937), pp.280–283.

Myers, R. a & Worm, B., 2005. Extinction, survival or recovery of large predatory fishes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 360(1453), pp.13–20.

Narum, S. *et al.*, 2013. Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, 22, pp.2841–2847.

Neat, F. *et al.*, 2008. The deepwater ecosystem of the continental shelf slope and seamounts of the Rockall Trough: a report on the ecology and biodiversity based on FRS surveys. *Fisheries Research Services*, February.

Nielsen, R, 2005. Molecular Signals of Natural Selection. *Nature Reviews Genetics*, 39(August), pp.197-218.

Nielsen, R. *et al.*, 2011. Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics*, 12(6), pp.443-451.

Norse, E. *et al.*, 2012. Sustainability of deep sea fisheries. *Marine Policy*, 36(2), pp.307–320.

Oscar, P. *et al.*, 2009. Estimating dispersal from genetic isolation by distance in a coral reef fish (*Hypoplectrus puella*). *Ecology*, 90(11), pp.3087–3098.

Papastamatiou, Y. & Lowe, C., 2012. An analytical and hypothesis-driven approach to elasmobranch movement studies. *Journal of Fish Biology*, 80(5), pp.1342–1360.

Patarnello, T. *et al.*, 2007. Pillars of Hercules: Is the Atlantic-Mediterranean transition a phylogeographical break? *Molecular Ecology*, 16, pp.4426–4444.

Pauly, D. *et al.*, 2002. Towards sustainability in world fisheries. *Nature*, 418(August), pp.689–695.

Peterson, B. *et al.*, 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE*, 7(5), e37135.

Pethybridge, H. *et al.*, 2011. Diet of demersal sharks and chimaeras inferred by fatty acid profiles and stomach content analysis. *Journal of Experimental Marine Biology and Ecology*, 409(1–2), pp.290–299.

Planes, S. & Fauvelot, C., 2002. Isolation by distance and vicariance drive genetic structure of a coral reef fish in the Pacific Ocean. *Journal of Zoology*, 56(2), pp.378–399.

Pogson, G. *et al.*, 2001. Isolation by distance in the Atlantic cod, *Gadus morhua*, at large and small geographic scales. *Evolution*, 55(1), pp.131–146.

Priede, I. *et al.*, 2013. The ecosystem of the Mid-Atlantic Ridge at the sub-polar front and Charlie – Gibbs Fracture Zone ; ECO-MAR project strategy and description of the sampling programme 2007 – 2010. *Deep sea Research Part II*, 98, pp.220–230.

Pritchard, J., *et al.*, 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155(2), pp.945–959.

Raj, A. *et al.*, 2014. FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, 197(2), pp.573–589.

Reid, J., 1961. On the temperature, salinity, and density differences between the Atlantic and Pacific oceans in the upper kilometre. *Deep Sea Research (1953)*, 7(4), pp.265–275.

Rex, M. & Etter, R., 2010. *Deep sea biodiversity : pattern and scale* 1st Edition, Cambridge, Massachusetts: Havard University Press.

Rice, W. & Hostert, E., 1993. Laboratory experiments on speciation: what have we learned in 40 years? *Evolution; international journal of organic evolution*, 47(6), pp.1637–1653.

Ritchie, H. *et al.*, 2013. Population genetic structure of the abyssal grenadier (*Coryphaenoides armatus*) around the mid-Atlantic ridge. *Deep sea Research Part II*, 98, pp.431–437.

Roberts, C., 2002. Deep impact: the rising toll of fishing in the deep sea. *Trends in Ecology & Evolution*, 17(5), pp.242–245.

Ruhl, H., 2004. Shifts in Deep sea Community Structure Linked to Climate. *Science*, 513(2004).

Rundle, H. & Nosil, P., 2005. Ecological speciation. *Ecology Letters*, 8(3), pp.336–352.

Saldanha, L. *et al.*, 1995. Observations on the Diet of Some Slope Dwelling Fishes of Southern Portugal. *International Review of Hydrobiology*, 80(2), pp.217–234.

Scheltema, R., 1968. Dispersal of Larvae by Equatorial Ocean Currents and its Importance to the Zoogeography of Shoal-water Tropical Species. *Nature*, 217(5134), pp.1159–1162.

Schultz, J. *et al.*, 2008. Global phylogeography and seascape genetics of the lemon sharks (genus *Negaprion*). *Molecular Ecology*, 17, pp.5336–5348.

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

Sethuraman, A. & Hey, J., 2016. IMa2p - parallel MCMC and inference of ancient demography under the Isolation with migration (IM) model. *Molecular Ecology Resources*, 16(1), pp.206–215.

Shendure, J. & Ji, H., 2008. Next-generation DNA sequencing. *Nature Biotechnology*, 26(10), pp.1135-1145.

Silva, C. *et al.*, 2017. Infection of the lesser spotted dogfish with *Proleptus obtusus* Dujardin, 1845 (Nematoda: Spirurida) reflects ontogenetic feeding behaviour and seasonal differences in prey availability. *Acta Parasitologica*, 62(2).

Simpfendorfer, C. & Kyne, P., 2009. Limited potential to recover from overfishing raises concerns for deep sea sharks, rays and chimaeras. *Environmental Conservation*, 36(2), pp.97–103.

Stevens, J., 2003a. *Centroselachus crepidater*. *SSG Australia & Oceania Regional Workshop, March 2003*. Available at: http://www.iucnredlist.org/details/46864/0 [Accessed December 31, 2015].

Stevens, J., 2003b. *Deania calcea*. *SSG Australia & Oceania Regional Workshop, March 2003*. Available at: http://www.iucnredlist.org/details/41798/0 [Accessed December 31, 2015].

Straube, N. *et al.*, 2011. Cryptic diversity and species assignment of large lantern sharks of the *Etmopterus spinax* clade from the Southern Hemisphere (Squaliformes, Etmopteridae). *Zoologica Scripta*, 40(1), pp.61–75.

Stuckas, H. *et al.*, 2017. Combining hydrodynamic modelling with genetics: Can passive larval drift shape the genetic structure of Baltic Mytilus populations? *Molecular Ecology*, In Press.

Strugnall, J. *et al.*, 2008. The thermohaline expressway: the Southern Ocean as a centre of origin for deep-sea octopuses. *Cladistics*, 24, pp. 853-860.

Tautz, D. *et al.*, 2010. Next generation molecular ecology. *Molecular Ecology*, 19(Suppl. 1), pp.1–3.

Teixeira, S. *et al.*, 2012. Panmixia in a fragmented and unstable environment: The hydrothermal shrimp *Rimicaris exoculata* disperses extensively along the Mid-Atlantic ridge. *PLoS ONE*, 7(6), e38521.

Varela, A. *et al.*, 2013. Global genetic population structure in the commercially exploited deep sea teleost orange roughy (*Hoplostethus atlanticus*) based on microsatellite DNA analyses. *Fisheries Research*, 140(October), pp.83–90.

Varela, A. *et al.*, 2012. Low levels of global genetic differentiation and population expansion in the deep sea teleost *Hoplostethus atlanticus* revealed by mitochondrial DNA sequences. *Marine Biology*, 159(5), pp.1049–1060.

Veríssimo, A. *et al.*, 2011. Population structure of a deep-water squaloid shark, the Portuguese dogfish (*Centroscymnus coelolepis*). *ICES Journal of Marine Science*, 68, pp.555–563.

Veríssimo, A. *et al.*, 2012. Genetic population structure and connectivity in a commercially exploited and wide-ranging deepwater shark, the leafscale gulper (*Centrophorus squamosus*). *Marine and Freshwater Research*, 63, pp.505–512.

Vitti, J. *et al.*, 2013. Detecting Natural Selection in Genomic Data. *Nature Reviews Genetics*, 47(November), pp. 97-120.

Wallberg, A. *et al.*, 2014. A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nature genetics*, 46(August), pp.1081–1088.

Ward, R. *et al.*, 2005. DNA barcoding Australia's fish species. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 360(September), pp.1847–1857.

Weir, B. & Cockerham, C., 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, 38(6), pp.1358–1370.

Wetherbee, B., 2000. Assemblage of deep sea sharks on Chatham Rise, New Zealand. *Fishery Bulletin*, 98(1), pp.189–198.

White, T. *et al.*, 2011. Genetic panmixia and demographic dependence across the North Atlantic in the deep sea fish, blue hake (*Antimora rostrata*). *Heredity*, 106(4), pp.690–9.

White, T. *et al.*, 2009. Unexpected panmixia in a long-lived, deep sea fish with well-defined

spawning habitat and relatively low fecundity. *Molecular Ecology*, 18, pp.2563–2573.

White, T. *et al.*, 2011. Comparative assessment of population genetics and demographic history of two congeneric deep sea fish species living at different depths. *Marine Ecology Progress Series*, 434, pp.155–164.

White, T. *et al.*, 2010. Local selection and population structure in a deep sea fish, the roundnose grenadier (*Coryphaenoides rupestris*). *Molecular Ecology*, 19, pp.216–226.

White, T. *et al.*, 2009. Unexpected panmixia in a long-lived, deep sea fish with well-defined spawning habitat and relatively low fecundity. *Molecular Ecology*, 18, pp.2563–2573.

Wright, S., 1943. Isolation by Distance. *Genetics*, 28(2), pp.114–138.

Willing, E. *et al.*, 2012. Estimates of Genetic Differentiation Measured by F_{ST} Do Not Necessarily Require Large Sample Sizes When Using Many SNP Markers. PLoS ONE, 7(8), e42649.

Wright, S., 1951. The Genetical Structure of Populations. *Nature*, 15(4), pp.323–354.

Yano, K., 1991. Catch Distribution, Stomach Contents and Size at Maturity of Two Squaloid Sharks, *Deania calceus* and *D. crepidalbus*, from the Southeast Atlantic off Namibia. *Bull. Japan Soc. Fish. Oceanogr.*, 55(3), pp.189–196.

Yau, C. *et al.*, 2002. Scavenging by megabenthos and demersal fish on the South Georgia slope. *Antarctic Science*, 14(I), pp.16–24.