Connecting the past, present and future: A population genomic study of Australasian snapper (*Chrysophrys auratus*) in New Zealand

By

Tom Oosting

A thesis submitted to the Victoria University of Wellington in fulfilment of the requirements for the degree of Doctor of Philosophy

Victoria University of Wellington
School of Biological Sciences
2021

Primary Supervisor:
Dr. Peter A. Ritchie Victoria University of Wellington

Secondary Supervisors:
Dr. Maren Wellenreuther Plant and Food Research Nelson
Auckland Univeristy
Dr. Nic J. Rawlance Otago University

Abstract

Advances in genomic methods now enable the study of wild populations and their evolutionary history at an unprecedented level. The genotyping of many thousands of genetic markers across the genome provides high statistical resolution. This enables the identification of adaptive genetic variation, providing novel insights into population demography and the processes driving population divergence. Marine fish are ideal candidates to study the processes driving evolutionary divergence because selection works efficiently in large populations, and marine populations can be distributed over large spatial ranges and occupy a range of environmental conditions. This thesis used whole-genome variant data to study the Australasian snapper (Chrysophrys auratus, tāmure) in New Zealand. Snapper is one of New Zealand's largest inshore fisheries and has experienced significant population reductions. The aims of this thesis were to investigate the genome-wide variation in snapper in New Zealand and 1) assess the neutral and adaptive population genetic structure, 2) reconstruct the demographic history, and 3) identify genomic regions, genes and their functions that show signs of selection.

Population genomic structure was assessed using whole-genome resequencing data from 350 individuals, and this data set resulted in 167,543 assumed neutrally evolving loci (SNPs). It was found that levels of genetic diversity were not significantly different between populations, suggesting that fishing pressure has not lead to local reductions in genetic variation. Levels of genetic differentiation between sampled populations was low, with significant evidence for isolation by distance ($R2 = 0.75$, $p = 0.002$). Pairwise FST estimates and PCA/DAPC showed the presence of two genetic clusters, one containing the northern and one containing the southern populations. Genetic disjunctions combined with mixing between the clusters was detected around the Mahia peninsula and Cape Reinga. The identification of adaptive loci enabled the identification of fine-scale population structure, reflecting currently recognized stocks. The ability to differentiate between stocks is fundamental for fisheries management. The patterns detected here show promising results for future implementation into fisheries management of snapper stocks.

Contemporary and ancient mitochondrial genomes were used to assess the demographic, and phylogeographic history of snapper. Analyses indicated that haplotype diversity was high (0.968-0.982), which is commonly observed in species with large populations sizes. Mitochondrial genomes showed the presence of two lineages that diverged approximately 650,000 (490,000 – 840,000) years ago. The separation was likely linked to reductions in sea level during glacial cycles. Estimates of changes in population size show strong support for an exponential population size increase after the last glacial maximum (LGM). Changes in population abundance based on the Bayesian Skyline plot indicated a strong population increase approximately 10,000 years ago. The steep increase in new branches in the phylogenetic tree suggests population sizes increase approximately 20,000 (7,000-35,000) years ago. A post-glacial expansion is the most likely explanation for the observed increase in population abundance. During this period, sea levels rose which presumably reconnected fragmented populations, and subsequent increased sea temperatures allowed for southward expansion.

Whole-genome sequences from contemporary snapper populations were used to identify genes under selection. Analyses were conducted to detect selection in a single genetic cluster (divergent selection), or both genetic clusters (nation-wide selection). In total, 101 genomic regions containing 253 different genes showed evidence for selection. Two genomic regions showed strong evidence for divergent selection between the northern and southern cluster (FST > 0.2). The regions contained two genes associated with glycolysis which are linked to (cell-) growth (i.e. mast2 and hk2). The regions containing hk2 showed a lack of rare alleles (TD > 2) in the southern cluster, consistent with balancing selection maintaining multiple alleles in the population. Variation in growth rate may be maintained throughout the genetic cluster because of a latitudinal gradient in sea temperature. Strong evidence for selective sweeps were detected in two genomic regions on a nation-wide level. Both regions contained genes associated with angiogenesis (mydgf and rnf213a), which has been shown to affect maturation in species of fish. While tentative, it is possible that intense size-selective fishing is selecting for early maturation in snapper, a life life-history commonly associated with fishing-induced evolution. A selection scan contrasting the population Tasman Bay and

Karamea Bight was performed to test for evidence of adaption to cold stress. Selection was detected in 123 genomic regions containing 296 genes, of which 197 potentially experience divergent selection. Two genes were located in regions that showed significant evidence of selection (camk2g and ksr2). Both genes have been associated with cold stress in previous studies, suggesting the Karamea Bight could represent an adaptive front at the southern range of the distribution of snapper.

This thesis presents the first population genomic study of Australasian snapper in New Zealand, a species with a diverse genetic landscape and a rich evolutionary history. The detection of fine-scale population structure through adaptive differences between populations highlights the promising application of genomics in fisheries management. The study of mitochondrial lineages showed the effect of glacial cycles, providing insights into how New Zealand's marine fauna has been affected by major changes in global climate. Finally, the identification of genes and associated biological traits under selection has provided fundamental new insights regarding the environmental conditions that drive adaptive change and act on phenotypes. Snapper is an ideal species for developing and integrating genomics into New Zealand fisheries management. A detailed understanding of fish stock demography and adaptive potential is critical to support improvement to fisheries management as wild stocks continue to face strong anthropogenic pressures (e.g. climate change and overexploitation). Genomics provides valuable insights into how stock assessments and harvesting levels can be better set to match the natural biological units of a species that are determined by gene flow and adaptive variation.

Acknowledgements

I want to start by thanking everyone who has been part of my life in last 3.5 years while doing my PhD. The opportunity to move to New Zealand and explore this amazing country has been incredible, and I thank everyone who was involved.

A massive thank you to all my supervisors. My primary supervisor Dr Peter Ritchie, I have an incredible appreciation and great respect for the way you guide your students. You have given me the opportunity to explore my academic interests and provided support and direction when needed. I am very excited we will be able to continue to work together. Thank you to all my secondary supervisors for providing support during different stages of my PhD. Dr Maren Wellenreuther, I have rarely met anybody who is so incredibly hard-working, your energy and engagement in research astounds me. Thank you for all the advice and support you have given me. Dr Nic Rawlence, thank you for guiding me through the world of ancient DNA research. The experiences of sampling ancient moa bones and holding the skull of the Haast eagle were truly unique. I also need to thank my non-official supervisor (but supervisor nonetheless) Dr Bastiaan Star. I have learnt so much during my time at Oslo University. You are one of the kindest person I have had the pleasure of getting to know during my PhD. Thank you to Professor Per Palsbøll and Dr Martine Bérubé who supervised me during my Bachelors and Masters. You were both instrumental in my early development as a scientist.

A big thanks to all the students of the Ritchie lab, especially Yvan Papa and Noemie Valenza-Troubat. We three started our PhD around the same time under the wings of Peter and Maren, and I am incredibly grateful to have gone through the experience with you two. Yvan, I have had great joy from working with you every day and having someone to go through all the troubles that a PhD brings. We work incredibly well together, and I hope we will keep doing so in the future. Noemie, although we did not see each other on day to day basis, I always appreciated your uplifting energy and kind personality. It is a rare event when two people hand in their thesis on the same day, and we will forever have that connection. Thank you to all the people here at the School of Biological Sciences for providing a great working, and more importantly, social environment. Thank you to all the friends that I have made here in Wellington during my PhD. You are all the reason that I can call Wellington home. There are too many

people to name, but special shout out to El, Randy, Tory, Dez, and Jack. A special thanks to Henry Jones, Charlotte Panton, Jack Ewing, Nevena Novakovic, and Vanessa Chalk. I could not have asked for better to people to live with, you have all become like family to me.

Thank you to all my friends back home in the Netherlands that I have kept in touch with over the last 3.5 years. It is a true testament to all the strong friendships I have back home and will keep forever. Special Thanks to Rik Veldhuis and Jurjan van der Zee, two of my dearest friends. The time we got to spend in New Zealand was incredible and I am happy you got to take part in this adventure. Jurjan, also thank you for being partner-in-crime in population genetics. We started our academic journeys together during our bachelors, and it has been great to have such a close friend to talk about some of the bigger things in life. I look forward to the day we have our own lab.

The biggest thank you goes to my family. Ep Oosting and Lammie Stoffer, I could not have asked for better parents. Thank you for all the love and support you have given me over the years. Thank you to Daan Oosting, I am incredibly proud to have you as my brother. It is amazing to see how you have created your own success with the devotion to things you truly care for.

External contributions

This research was funded by the Marsden Foundation (16-VUW-040), under the title "Testing for Fishing-Induced Evolution using DNA from Ancient and Modern Snapper". All laboratory work, analysis and writing were conducted by the author with the help of supervisors Dr. Peter Ritchie, Dr. Maren Wellenreuther, and Nic J Rawlence.
Several other important contributors are acknowledges below.

Sample collection:
Ministry of Primary Industries, and their team of observers of commercial fishing vessels collected the majority of samples.

Gisborne Tatapouri Sports Fishing Club, Hawkes Bay Sports Fishing Club, and Petone Working Man's Club provided the opportunity to collect samples at one of their fishing competitions.

National Institute of Water and Atmospheric Research (NIWA) provided samples, including samples collected during their 2019 trawling survey in Tasman Bay, Golden bay, and West Coast (KAH1902). Additional samples were provided by Darren Parsons.

Ancient DNA work:
Māori iwi including Ngāti Kuia, Ngāi Tai Ki Tāmaki, Ngāti Toa, Ngāti Hei and Ngāti Koata supporting the use of archaeological samples from their rohe

Te Papa Museum and Carolyn McGill for providing archaeological samples

Ancient DNA extractions were performed at Otago University under supervision of Dr. Nic Rawlence by Alex Verry, MSc and Lachie Scarsbrook, MSc.

Library preparation and sequencing of ancient DNA samples was performed at the University of Oslo under the supervision of Dr. Bastiaan Star and Dr. Sanne Boessenkool by Dr. Giada Ferrari.

Table of contents

Data availability statement

Custom made scripts were used throughout this thesis. When mention in the text, reported scripts can be found on https://github.com/tomoosting/PhD_thesis.

Chapter 1. General introduction

1.1. Population & evolutionary genetics

1.1.1. Genetic diversity and populations

Biodiversity on Earth can be classified based on the phenotypic similarities of individuals, which are primarily determined by genetics. Groups of sexually reproducing individuals exchange and mix their genetic material and consequently share genetic similarities and belong to the same gene pool of a population (Palsbøll et al., 2007). The genetic diversity present within populations reflects multiple processes, and includes past and present processes (Ellegren & Galtier, 2016). Genetic diversity within a population is influenced by four key evolutionary processes: 1) mutation and recombination, 2) genetic drift, 3) selection and 4) gene flow. It is important to state that when talking about populations in a genetic context, population sizes are expressed as the effective population size (the number of reproducing individuals within a population; Ne) and not the census population size (physical number of individuals in a population; N_c) (Lande & Barrowclough, 1987). Genetic diversity in a population reflects the individuals that have contributed to the genetic pool through reproduction.

Genetic variation is generated through *mutations* and *recombination*. Mutations generate genetic variation by introducing new genetic variants in a population, which occurs at a very slow rate. In sexually reproducing organisms, new combinations of genetic variation are generated by combining genetic variants from both parents through recombination. Harmful genetic variation (deleterious) will likely be filtered out of the population, and only a very small fraction of genetic variation will provide an adaptive advantage. Most genetic variation that persists in the population is neutral or nearly neutral (Kimura, 1968; Ohta, 1992), implying that they either do not influence the phenotype, or if they do, that they do not give the individual an adaptive advantage in their current environment (Barrett & Schluter, 2008). Neutral genetic variation makes up the majority of genetic diversity within a population and it contains the adaptive potential for when a population experiences a new environmental challenge. Neutral genetic variation responds to demographic changes in ways we can predict using population and evolutionary genetic theory, and provides is with the ability to infer

population sizes (Charlesworth et al., 2003), connectivity between populations, phylogenetic relations, and past demographic events (discussed below).

Genetic drift is the stochastic loss of genetic variation from a population (Lande, 1976). The amount of genetic variation populations can hold depends on the strength of genetic drift, which is related to the size of the population. The effect of drift is inversely related to the size of the population. This means that in large populations the stochastic loss of genetic variation is small, allowing higher levels of genetic diversity to be maintained in the population. In contrast, the effect of stochastic loss is high in small populations, which limits the level of genetic variation that can be maintained in the population.

Selection favours the propagation of certain variants of genetic variation (alleles) based on the relative number of offspring that survive with a given genome. Adaptive alleles spread through the population, potentially removing all (hard sweep) or most (soft sweep) alleles in that part of the genome, reducing variation (Vitti et al., 2013). Selection and genetic drift are often opposing forces. If selection is too weak to overcome the stochastic loss of genetic variation, advantageous alleles can still be lost. Thus, selection is much more efficient in large populations where the effect of drift of small.

Gene flow is the exchange of alleles and genotypes between populations (Slatkin, 1987). Genetic diversity present within a population can be influenced by gene flow. Here, the movement of individuals to a population that was not their natal group and sexual reproduction with resident individuals results in either the introduction of new variants and/or the homogenisation of genetic diversity between populations. Inversely, when there is no gene flow among populations then drift (and selection) can operate independently leading to genetic differentiation. Population differentiation will continue to increase over time between two groups that remain isolated from each other.

1.1.2. Environmental factors influencing genetic diversity

Geographical and biological changes in the environment can cause populations to expand, contract, which can result in the disruption of gene flow and the emergence of separate populations. This influences how genetic variation is temporally and spatially distributed (genetic structure). By studying the genetic structure and levels of diversity,

we can gain fundamental insights into what processes have shaped the evolution of populations.

Geographic processes, such as climate change, play a major role in the formation of genetic structure (Hewitt, 2000). For example, the separation of populations on each side of the equator (anti-tropical distribution) is associated with high ocean temperatures at low latitudes during periods of global warming (Briggs, 1987; Koufopanou et al., 1999). Increased ocean temperatures have been suggested to disrupt ocean circulation, generating isolated pockets of high primary productivity (Pastene et al., 2007). These fragmented populations are then separated geographically and can become genetically diverged. During periods of global cooling (e.g. ice ages), glaciers fragment terrestrial habitats, forcing populations to contract and survive in small isolated refugial populations (Shafer et al., 2010; Skog et al., 2009). Similarly, sea levels can drop by up to 120 meters during glacial periods (Peltier & Fairbanks, 2006), exposing land that act as barriers between previously connected ocean habitats, separating populations. For example, the genetic diversity in multiple species in the Indo-Pacific region suggests there was a barrier between the Indian and the Pacific Ocean during the Pleistocene glacial cycles in the past 470,000 years (Bowen et al., 2016; Siddall et al., 2003). Populations that become fragmented and survive in isolated regions can experience strong reductions in population size, also known as a bottleneck. This reduction in population size increases genetic drift and reduces genetic diversity. Patterns that arose tens of thousands of years ago can still be detected in present-day populations using genomic sequencing approaches (Fauvelot et al., 2003).

Biological factors that influence genetic structure include changes in migration patterns, predation pressure, diet, and/or mating behaviour. For example, multiple species of Malawi cichlids (*Amphilophus sp.*) found within the same African lake have genetically diverged and developed into multiple species. This example of genetic differentiation was driven by sexual selection and niche separation (Barluenga et al., 2006). Likewise, three different ecotypes (highly diverged populations but no yet different species) of killer whales (*Orcinus orca*) are present within the Eastern Pacific Ocean (Ford et al., 1996). These three groups overlap in their distribution, yet are demographically independent and genetically distinct populations and have different

feeding strategies (Foote & Morin, 2016; Ford et al., 1998). It is not clear whether biological factors can fully explain the genetic divergence between the ecotypes (Foote & Morin, 2015). Foote and Morin (2015) argue that temporary dispersal into the Atlantic Ocean could have contributed to the diversification of the different ecotypes. However, strong behavioural differences between the populations are maintaining the genetic differentiation between ecotypes.

1.2. Genetics

1.2.1. Genetic markers

Genetic markers (loci, or locus when singular) are used to estimate the levels and patterns of genetic variation of a population. A wide variety of types of genetic markers have been developed during the advancement of molecular methods (Doveri et al., 2008). The following section summarises the markers relevant to this thesis, and describes their key characteristics. In animals, genetic marker sets are commonly obtained from two genomic sources: mitochondrial DNA and nuclear DNA.

Mitochondrial markers

Mitochondrial markers have been popular throughout modern population genetic research (Moritz, 1994a). Key characteristics that make mitochondrial DNA so useful are its haploid nature, high abundance in cells, and regions that have different evolutionary rates.

All organisms have a mitochondrial genome that is clonally inherited through the maternal lineage (Elson & Lightowlers, 2006). The lack of recombination makes it very useful for investigating population processes that occurred thousands of year ago as the pattern is not erased due to mixing by sexual reproduction. Mitochondrial markers only represent the evolutionary history of female lineage in a population. Clonal inheritance means that offspring receive their mother's copy of mitochondrial DNA, and these only changes when mutation occur. This form of inheritance does not produce recombinants, where genetic material from both parents is combined to generate new combinations of mutants (Harrison, 1989). All variation on the mitochondrial genome is linked and cannot independently propagate through a population.

Mitochondrial DNA is highly abundant in cells (between 100 and 10,000 copies) making it much easier to isolate than nuclear DNA and more often present in degraded

historical samples (Bendich, 1987; Rizzi et al., 2012). Mitochondrial loci are suitable markers to use from sources such as faecal material, shedded cells, environmental DNA (eDNA), and ancient DNA (aDNA) (discussed below).

While the mitochondrial genome is typically about 16,000 bp and circular, which is extremely small when compared to the nuclear genome, it contains a variety of genetic markers with different evolutionary rates (Mueller, 2006). Depending on the evolutionary rate of a locus, evolutionary processes that occur on different time scales can be investigated. Loci with slow evolutionary rates are more suited for studying long-term evolutionary processes such as phylogenetic relationships that involve multiple taxa (Harrison, 1989). The versatility of mitochondrial loci combined with the relative ease of resolving its DNA sequence makes mitochondrial DNA a popular choice for a range of population and phylogenetic studies.

Nuclear markers

The nuclear genome makes up the vast majority of the genetic material in a cell (Morin et al., 2004), and behaves in fundamentally different ways compared to the mitochondrial genome. Most relevant is that the nuclear genome in many species contains two copies (diploid), and because of sexual reproduction the nuclear DNA experiences recombination (Figure 1.1A). Many types of genetic markers have been developed for the nuclear genome including allozymes, microsatellites, randomly amplified polymorphic DNA (RAPD), and restriction fragment length polymorphism (RFLP) (Doveri et al., 2008). However, the genetic markers relevant for this thesis and most dominant in modern population genetic research are single nucleotide polymorphisms (SNPs) (Morin et al., 2004).

SNPs are one of the most abundant types of variation in the genome (Brookes, 1999;

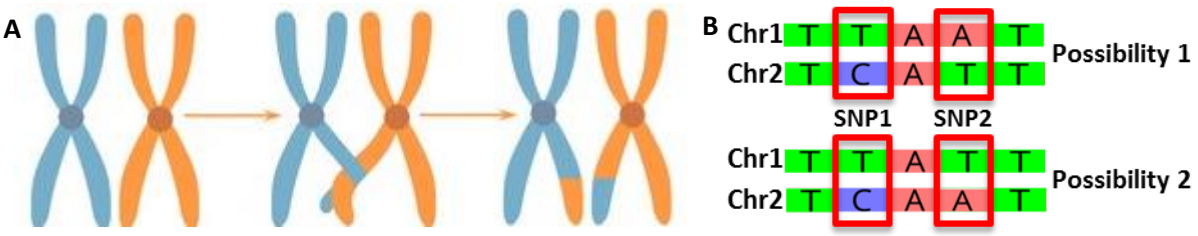

Figure 1.1: A) The process of recombination where genetic variants are being broken up and new combinations of genetic variation are generated. B) Schematic representation showing the two possibilities of how alleles from two single nucleotide polymorphisms can distributed over two chromosomes, creating uncertainly around which genetic variants are inherited together. Chr = chromosome, SNP = single nucleotide polymorphism, nucleotides (A,C,T) show different alleles.

Catanach et al., 2019). When SNPs are identified it is often difficult to determine, alleles are located on the same chromosome and thus inherited together (Figure 1.1B). This is because recombination breaks up genetic association over time (Figure 1.1A). The chance that two SNPs on the same chromosome are broken up in a population increases with their physical distance to each other. This is important, because it means the SNPs on the nuclear genome can propagate independently from each other through a population when they are far apart (or on separate chromosomes). This is in contrast to mitochondrial DNA where genetic associations cannot break up. The level of physical association between two SNPs (linkage disequilibrium) can be estimated as r^2 (Reich et al., 2001), and when this value is small two loci can be considered independent. SNPs that are independent from each other are more likely to have different evolutionary histories and are often assumed to be independent data points to infer evolutionary processes.

1.2.2. From genetics to genomics

Modern genetic research started with the development of Sanger sequencing and the polymerase chain reaction (PCR) (Saiki et al., 1985; Sanger et al., 1977). These developments allowed the determination of the consecutive sequence of nucleotides (i.e. DNA sequence) at large scales. It was these technological advances that eventually enabled whole genomes to be sequenced (Venter et al., 2001). Genetic research traditionally involved a small number of genetic markers (1-100 loci) that were usually of specific interest to the particular study. In population genetics, this commonly involves using a small number of (assumed to be neutrally evolving) loci to investigate evolutionary processes. Here, genetic loci are known or assumed to be genetically unlinked so they represent different data point. The small number of loci often implies limited statistical power to detect population differences, unless genetic differences are substantial.

In a similar way that genetic research accelerated with the development of PCR and Sanger sequencing, the field of genomics started with the development of massive-parallel sequencing (Mardis, 2017), also known as next-generation-sequencing (NGS). In this process, millions of sequencing reactions occur simultaneously generating a vast amount of genetic data. Genomics involves the identification of many thousands of

genetic markers across the genome and studies their interactions. This enables the study of all processes that influence genetic diversity across the genome (i.e. mutations, recombination, drift, selection, and gene flow). Adaptive genetic variation can now be detected by searching the genome for genetic markers that deviate from neutral expectations. The large number of genetic markers provides a better representation of the genetic diversity in the population as they cover the genome more densely (Luikart et al., 2003; Luikart et al., 2018) and thus provide more statistical resolution to detect small differences between populations, or more subtle demographic changes. One important development is that reference genomes are becoming available for a large number of non-model organisms (Koepfli et al., 2015; Lewin et al., 2018). For instance, it provides improved identification of loci, linking genetic markers to candidate genes, and haplotype-based analyses (Brandies et al., 2019; Hohenlohe et al., 2020). Genomics can be used to identify a range of genetic markers including structural variants which are mutations that change the length of the genome through inserting, deleting and duplicating genetic sequences, or entire genomic regions, or reverse their orientation (inversions) (Wellenreuther et al., 2019b). These types of variation have been shown to play an important role in the adaptation to new habitats (Prunier et al., 2019; Wellenreuther & Bernatchez, 2018). However, this thesis solely utilises the identification of SNPs to help explain evolutionary processes.

1.2.3. Ancient DNA

Ancient DNA is the use of DNA from biological material that is typically over 100 years old. The field of ancient DNA is a set of methods that are used to capture DNA from samples that are generally considered to contain a low number of DNA copies and study the genetic variation of past populations or extinct species. This approach has provided insights that would otherwise have been impossible to obtain from only sampling modern DNA samples. For example, the extraction of aDNA from Neanderthals and other hominoids has provided unique insights into the evolution of humans, including that Neanderthals contributed DNA to modern humans (*Homo sapiens*) (Green et al., 2010; Meyer et al., 2012; Prüfer et al., 2014). It also provided fascinating insights into the evolution of mammoths, which showed that genes associated with cold adaptation were already present one million years before that type

of environment had appeared (van der Valk et al., 2021). Ancient DNA also provides valuable information about extant species by showing how populations have changed over time, or capture genetic variation that has since been lost (Hutchinson et al., 2003; Speller et al., 2012).

The main challenge with the application of aDNA is that DNA rapidly degrades over time, making it much more difficult to obtain long DNA sequences and comprehensive coverage of genomics areas. Generally after an organism dies, nucleases cut and hence degrade DNA into short fragments (Graham et al., 2015; Oosting et al., 2020). Under some conditions, the nucleases themselves are destroyed or immobilized early in this process. In addition, DNA can be degraded by oxidation, hydrolysis and UV radiation, and the rate of decay depends on the environmental conditions (Schroeder et al., 2006). Cold and dry conditions are generally good for preserving DNA, such as Polar Regions and some caves. Under perfect conditions, it is estimated that DNA could be extracted from biological materials up to one million years old (Hofreiter et al., 2001). Another issue with aDNA samples is they can become contaminated with DNA from other organisms (Carpenter et al., 2013). This also includes contamination from microorganisms or DNA present in the environment while aDNA is extracted. To minimize contamination, samples need to be treated with extreme care. This includes careful retrieval from their site of preservation. Ancient DNA is typically isolated in a dedicated laboratory with strict hygiene protocols that are designed to avoid the introduction of modern DNA (Hofreiter & Shapiro, 2012).

Mitochondrial DNA has proven to be extremely valuable for aDNA research because cells contain between 100 and 10,000 copies of mitochondrial DNA, compared to only two copies of the nuclear genome in a cell (Bendich, 1987; Rizzi et al., 2012). This relatively high copy number makes it likely that enough copies of the region of interest have survived the process of degradation to enable retrieval for DNA sequencing and subsequent error checking (because multiple reads can be achieved).

The development of massive-parallel-sequencing has revolutionised the way aDNA sequences can now be retrieved from samples (Der Sarkissian et al., 2015). Instead of targeting a small section of the genome, approaches such as shotgun sequencing enable all of the DNA in a sample to be sequenced. One key problem is the level of

endogenous DNA (fraction of DNA belonging to the species of interest) in the sample. In the early days of aDNA research only samples with high levels of endogenous DNA were suitable for genomic research. However, the use of DNA enrichment approaches can be used to pull out endogenous DNA fragments from samples even though it is only present in low quantities (Carpenter et al., 2013).

1.3. Fisheries research

1.3.1. Fishing

Fishing has been central to human existence, with extensive evidence of fishing activity going as far back as 40,000 years (Gartside & Kirkegaard, 2010). Most cultures around the world have a rich history of fishing (Sahrhage & Lundbeck, 2012). Fish were so abundant throughout human history that they were thought to be inexhaustible. This idea was supported by the famous evolutionary biologist Thomas H. Huxley, a strong advocate of Darwin's theory of evolution. While being on the UK Royal Commission on the Sea Fisheries in 1866, he proclaimed that *"Probably all the great sea fisheries are inexhaustible; that is to say, that nothing we do seriously affects the numbers of fish. Any attempt to regulate these fisheries seems consequently, from the nature of the case, to be useless".* In 1883, during his inaugural speech at International Fisheries Exhibition in London he said that "*it is inconceivable that the great sea fisheries, such as those for cod, herring and mackerel, could ever be exhausted*" (Sims & Southward, 2006). We now know that Atlantic Cod was the first major fisheries to collapse due to overfishing (Myers et al., 1997).

The development of fishing on an industrial scale during the 1950s has led to large reductions in species abundance. This high level of extraction has been driven by expanding markets to feed an exponentially growing human population (FAO, 2020). Fish are an important protein source that humans still directly harvest from wild populations, which started thousands of years ago. Fishing effort has drastically increased since the onset of industrialisation in the 1950s (Sahrhage & Lundbeck, 2012). In 2018, global fish production was estimated at 178.5 million tonnes, of which 96.4 million tonnes (54%) was extracted from wild populations, and 82.1 million tonnes (46%) was produced from aquaculture (FAO, 2020). Fish account for 17% of the global intake of animal protein, providing over 3.3 billion people (42% of the human

population) with 20% of their average intake of animal protein. Particularly for developing countries in Asia and Africa, fish often provides over 50% of the average intake of animal protein. Maintaining viable fish stocks is therefore key to continue to supply people with protein and preserving functioning marine ecosystems.

Over 34% of fished species were harvested at unsustainable levels in 2017 (FAO, 2020). This intense fishing effort has already caused the collapse of numerous fish stocks (Pauly et al., 2002). Best known is the collapse of the North Atlantic cod fisheries, which was once was the largest fisheries in the world (Myers et al., 1997). Previous studies have suggested that global collapse of the world's fisheries could occur by 2048 (Worm et al., 2006). Stock collapses have serious consequences for the fish species, the ecosystems they live in, and the people that rely on them.

1.3.2. Fisheries management

Fisheries management aims to protect the long-term viability of fished stocks while sustainably extracting a portion of the population for consumption. Documented requests for regulation of fishing practices goes back as far as 1376 (Caddy & Cochrane, 2001), where a petition expressed concerns about the use of an early type of small mesh beam trawl. Some of the earliest fishing regulations included the closing of fisheries during spawning, minimum size rules, and bans of specific types of fishing gear. Since the 1960s, many countries have introduced total allowable catch (TAC) rates (Figure 1.2), which are used to regulate the total amount of fish that can be extracted annually. An extension to this approach is the use of individual transferable quotas (ITQs), where proportions of the total allowable commercial catch (TACC) are sold to fishing enterprises. This allows them to harvest their proportion of the TACC. ITQs have been implemented by most developed countries, with New Zealand having the most developed and widespread use of ITQs (Beddington et al., 2007; Hilborn et al., 2020). To set TACs, fisheries managers often utilise management strategy evaluation (MSE), where management strategies are being evaluated against operating models that are based on the biological components of the fisheries in question (e.g. growth rates, recruitment, and mortality rates) (Punt et al., 2016). The implementation of these management tools have shown mixed results. Assessments suggest that the

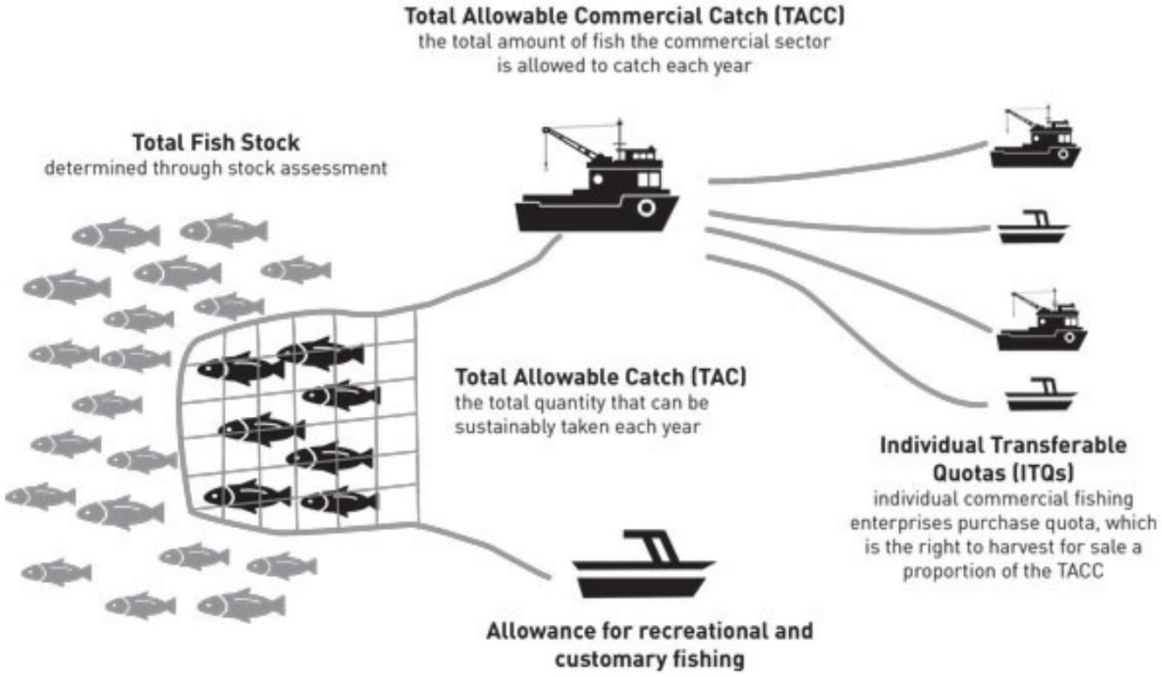

Figure 1.2: Overview of the quota management system (QMA) of New Zealand showing how total allowable catch (TAC) based on stock assessments is distributed between the commercial and recreational fishers. The implementation of individual transferable quotas (ITQs) to distribute fishing quotas among fishers has adapted by a select number of countries, including New Zealand. Source: (Ministry of Fisheries, 2011)

implementation of TACs are broadly reducing fishing pressure and managed stocks are showing increases in stock abundance (Costello et al., 2016; Hilborn et al., 2020; Hilborn & Ovando, 2014). However, reports from the Food and Agriculture Organization (FAO) suggest that the fraction of overfished stocks has increased from 27 to 33 per cent between 2000 and 2017 (FAO, 2020; Hilborn et al., 2020). This can primarily be attributed to countries that have not implemented MSE based strategies to manage fisheries stocks (Hilborn et al., 2020). These studies highlight the value of effective fisheries management. Even in countries that have already implemented MSE based strategies, fisheries management can be improved through regular stock assessments that refine operating models (Punt et al., 2016).

Stock assessments provide valuable information which is used to determine what proportion of the population can be sustainably harvested. A common management strategy aims to maximize long-term yields while maintaining a constant population size, which is commonly referred to as the maximum sustainable yield (MSY) (Figure 1.3) (Hilborn, 2012). Here, the population is kept at a population size (B_{MSY}) where population growth is highest. To sustainably harvest these stocks, only the annual net

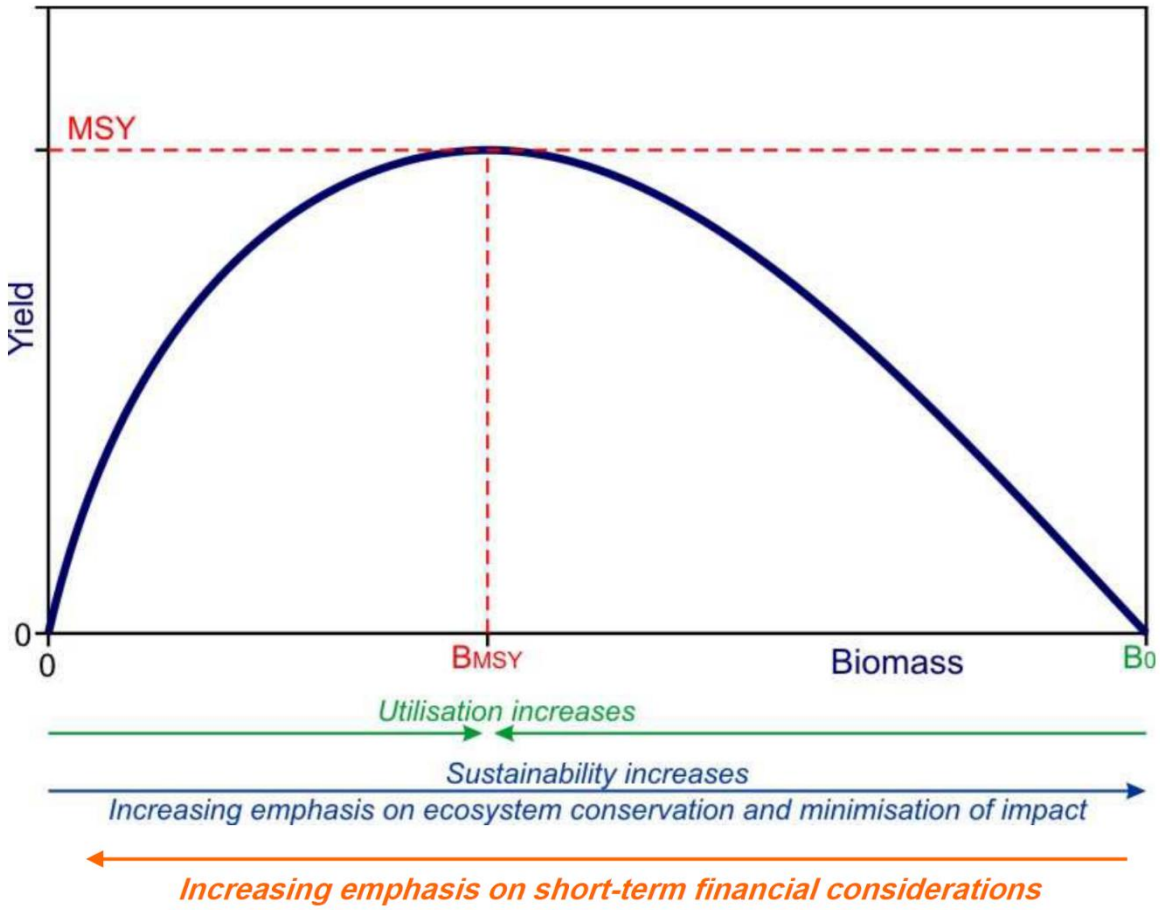

Figure 1.3: Plot showing the theoretical framework behind maximising catch rates while maintaining a sustainable fisheries. Maximum sustainable yield is maximum catch which can be sustained harvested, it represents to point where population growth of the fished stock maximised and only that proportion is harvested annually. MSY = maximum sustainable yield, BMSY = biomass at maximum sustainable yield, Bo = biomass at maximum carrying capacity/virgin biomass. Source: Ministry for Primary Industries (2021)

growth in biomass is extracted for consumption (MSY). When populations are overfished, management actions are required to assure stocks can recover (Ministry of Fisheries, 2011). When the biomass of the fished stock drops below the soft limit (½B_{MSY}, or 20% of the virgin biomass (B_o), whichever is higher), fishing quotas should be adjusted for the stocks to recover. If the biomass of a fished stock drops below the hard limit (¼B_{MSY}, or 10% B_o), whichever is higher), the fished stock is considered collapsed and should be closed until it has recovered (Ministry of Fisheries, 2011).

For these fisheries management strategies to be effective it is crucial to identify demographically independent stocks (Begg et al., 1999). The delineation of stock boundaries provides crucial information that is required to determine how many fish can be sustainably extracted. The identification of fish stocks can be performed using a range of methods. These include direct approaches such as mark-recapture and

telemetry methods that directly assess the movement of individual fish (Pine et al., 2003; Thorstad et al., 2013). Other methods include the detection of differences in growth, age structure, and diet to identify demographically separated stocks (Fisheries New Zealand, 2018). Genetic stock identification has been widely used to attempt to delineate stocks boundaries (Araujo et al., 2014). However, genetic approaches have limited power to detect separate stocks when genetic differences are small (Ward, 2000). Here, genomic approaches show great potential for the management of fisheries stocks.

1.3.3. Genomics and fisheries management

So far, the application of genomics in fisheries has been largely underutilised (Bernatchez et al., 2017; Papa et al., 2020). This is partly due to the lack of power that conventional genetic methods have to address fisheries management issues. Fish have large population sizes and can have high rates of dispersal, resulting in very small genetic differences between stocks (Ward, 2000). A relatively small number of migrants between populations are needed to provide gene flow that homogenises genetic variation across demographically independent stocks. Here, the statistical power to identify population differentiation is often lacking when only a relatively small number of neutral loci are available. Genomics enables the identification of a large number of neutral SNPs that can provide a much higher statistical power to identify small genetic differences (Benestan et al., 2015). The use of genomics also enables the identification of adaptive SNPs that may be specific to demographically independent stocks due to local adaptations (Benestan et al., 2015; Euclide et al., 2021; Larson et al., 2014).

The identification of genetic variants related to local adaptation can provide new insights into environmental factors that are driving the reproductive isolation between stocks, and genes that are related to these adaptations. Stock structure based on adaptive SNPs would better align the productivity or reproductive success of units within the population to the management of fishing pressure. Once informative SNPs have been identified, it would allow for more low-cost sequencing approaches to be developed that specifically target samples for these SNPs using very large sample sizes. While stocks identification through adaptive SNPs is easiest to scale up, it is possible that adaptive differences between demographically independent stocks are not captured with SNPs. Ecological adaptation to local environments is often driven by

structural variants, making them useful for the identification of separate stocks (Wellenreuther et al., 2019b). For instance, chromosomal rearrangements in Atlantic cod (*Gadus morhua*) have been used to identify separate populations that have adapted to their local environment (Barth et al., 2017; Berg et al., 2016; Kirubakaran et al., 2016). Unfortunately, methods for the detection of structural variation are underdeveloped compared to SNP detection, and long-read sequencing technologies that enable the detection of large structural variants are expensive (Wellenreuther et al., 2019b). As the field of genomics advances, structural variants will likely play a valuable role in fisheries management. Adaptive variation can also be used to detect which genes and phenotypes are under selection to monitor what environmental conditions are facilitating evolutionary change.

The wealth of information present in genomic data sets has a multitude of applications for fisheries management. As presented above, the ability to differentiate between populations with low levels of genetic differentiation can help evaluate accuracy of stock boundaries. In addition, increased statistical power allows for better estimates of migration rates (gene flow) between identified stocks, this could help resolve questions regarding the level of demographic independence of stocks. Genomics can also help develop marker panels that can be used to genetically identify fish. Genetic tagging could be an alternative to mark-recapture methods that use physical tagging, which is known to suffer from post-release mortality, loss of tags, and low report rates (Ovenden et al., 2015). Such methods can be used to assess the dispersal of individuals, and help estimate population sizes (Bravington et al., 2016; Manel et al., 2005). Long-term genetic monitoring also provides the opportunity to monitor changes in genetic diversity (Ovenden et al., 2015). The health of fished stocks are primarily assessed using abundance estimates (biomass). However, monitoring levels of genetic diversity would provide important information about the health and resilience of a fished stock. The level of genetic variation present within populations allows them to adapt to environmental pressures (Barrett & Schluter, 2008). Exploited stocks may experience loss of genetic diversity (Hauser et al., 2002), lowering their resilience. Integrating genetic monitoring into fisheries management helps to assure that fished stocks

maintain their resilience. One prominent environmental pressure that stocks are exposed to is fishing itself.

1.3.4. Fishing-induced evolution

Intense size-selective fishing pressure is hypothesised to be a powerful driver of evolutionary change (Heino et al., 2015). Fishing is an inherently non-random process, designed to catch larger fish. As a consequence, small fish have a higher chance of survival compared to larger fish. High mortality rates imply that individuals that reproduce early have a better chance of passing on their genes to the next generation. As a result, intense size-selective fishing pressure selects for fast life-history traits such as small body size and reproduction early in life (Figure 1.4) (Enberg et al., 2012). This leads to populations where fish have reduced growth rates and faster maturation.

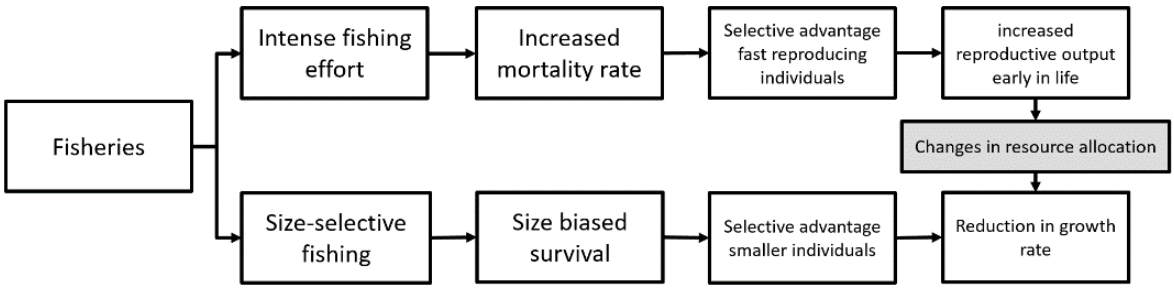

Figure 1.4: Predicted responses of fish populations to intense and size-selective fishing effort. Grey box indicates indirect effect on life-history trait.

Reductions in growth rate are also predicted to be affected indirectly through shifts in resource allocation towards reproduction (grey box in Figure 1.4) (Enberg et al., 2012). Changes in life-history traits associated with fishing-induced evolution have been observed in a range of species (extensively reviewed by Heino et al. (2015)).

The observed phenotypic changes associated with fishing could be caused by either density-dependent processes (phenotypic plasticity), or directional selection on allele frequencies (genetic adaptation). The distinction between the two processes is of crucial importance as phenotypic changes through genetic adaptation may be difficult to reverse if the allele responsible reaches or becomes close to fixation in the population. Although the theoretical framework that underpins fishing-induced evolution is well established, providing empirical genetic evidence of selection has proven to be far more difficult. Establishing evidence for fishing-induced evolution requires i) a causal link

between (size-selective) fishing and phenotypic changes in the population, ii) excluding (at least partially) other confounding factors that could influence the phenotypic traits, iii) providing evidence of genetic change for the observed phenotypic change, and iv) showing this has occurred in a wild exploited population.

Controlled size-selection experiments were able to show a causal link between removal based on fish size and changes in growth (Conover & Baumann, 2009; Conover & Munch, 2002; Diaz Pauli & Heino, 2014). Other studies were also able to relate the observed changes in growth with genetic changes in wild populations (Therkildsen et al., 2019; Uusi-Heikkilä et al., 2017; Uusi-Heikkilä et al., 2015; van Wijk et al., 2013). While selection experiments are useful to exclude confounding factors, they are conducted in laboratory conditions that are unrealistic in nature. Ecological interactions, such as predation, are known to affect the same life-history traits. In wild populations, evidence in support of fishing-induced evolution is based on phenotypic data (e.g. body size, growth rates, maturation rates), without linking the observed phenotypic changes to genetic change. For example, intense fishing effort has resulted in changes in maturation rate in Atlantic Cod (*G. morhua*) (Barot et al., 2004; Heino et al., 2002; Marty et al., 2014; Olsen et al., 2004; Swain et al., 2007). For an extensive review in fishing-induced evolution see Heino et al. (2015). New Zealand has a history of heavy size-selective fishing, which increased significantly with the mechanisation of fishing methods in the 1950s. However, the possibility of whether size-selective fishing has induced any evolutionary changes in New Zealand fish stocks has never been investigated.

1.4. Australasian snapper (*Chrysophrys auratus*)

The Australasian snapper (*Chrysophrys auratus*), referred to as snapper hereafter, has been one of the most heavily fished species in New Zealand using a size-selective fishing approach. Snapper belongs to the family Sparidae with the red seabream (*Pagrus major*) occurring in Japanese waters as its closest relative. In Te Reo Māori, snapper is known as tāmure. Snapper are found in the waters around the North Island and top of the South Island of New Zealand, and southern half of Australia (Parsons et al., 2014). They are serial broadcast spawners, where females can release over 60,000 eggs in one batch (Zeldis, 1993; Zeldis & Francis, 1998). Less than two per cent of

larvae survive the open water stage before they settle as juveniles in shallow estuarine like habitats after one month (Parsons et al., 2014). Around one year of age, snapper start moving into shallow coastal habitats and primarily occur at depths up to 50 meters, with 200 meters considered as their maximum depth. Snapper are protogynous hermaphrodites, which means that all individuals start as immature females and develop into sexually mature males or females between three to four years of age at around 20-30 cm fork length, and thought to reach up to 60 years of age (Parsons et al., 2014).

Snapper is one the New Zealand's biggest inshore fisheries with an annual total allowable commercial catch (TACC) of 6,400 tonnes (Fisheries New Zealand, 2018). Snapper fisheries are managed using six quota management areas that are thought to reflect biologically separated groups (Figure 1.5) (Fisheries New Zealand, 2018). However, multiple demographically independent stocks are recognized within some of these management areas. SNA1 which is the largest management area in terms of TAC (8,050 tonnes) is thought to consists of three separate demographically independent stocks (i.e. Northland, Hauraki Gulf, and Bay of Plenty) (Walsh et al., 2011). In SNA2, two stocks are thought to be present that are separated around the Mahia peninsula (Walsh et al., 2012). SNA7 likely consists of two separate stocks present in Tasman Bay and the Marlborough Sounds (Fisheries New Zealand, 2018). SNA8 is thought to consist of on single reproductive stock (Fisheries New Zealand, 2018). Snapper abundance is low in management areas SNA3, and fishing on the Kermadec (SNA10) is highly restricted.

Trawling in New Zealand started around the beginning of the 20th century, and fish stocks were subjected to intense fishing effort in the 1960s. In snapper, this has led to drastic population declines that caused some stocks to drop below 20% of the original biomass (Figure 1.6), indicating the stock is overfished. In 1984, New Zealand started managing its stocks with the quota management system (Annala, 1996). Since then, stocks have shown signs of recovery (Fisheries New Zealand, 2018). All three stocks in SNA1 are thought to be around 20% of their original biomass, however high commercial and recreational catches are predicted to cause overfishing (Fisheries New Zealand, 2018). The status of the stocks in SNA2 are unknown. Snapper abundance is showing

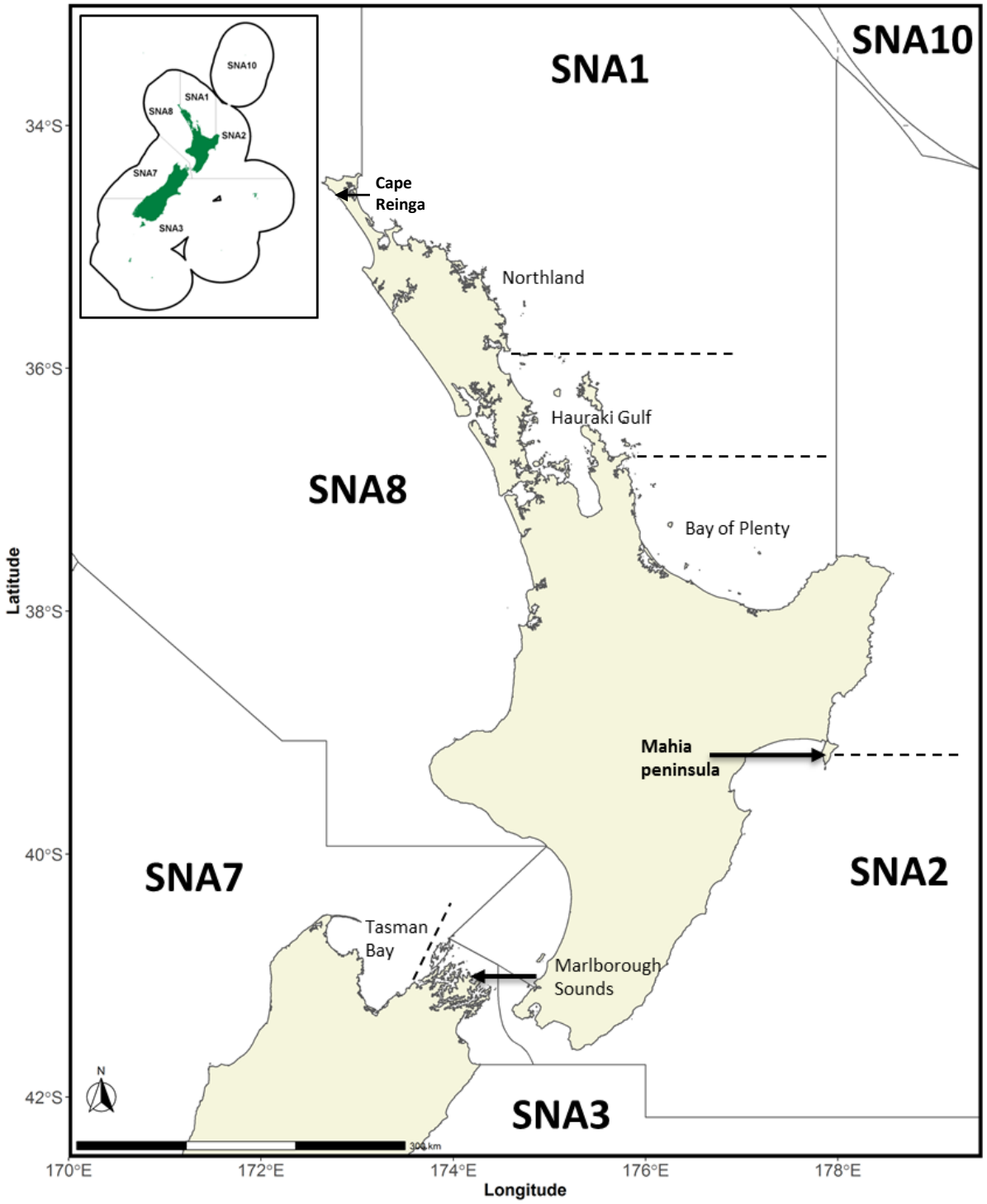

Figure 1.5: Overview of snapper management areas and stocks. Dashed black lines indicate boundaries between different reproductively isolated stocks. Source top left figure: Fisheries New Zealand (2018).

good signs of recovery in SNA7 which was estimated to be at around 38% of its original biomass. SNA8 is showing signs of recovery, but it has likely not reached 20% of its original biomass. While snapper experienced the highest fishing pressure during the last century, it has a long history of being fished in New Zealand.

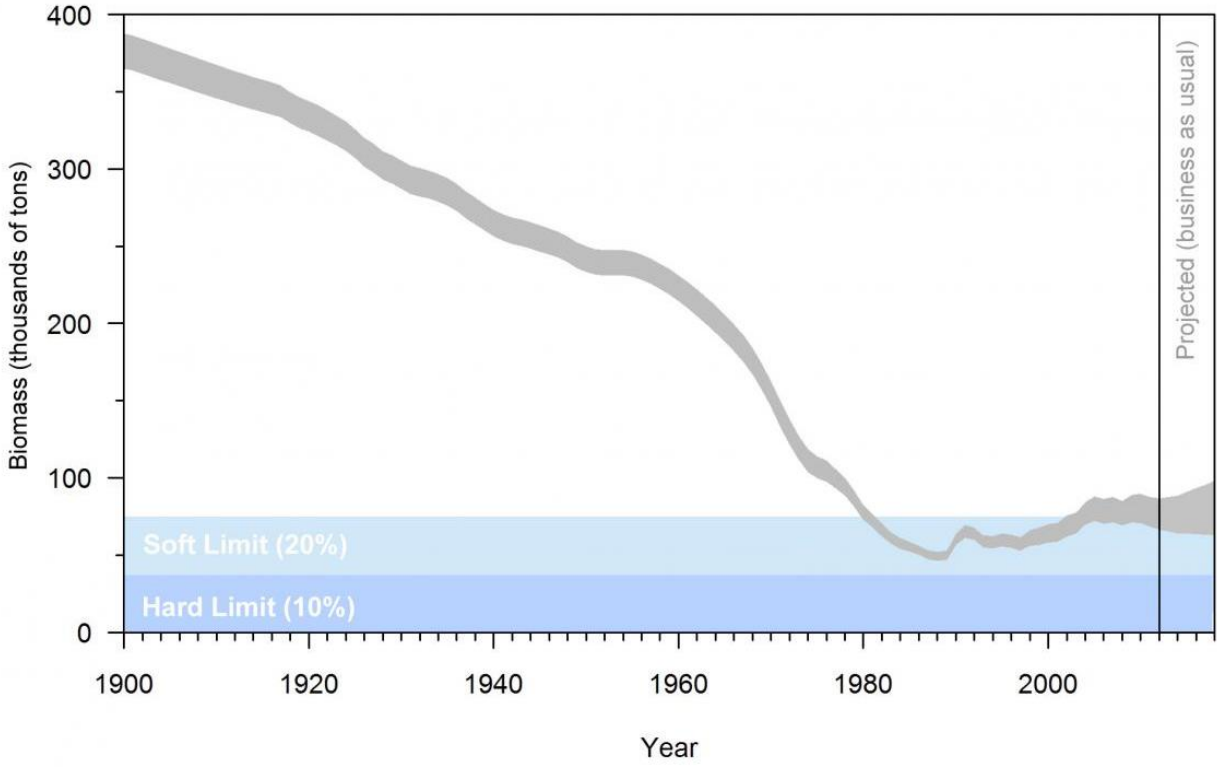

Figure 1.6: Abundance trend of New Zealand snapper in SNA1. Soft limit indicates the biomass level (20% of the original biomass) below which a fish stock is considered to be overfished or depleted and needs to be actively rebuilt using . Hard limit indicates the biomass level (10% of the original biomass) below which a fish stock is considered to have collapsed and fisheries may need to be closed to rebuild at the fastest possible rate.

Snapper has been an important food source for Māori since they arrived in New Zealand (1280-1450 CE) (Leach, 2006; Leach & Davidson, 2000). Evidence of fishing in New Zealand can be found as early as 1450 CE (Leach, 2006). Snapper comprises roughly 16% of the archaeological fish material in New Zealand (Leach, 2006), highlighting the role this species played in Māori culture. Early signs of the exploitation of snapper can also be found in the archaeological record, where the representation of snapper bones in archaeological record reduced significantly over time (Fig. S3.2B). Rich archaeological deposits of snapper remains have been found at archaeological sites. Morphometric analyses showed that these historical snapper were much larger compared to their contemporary counterpart (Leach, 2006). Also, anecdotal evidence about the exploitation of snapper in New Zealand in the 19th century suggests populations were heavily fished prior to the start of catch records (Parsons et al., 2009). This also suggests that original biomass estimates (biomass when fishing records started) used for monitoring stocks are likely an underestimation of the true original biomass.

Previous studies on snapper reported evidence of weak genetic structure around New Zealand (Ashton, 2013; Bernal-Ramírez et al., 2003; Smith et al., 1978). While no

genetic structure was able to be detected using mitochondrial markers, significant population differentiation was detected using nuclear makers (i.e. allozymes and microsatellites). Genetic disjunctions were reported around to the top of the North Island (i.e. Cape Reinga), and on the East Coast of the North Island (i.e. Mahia Peninsula) (Figure 1.5). Populations on the West Coast and Hawke's Bay showed low levels of genetic differentiation which was remarkable as they are situated on opposite sides of the North Island. It was hypothesized that low genetic differentiation was facilitated through high levels of gene flow via the D'Urville current. Bernal-Ramírez et al. (2003) also showed that Tasman Bay was genetically isolated from other populations. This corresponds with another study that showed Tasman Bay experienced loss of genetic variation (Hauser et al., 2002). This was likely caused by high levels of genetic drift as the population size was significantly reduced as a consequence of overfishing.

1.5. Aims and thesis structure

The aims of this thesis were to *1)* assess the population genetic structure of snapper in New Zealand, *2)* reconstruct the demographic history of snapper, and finally to *3)* identify genes and their functions in the snapper genome that show signs of selection (Figure 1.7). To achieve these aims, DNA samples were collected from both contemporary and ancient material. Each data chapter is dedicated to addressing one of these questions (Chapters 2, 3, and 4). **Chapter 1** (this chapter), is an introduction to provide an overview of the framework of the research. It contains a description of the key evolutionary processes and genetic methods. **Chapter 5** presents a General Discussion of Chapters 2, 3 and 4 to summarise the results and to place the findings into a broader context.

Chapter 2 addresses how genetic variation is spatially distributed within and between modern snapper populations. It describes how contemporary samples were collected and sequenced to obtain a population-wide whole-genome data set. Population structure is inferred using both neutral and putative adaptive loci. The results are used to describe the connectivity between different areas and to describe which populations possibly constitute demographically independent units.

Chapter 3 addresses temporal demographic changes using contemporary and aDNA samples. The first objective was to test whether it is possible to extract DNA from 600-

year-old snapper bones and obtain complete mitochondrial and nuclear genomes. Combined with contemporary data, this chapter looks at what demographic events played important roles in the phylogeographic structure of snapper, and may have led to changes in population abundance through time.

Chapter 4 uses genome-wide scans to identify regions that show evidence of selection. Subsequent identification of genes within these regions was used to infer which traits could be experiencing selection. Finally, the discussion proposes some of the possible environmental factors that could be causing the observed patterns of selection.

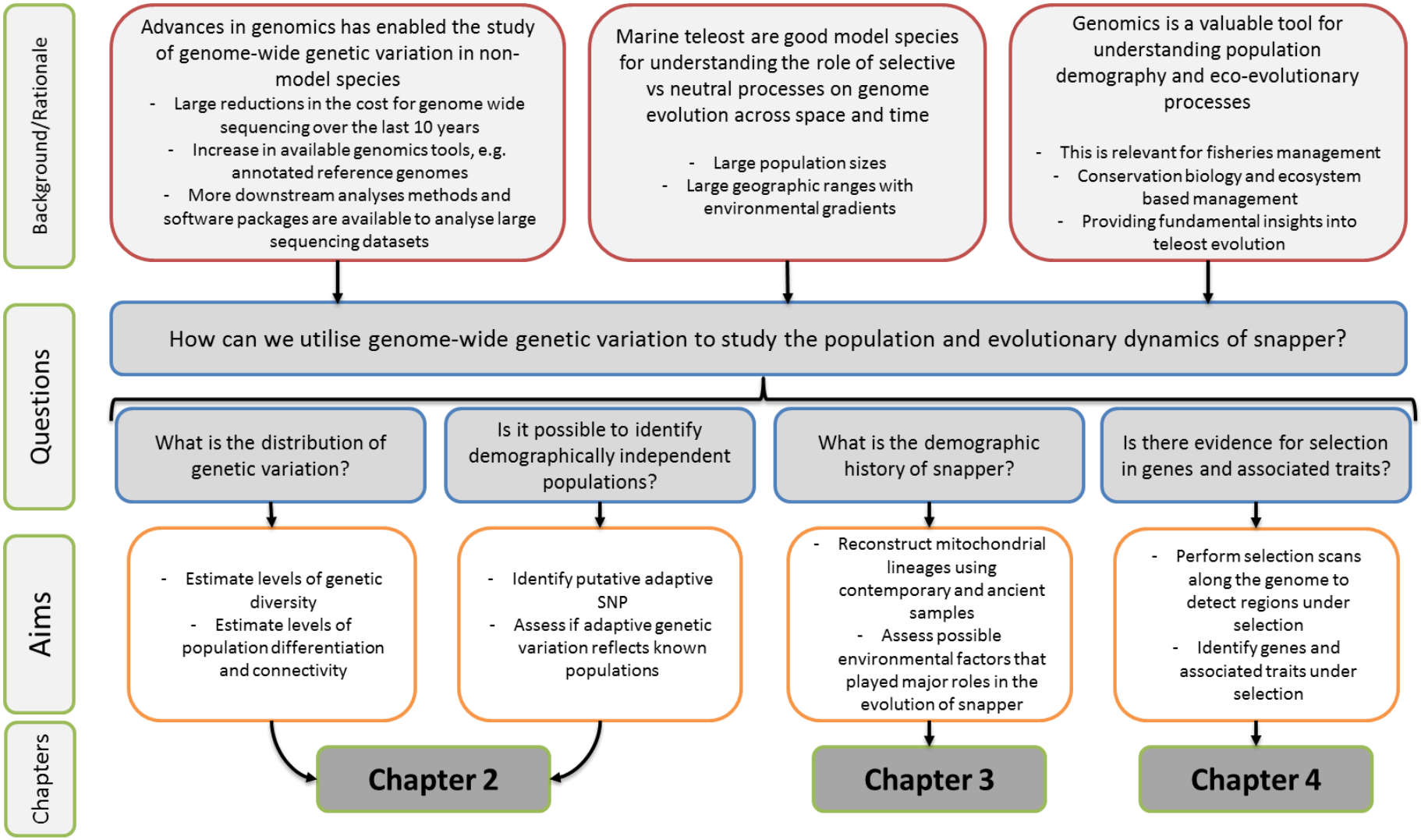

Figure 1.7: Flowchart diagram of thesis

Chapter 2. Fine-scale population structure in the Australasian snapper (*Chrysophrys auratus*) identified using whole-genome-resequencing

2.0. Abstract

The assessment of the genetic structuring of biodiversity is crucial for management and conservation. For species with large effective population sizes, a low number of markers may fail to identify population structure. The recent ability to genotype thousands of markers on a genome-wide scale adds significant power to uncover even subtle structure in populations. Here we use a dense whole-genome data set to explore the population genomic structure of the marine teleost species, the Australasian snapper (*Chrysophrys auratus*), which is characterised by large population sizes. Whole-genome resequencing data of 350 individuals across six million high-quality SNPs were analysed, which spanned the majority of the geographic range of this species in New Zealand. Population analyses using 167,543 neutrally evolving SNPs detected low but significant levels of differentiation and two distinct genetic clusters. Genetic disjunctions were observed around Cape Reinga and the Mahia peninsula, creating a northern and southern genetic cluster. Mixing of individuals from both genetic clusters was observed around Gisborne and West of Cape Reinga showing movement of individuals across genetic disjunctions. This was supported by AMOVA based on neutral loci showing that 95% of genetic variation was present within samples, suggesting large-scale homogenisation of genetic variation across snapper populations. Genetic structure also showed a significant pattern of isolation by distance, indicating that there are few barriers to gene flow, and genetic dissimilarities are driven by the physical distance between populations. More fine-scale population structure was detected using putative adaptive SNPs which possibly reflect demographically independent populations. Five distinct clusters were able to be identified in both the northern and southern cluster. However, high tolerance for false discovery rates (50%) were required to separate three locations in the northern cluster (i.e. Northland, Hauraki Gulf, and Bay of Plenty). Other genetic clusters suggest the presence of an adaptive front in Karamea Bight, and a previously unsampled population on the Kapiti

Coast. This study shows the potential for genomics to identify genetic structure in fish with low genetic differentiation between populations. This information should be incorporated into fisheries management to maintain viable fish stocks.

2.1. Introduction

The assessment of biodiversity, including the intragenic structuring of species, is one of the major goals of population management and conservation biology (Moritz, 1994b). The detection of genetic structure is crucial to identify isolated units and to assess the degree of connectivity among populations. Specifically, the identification of management units is important for the conservation of biodiversity (Palsbøll et al., 2007), including fisheries management (Allendorf et al., 2008; Papa et al., 2020). Populations genetically diverge through adaptation to local environments, genetic drift, and the introduction of new mutations to a population. Selection plays a particularly important role in species with high effective population sizes (e.g. fish) as genetic drift is relatively weak compared to species with small effective population sizes. Fish populations are difficult to monitor due to their large population sizes, high potential for dispersal, lack of obvious physical barriers in oceans, and the inaccessibility of the marine environment. Many species have been characterized by high gene flow and low levels of genetic differentiation between populations (Cano et al., 2008). This homogenisation of the genetic variation can be maintained with only a small number of migrants exchanged every generation (Palsbøll et al., 2007). When only a small number of migrants are involved in homogenising genetic differentiation, it is often inconsequential when it comes to rebuilding depleted stocks because demographic coupling requires typically more than 10% exchange between populations (Lowe & Allendorf, 2010; Waples, 1998).

Small data sets that use neutral loci have been widely used to analyse population structure, gene flow, and demographic changes over time (Allendorf et al., 2010). Traditionally, population genetic approaches used a small number of microsatellites and mitochondrial sequences to assess population genetic structure, however, the low numbers of markers often lacked the statistical power to detect low levels of genetic differentiation in high gene flow environments (Morin et al., 2009). This is a strong argument for implementing genome-wide sequencing approaches in fisheries

management to provide the level of resolution required for fisheries species and more accurately identify genetically independent stocks (Benestan, 2019; Bernatchez et al., 2017). For instance, a comparative study reported that between 100 and 200 SNPs were required to provide the same level of statistical power as 13 microsatellites for correct genetic stock identification (Hess et al., 2011). Another study found that for assignment tests, 111 microsatellites had the same statistical power as highly selected 500 SNPs (Layton et al., 2020). The authors also noted that highly adaptive SNPs were required to detect fine-scale differences between populations (Layton et al., 2020; Sylvester et al., 2018). The power of SNPs to detect fine-scale population structure was shown by Cheng et al. (2021), who reported previously undetected structure in a commercial squid (*Doryteuthis opalescens*) using 662 SNPs. The integration of genomics into fisheries management has been slow (Waples et al., 2008), despite the high potential for informing management decisions (Bernatchez et al., 2017).

Snapper (*Chrysophrys auratus*, tāmure in Māori) is a commercially, recreationally, and culturally important fisheries species found in New Zealand and Australian waters. Snapper occur around the North Island and top of the South Island of New Zealand (Figure 2.1) and is one of the largest inshore fisheries. A century of intense fishing effort has greatly reduced snapper stocks (Fisheries New Zealand, 2018), and brought some stocks below 20% of the original biomass (Figure S2.1). Previous genetic studies have identified weak but significant population structure, and possible oceanographic barriers to gene flow (Bernal-Ramírez et al., 2003; Smith et al., 1978). The most northern point of New Zealand (Cape Reinga), and Mahia Peninsula (on the East coast of the North Island) were identified as possible barriers to gene flow. Population structure was reported using nuclear markers (i.e. allozymes and microsatellites), however, mitochondrial DNA markers did not show any significant population structure. The analyses of nuclear loci concluded that the Hawkes Bay population was genetically closest to the West Coast population (Bernal-Ramírez et al., 2003; Smith et al., 1978), which was a surprising finding because these areas are not adjacent to each other. Smith et al. (1978) hypothesized that this was associated with similar water temperatures, and natural selection was acting on a genetically linked locus (esterase) (Smith, 1979). However, Bernal-Ramírez et al. (2003) showed a similar pattern using six

neutral markers, suggesting ongoing or historic gene flow between the two populations. While this was an important finding, the small number of loci used do not provide the level of support that is needed to substantiate such a finding. Lastly, the Tasman Bay snapper population was observed to be significantly differentiated from all other populations using seven microsatellite markers (Bernal-Ramírez et al., 2003; Smith et al., 1978), suggesting the Cook Strait acts as a barrier to gene flow. Based on non-genetic methods, multiple biologically separated fish stocks are recognized within these genetic clusters (Parsons et al., 2014; Walsh et al., 2011; Walsh et al., 2006). An important question is whether genomics is able to provide the resolution needed to identify these biologically separated stocks for effective genetic monitoring of snapper.

This study investigated the genetic stock structure of New Zealand snapper using 350 whole-genome sequences from samples collected from ten locations, spanning most of the geographic range of this species around New Zealand. The overall goal of this study was to establish the stock structure of snapper in New Zealand using high-resolution genomic markers, which can be used as a framework for stock assessments and better match the management stock structure to their natural reproductive stocks. To achieve this we first assessed levels of genetic diversity between ten populations. Second, a test for population genetic differentiation was conducted using both neutral SNPs and putative adaptive SNPs. Our results are compared and discussed in light of previous studies of this, and other teleost species.

2.2. Methods

2.2.1. Sampling

Fin clips were collected from commercial and recreational fishing sources (Figure 2.1 & Table 2.1). All samples were collected using fish already caught for other purposes. Sample sources including the Ministry of Primary Industries (MPI), the National Institute of Weather and Atmospheric Studies (NIWA), local fishing competitions (i.e. Wellington; Hawkes Bay fishing club - Napier; Tatapouri fishing club – Gisborne), and Fishing sheds in Auckland. Samples collected through the MPI were obtained using fin clip sampling kits specifically designed for fisheries observers stationed on commercial fishing vessels. These vessels primarily operate in management areas SNA1 and SNA8 (Figure 2.1, top left panel). GPS locations of the fishing transact were recorded to get an

estimate of the sampling location. Also, location data and fork length were recorded for each individual as part of their survey. Tissues collected through MPI would often spend an extended amount of time (up to 2 months) unrefrigerated during transport before

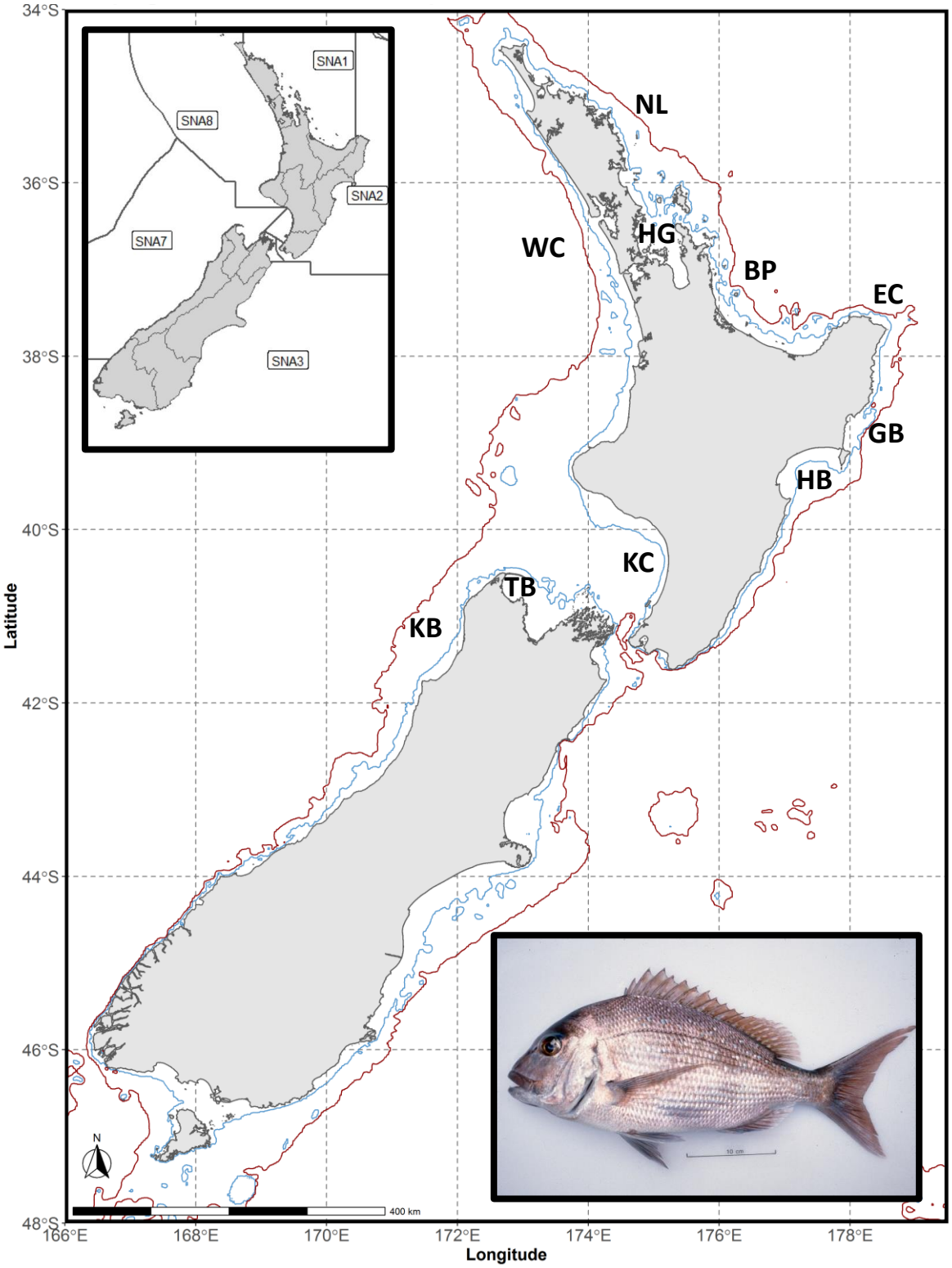

Figure 2.1: Map showing sample distribution. The blue contour line shows 50m line where snapper primary occur, and red shows 200m indicating the maximum depth where snapper is found. Top left panel shows the four different management areas in which snapper are currently managed.

being extracted or stored at -20 °C. To preserve samples under such "sub-optimal" conditions, all samples were stored in a solution of 20% DMSO, 0.25 M EDTA, Sodium Chloride saturated solution (DESS). This solution was shown to significantly outperformed another commonly used preservative: 98% Ethanol, see Oosting et al. (2020).

Table 2.1 Information Samples sites

Management area	N	Sampling location	Code	N	Source samples
SNA1	160	Northland	NL	50	MPI / Fishing sheds Auckland
		Hauraki Gulf	HG	40	Ministry of Primary Industries
		Bay of Plenty	BP	70	Ministry of Primary Industries
SNA2	55	East Cape	EC	20	Ministry of Primary Industries
		Gisborne	GB	15	Fishing competition - Gisborne
		Hawkes Bay	HB	20	Fishing competition - Napier
SNA7	52	Tasman Bay	TB	33	NIWA (WCSI 2019 survey)
		Karamea Bight	KB	19	NIWA (WCSI 2019 survey)
SNA8	83	Kapiti Coast	KC	17	Fishing competition Wellington
		West Coast	WC	66	Ministry of Primary Industries

Notes: NIWA: the National Institute of Weather and Atmospheric Studies, for a map of sampling locations see Figure 2.1

2.2.2. DNA extraction

In total, 1324 tissues were collected, and DNA was extracted from 1099 samples. DNA was isolated using a high-salt extraction protocol adapted from Aljanabi and Martinez (1997) (see Appendix 1 for protocol). Purified DNA was stored in 10 mM Tris-HCl pH 8, 0.1 mM EDTA (TE), and stored at -20 °C. A low concentration EDTA (0.1mM) was used to collate enzymatic inhibitions during library preparation. The DNA quantity of 1004 samples was determined using Qubit broad range kit (Invitrogen©), following manufactures instructions. Samples that yielded less than 20 ng/µl were not used. Impurities in the extractions were measured using an Implen© NP80 spectrophotometer. Samples with 260/280 above 2.0 and/or 260/230 reading under 2.0 were discarded. Samples that passed the quality assessments were then visualized using gel-electrophoresis to determine whether they contained high-molecular-weight DNA

(HMW, >20kb). Electrophoresis was conducted in TBE-buffered gels using 5.0 V/cm and 400 mA for 80 min and a HindIII high-molecular-weight standard was used as a reference. Samples were arbitrarily classified based on weight as either high, medium, or low molecular weight. Where possible, only high molecular weight samples were then used for sequencing. Samples that were classified as medium were only used when sample sizes for their location were low. No samples classified as low were used. The large number of extracted tissues allowed for careful selction of samples that best reflect the distribution of snapper aournd New Zealand, and research goals of this thesis.

2.2.3. Whole-genome-sequencing

Three hundred and fifty snapper were selected for whole-genome re-sequencing from the extracted samples. This included twelve individuals that had already been sequenced before this study using 100 bp paired reads at a depth of 30x. These samples were chosen to give an even representation of the known species distribution (Figure 2.1). Sample sizes from SNA1 were higher compared to other areas because this is the largest fisheries stock (Fisheries New Zealand, 2018). Library preparation and whole-genome-sequencing were performed by the Australian Genome Research Facility (AGRF). The samples were diluted to equal concentrations before shipping with a minimum of 20 ng/µl and 1500 ng total DNA per sample. Library preparation was done using the Nextera flex protocol with unique dual indexing. Sequencing was performed using the Illumina NovaSeq 6000 S4, performing 300 cycles. Each lane contained 96 individuals. Each run involved sharing with a minimum of two species (i.e. Tarakihi, *Nemadactylus macropterus*) from the projects of other researchers. Shared runs helped check whether sample labelling was correctly maintained throughout the sequencing process and de-multiplexing. Each lane was estimated to yield between 700-800 Gigabases (Gb) of data, generating ~8 Gb of data per individual, to give an average sequencing depth of ~11X per sample based on the estimated snapper genome size of 739.7 Mb (Catanach et al., 2019).

2.2.4. Read QC and alignment

Read files were analyzed in fastqc v0.11.7 (Andrews, 2010), and compiled using multiqc v1.7 (Ewels et al., 2016). Samples were aligned to both the nuclear and mitochondrial snapper reference genome in paleomix v1.2.13.3 using the bam pipeline

(Ashton, 2013; Catanach et al., 2019; Schubert et al., 2014). The 378MB reference genome for snapper consists of 24 linkage groups/pseudo-chromosomes (Catanach et al., 2019). The paleomix was designed for ancient DNA data, but also provides a highly automated approach for aligning contemporary genetic data. The pipeline uses a wide array of established programs to generate binary alignment (BAM) files. First, Adapter removal v2.2.3 was used for trimming of Illumina adapters.

Adapter1:
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG

Adapter2:
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG

Allowing a 33% mismatch rate (--mm 3) and a minimum read length of 25 bp. N's and low-quality bases were trimmed of all 5' and 3' ends.

Burrows-Wheeler Aligner (BWA) v0.7.15 (bwa-mem algorithm) was used to map reads to the reference genome. Unmapped reads and reads with a mapping quality of <25 were removed. PCR duplicates were removed using MarkDuplicates, Picard tools v2.18.20. Finally, the pipeline used the indel realigner (GATK v4.0.8.1) for local realignment. In addition to the paleomix pipeline, soft-clipped reads were removed using samtools v1.9.

2.2.5. Genotyping

Data sets based on called genotypes were generated using joint calling with GATK v4.1.4.1. (Poplin et al., 2018). First, individual g.vcf were generated from each alignment file using HaplotypeCaller with the –ERC GVCF parameter set (other settings were kept at default). Following GATK's best practices, g.vcf files were merged using GenomicsDBImport. This tool only accepts separate genomic intervals (chromosomes or linkage groups) and forces separate databases to be made for each of the 24 linkage groups. indexes were created for coordinate sorting of the vcf files --create-output-variant-index, and other settings were left on default. Genotypes were jointly called for all samples for each linkage group using GenotypeGVCFs. Variances were annotated (-G StandardAnnotation), and allele frequencies were estimated using GATK's new quality model (-new-qual). Genotyping crashed if the new quality model was not used.

Additional annotations were added using bcftools +fill-tags v1.11.2 (Li, 2011). VCF files from each linkage group were merged into a single SNP dataset using picard tools v2.18.20 MergeVcfs.

First, GATK VariantFiltration was used to set any individual genotypes called with an allelic depth of two or less to missing (i.e. no genotype called). After, vcftools v0.1.16 was used to filter sites across individuals (Danecek et al., 2011). Only biallelic sites were retained by removing all Indels (--remove-indels) and multiallelic sites (--max-alleles 2 --min-alleles 2). Sites were filtered on quality score (--minQ 600). Only sites that were present in 95% of individuals were retained (--max-missing 0.95). A minimum allele frequency of 0.01 (--maf 0.01) was applied. A mean sequencing depth between 8 (--min-meanDP 8) and 25 (--max-meanDP 25) was applied. Sites with a mean depth above 25 were removed to filter out potentially duplicated regions. Subsequently, sites showing allelic imbalance were identified using a custom script. Genotypes (GT) and allelic depth (AD) were extracted for each site and used to apply a two-sided binomial test on heterozygote sites. Sites were tested for a significant deviation from equal proportions between the reference and alternative alleles (p=0.5). Sites identified to be subjected to allelic imbalance were filtered out using vcftools v0.1.16 (--exclude-positions).

The data was filtered for sites deviating from Hardy-Weinberg equilibrium (--hwe 0.05), and thinned down to 1 SNP every 1,500 bp (--thin 1500), and finally pruned using Plink (--indep-pairwise 50 5 0.2). Sites under selection were filtered out with Bayescan v2.0, using outputting 10,000 iterations with a thinning interval of 10. The prior odd was set to 10,000 to minimize false positives (Benestan et al., 2015; Lotterhos & Whitlock, 2014). Sites were again excluded using vcftools. VCF files from each linkage group were merged into a single SNP dataset using picard tools v2.18.20 MergeVcfs.

Outlier loci were identified using the R packages outflank (Whitlock & Lotterhos, 2015). Detection for outliers was done three times, using subsampling of detected clusters to identify SNPs with more fine-scale population structure. Outlier detection including all ten sampling locations was performed using a false discovery threshold of 1% (qthreshold 0.01), minimum heterozygosity of 0.1 (Hmin 0.1), and left and right trim fraction of 0.35 and 0.1 respectively (LeftTrimFraction 0.35, RightTrimFraction 0.1).

Identified outliers were thinned by selecting the SNP with the highest F_{ST} in 500,000 bp non-overlapping sliding windows. Higher false discovery thresholds were accepted as more fine-scale population differences were assessed with smaller sample sizes (Table 2.2)

Table 2.2: Information regarding outlier analyses for identifying putative adaptive SNPs

Location	N locs	qthreshold	N outliers	PC1 %	PC2 %	Plot panel
All	10	0.01	134	7.00	1.34	A
North	5	0.1	30	11.08	5.36	B
South	5	0.1	53	9.25	8.95	C
SNA1	3	0.5	40	10.01	7.13	D
SNA7 & 8	2	0.01	120	17.05	3.49	E

Notes: plot panel refers to PCA plots shown in Figure 2.6.

2.2.6. Data analyses

An ANOVA was performed to check for significant differences in heterozygosity between populations, to test for local loss of genetic variation. Principle component analyses (PCA) were performed using SNPRelate (Zheng et al., 2012), and visualized using a custom script obtained from Clucas et al. (2019), and Discriminant analysis of principal components (DAPC) were performed using adegenet (Jombart, 2008). Subsequently, clusters within both plots were analysed separately to detect more refined substructuring. Population structure was assessed using faststructure (Raj et al., 2014). The number of populations (K) tested ranged between 1 and 10, using the simple prior parameter and a tolerance threshold of $10e^{-6}$ (default). The number of populations that best explain the structure in the data set was evaluated using chooseK (distributed with faststructure).

Population differentiation (weighted F_{ST}) was estimated for each SNP using snpgdsFST from the R package SNPrelate, using the Weir & Cockerham, 1984 method (Weir & Cockerham, 1984; Zheng et al., 2012). Then, 95% confidence intervals were obtained by bootstrapping the values 10,000 times. A heatmap was used to visualize the F_{ST} estimates using the R-package pheatmap (Kolde & Kolde, 2015). The package was also used to run *hclust* to generate a dendogram based on the F_{ST} values. A hierarchical analysis of molecular variance (AMOVA) based on both genetic clustering and sampling region was performed using the R-package poppr (Kamvar et al., 2014). Here, an

hierarchical model was applied that compared the variation between management areas and sampled regions (poppr.amova(data,~management area/ sampling location)).

Significance levels were tested using randtest from R-package Adegenet, using 999 permutations. A Mantel test was performed to detect a possible significant correlation between genetic and geographic distances using the R-package ade4 (Thioulouse et al., 2018). Distances between sample locations were approximated by calculating the least-cost distance using the R-package gdistance (van Etten, 2017). Locations were determined by taking the mean longitudes and latitudes of all samples for each location. In locations where GPS data was not available, an arbitrary point representing the area was chosen (i.e. Gisborne and Hawkes Bay).

2.3. Results

2.3.1. Data quality and filtering

FastQC results from 350 individuals showed high-quality reads (>30 Phred Score) for all individuals (Figure S2.3). All samples that were prepared using the Nextera flex protocol were flagged for their per base sequence content. This was caused by an uneven distribution of nucleotides in the first ~15 bases of all reads (Figure S2.4), which is a bias induced by non-random priming of hexamers (Hansen et al., 2010). Reads did not have to be trimmed as the sequences still represent accurate genomic sequences.

In total, 13.6 billion reads were generated and ranged between 22.7 million and 372.9 million per sample (Figure S2.5A). This provided an average sequencing depth between 6.3x and 60.1x per individual (Figure S2.5B&C). Twelve additional samples with an average coverage ranging from 29x to 36x were always available and added to the data set. One individual (i.e. CA119052) had an average coverage of 60.1 x. In total, 17 (4.8%) samples had an average sequencing depth below 8, which was set as the lower boundary for SNP filtering.

After filtering, 12.9 million polymorphic sites were obtained with an average depth 10.2 x (Table 2.3 & Figure S2.6). After initial quality filtering, six million high-quality biallelic SNPs were retained, and 1.5 million sites were removed due to allelic imbalance. The 13 individuals that were sequenced to 30x showed higher levels of heterozygosity (Figure S2.6). Subsampling of reads from these individuals showed levels

of heterozygosity in all other samples was consistently underestimated (Appendix 2). Notably, one individual with an average sequencing depth of ~10x was observed to have a heterozygosity level of 0.28, possibly because the sample contained contamination. One individual had a lower level of heterozygosity (0.18), and a high level of missing data (3%) when compared with other samples (Figure S2.6C). Data pruning was conducted to ensure only independent segregating sites were used which removed an additional 4.4 million sites. Finally, Bayescan identified 45 SNPs potentially influenced by selection and were removed from the data set. The final data set comprised 350 individuals, and consisted of 167,543 independently segregating, neutrally evolving bi-allelic SNPs. See appendix 3 for a detailed discussion on SNP filtering.

Table 2.3: SNP filtering

Filter	SNPs removed	SNPs retained
Raw data	-	12,941,093
Quality control	6,906,211	6,034,882
Allelic imbalance	1,450,440	4,584,442
Pruning	4,416,854	167,588
Sites under selection	45	167,543

Notes: SNPs removed indicate the number of SNPs filtered out in the respective filtering step, SNPs retained are the number of SNPs left in the data set after filtering.

2.3.2. Genetic diversity

Levels of heterozygosity were similar among populations, with an average of 0.210 (0.206-0.215) after quality filtering, and 0.196 (0.192-0.199) after filtering for neutrality (Table 2.4). There was no significant difference in heterozygosity between sampled populations (p = 0.069) and genetic clusters (p = 0.521) when using the quality controlled data set, or neutral data set (0.061 and 0.393, respectively). Karamea Bight displayed the lowest levels of genetic variation, followed by Gisborne.

Table 2.4: Sample collection location and estimates of genetic diversity

Population	Code	Lat	Long	N_{ind}	H_{QC}	$H_{neutral}$
Northland	NL	-36.007	174.778	50	0.213	0.199
Hauraki Gulf	HG	-36.666	175.254	40	0.211	0.197
Bay of Plenty	BP	-37.276	176.369	70	0.209	0.196
East Cape	EC	-37.550	178.516	20	0.210	0.197
Gisborne	GB	-38.699	178.126	15	0.205	0.193
Hawkes Bay	HB	-39.528	177.240	20	0.209	0.195
Tasman Bay	TB	-40.834	173.156	33	0.209	0.196
Karamea Bight	KB	-41.163	172.015	19	0.205	0.192
Kapiti Coast	KC	-40.833	174.984	17	0.210	0.197
West Coast	WC	-36.755	173.938	66	0.210	0.196
All	-	-	-	350	0.210	0.196

Notes: Lath = Latitude, Long = Longitude, N_{ind} = number of individuals, H_{QC} = observed heterozygosity based on sites after quality control, $H_{neutral}$ = observed heterozygosity based on neutral data set.

The AMOVA based on neutral SNPs showed populations to be significantly differentiated (Table 2.5), both between management areas (p = 0.008), and sample populations within management areas (p = 0.001). However, nearly 95% of the observed variation was observed within samples. When using putative adaptive loci, the AMOVA showed that only 77% of genetic variation was present within samples (p = 0.001), and 12.25% of the variation is present between management areas (p = 0.008), 2.13% between sample locations within management areas (p 0.001), and 8.37% between samples within sample locations (p = 0.001) (Table 2.6).

Table 2.5: AMOVA among 10 populations using 167,543 neutral SNPs

Source of Variation	d.f.	Sum of squares	Variance components	Percentage of variation	ϕ	*p*
Between management areas	1	33,304	43.0	0.25	0.0504	0.008
Between sample locations within management areas	8	146,924	3.2	0.02	0.0478	0.001
Between samples within sample locations	340	6,172,864	829.6	4.77	0.0002	0.001
Within samples	350	5,773,669	16496.2	94.95	0.0025	0.001
total	699	12,126,761	17372.1	100.00		

Notes: d.f. = degrees of freedom, ϕ i = phi statistics, *p* = p-value

Table 2.6: AMOVA among 10 populations using 134 adaptive SNPs

Source of Variation	d.f.	Sum of squares	Variance components	Percentage of variation	ϕ	*p*
Between management areas	1	1264.3	3.42	12.25	0.2275	0.008
Between sample locations within management areas	8	519.5	0.59	2.13	0.0977	0.001
Between samples within sample locations	340	8943.2	2.34	8.37	0.0243	0.001
Within samples	350	7566.5	21.61	77.25	0.1224	0.001
total	699	18293.6	27.98	100.00		

Notes: d.f. = degrees of freedom, ϕ = phi statistics, *p* = p-value

2.3.3. Population differentiation – neutral loci

Both the PCA and DAPC analyses showed a split of the data into two clusters (Figure 2.2 & Figure 2.3). The first principle component (PC1) shows a separation of the sample locations into two clusters, explaining 0.5% of the total genetic variation. Scree plot of the discriminant analyses also shows that the first component primarily contributes to the observed structure. Multiple individuals that grouped within one of the two clusters in the PCA/DAPC were geographically related to the other group, showing the movement of individuals across the genetic boundaries. For example, individuals sampled from the West Coast area are grouped within the cluster containing samples collected from northern areas. Faststructure did not show any significant groupings, suggesting that the number of populations that maximizes both the marginal likelihood and best explains the structure was one.

All sample locations were found to be genetically differentiated (Figure 2.4). The average pairwise F_{ST} estimate was 0.0023 and ranged between 0.0003 and 0.0047. Within the clusters identified by the PCA/DAPC genetic differentiation between sampled sites was low (0.0003 - 0.0017), except for Gisborne. Gisborne grouped with the northern cluster, the differentiation between the southern populations is much lower. The three most southern populations (i.e. KC, KB, and TB) were shown to be most significantly isolated from the northern cluster. Notable, the populations Hawkes Bay and West Coast are closely related while being located on separate sides of the North Island, and having the Kapiti Coast population in between.

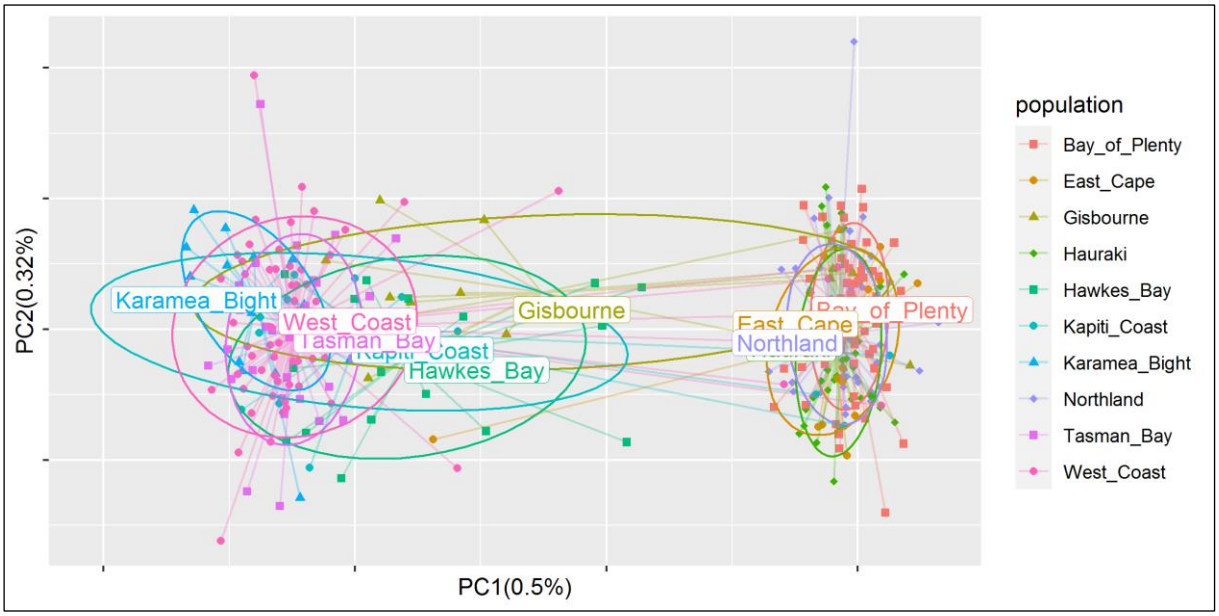

Figure 2.2: Principle component analyses showing the genetic clustering of individuals grouped per samples locations. The percentage shown on each axis indicate what proportion of the total amount of variation is explained by that particular principal component.

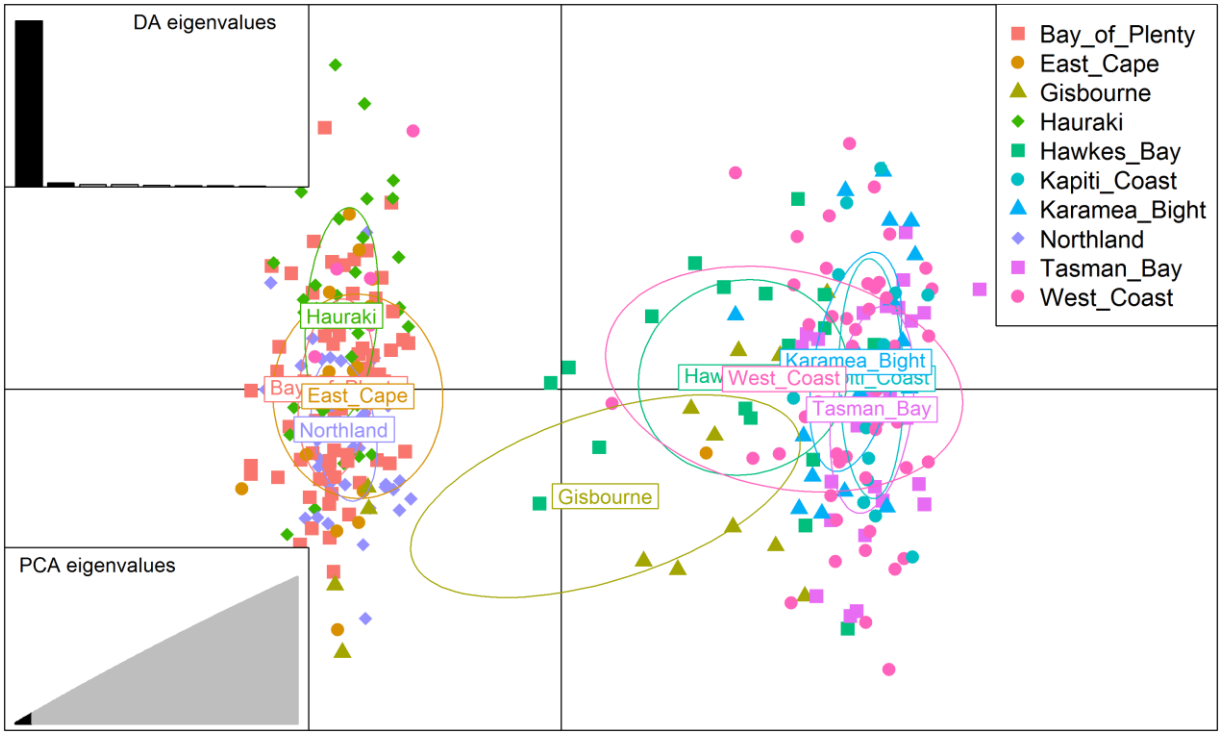

Figure 2.3: Discriminant analyses of principle components (DAPC). Top-left plot: representation of the relative differences of the first 9 eigenvalues of the discriminant analyses. Bottom-left plot: representation of the relative differences of the first 20 eigenvalues of the principle component analyses.

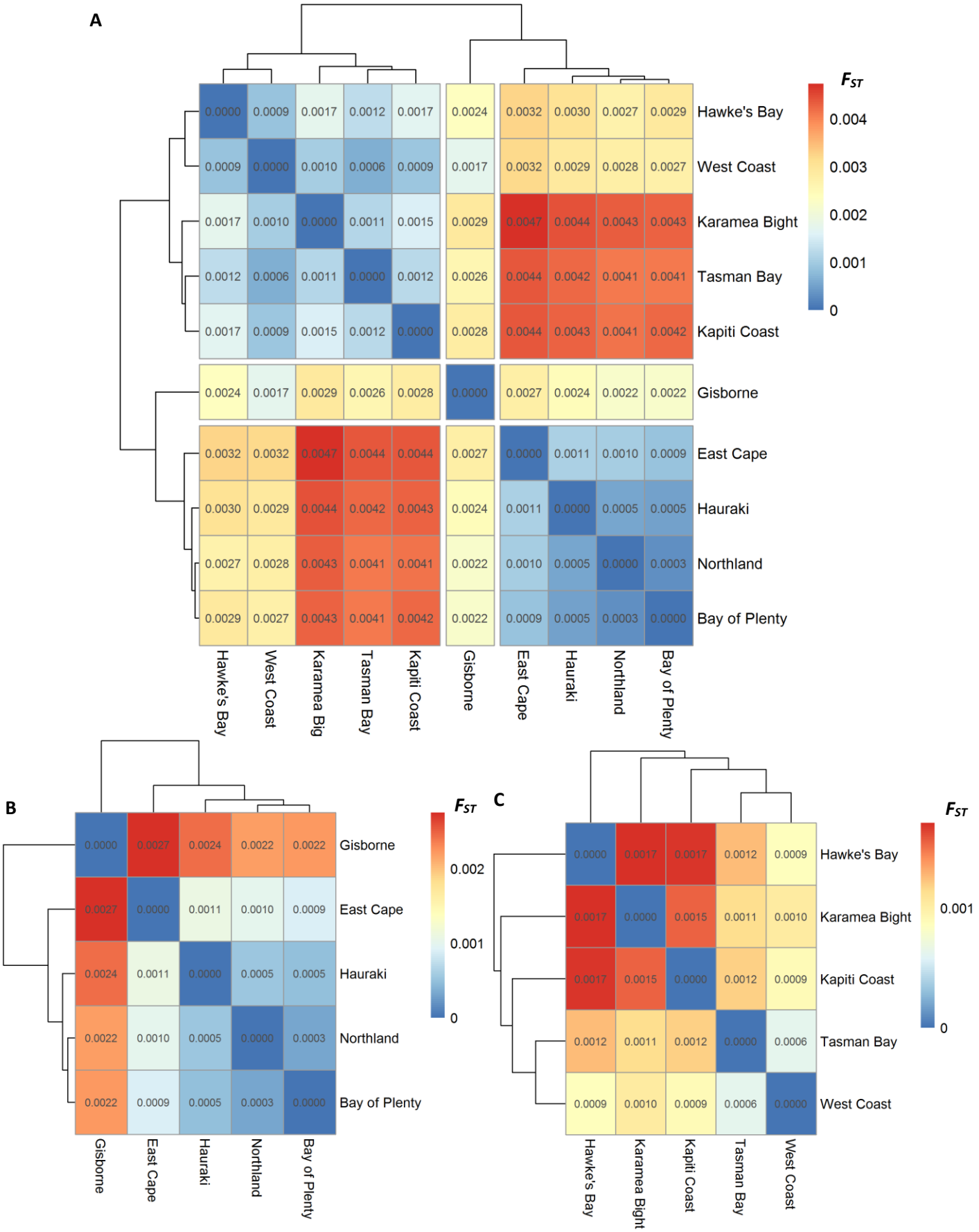

Figure 2.4: Heatmap showing pairwise weighted F_{ST} estimates for samples populations. The dendogram shows the relationship between samples populations. Values above and below the diagonal show pairwise FST estimates. Heatmaps of A) all sampled sites, B) northern cluster, C) Southern cluster.

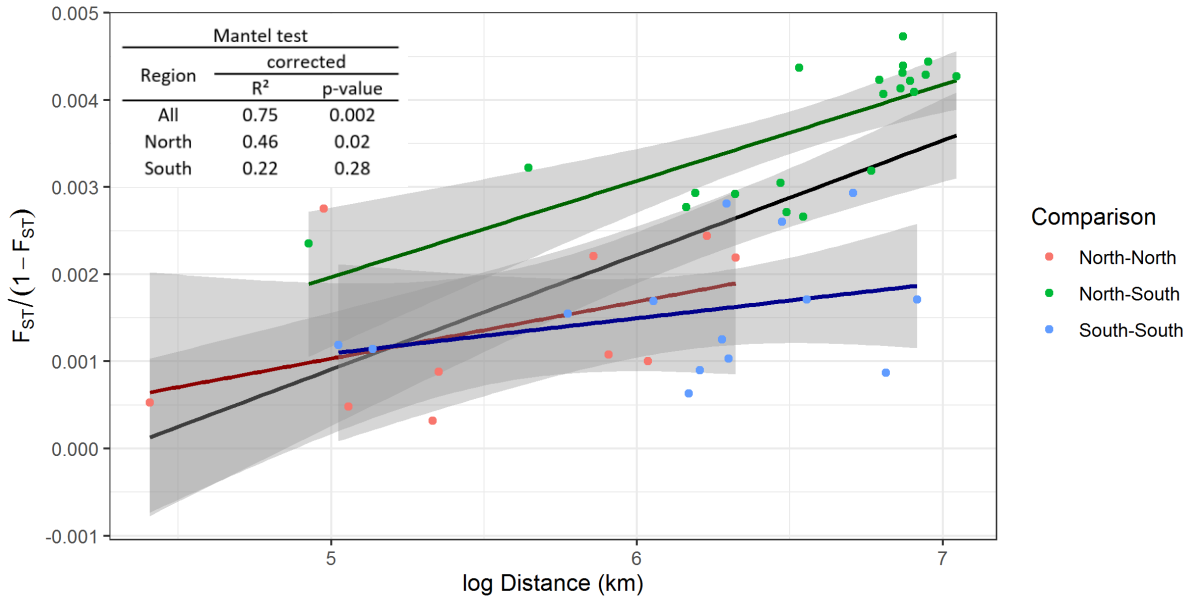

Figure 2.5: Genetic distance (weighted F_{ST}) between populations plotted against geographic distance. The lines shows the linear regression between points. Colours indicate comparisons made between populations from the same cluster, or between different clusters. The black line shows the regression line between all points.

A significant correlation was observed between genetic and geographic distance (R^2 = 0.75, p-value = 0.002) (Figure 2.5). This significant correlation was not present in southern cluster (R^2 = 0.22, p-value = 0.28), and weakly significant in the northern cluster (R^2 = 0.46, p-value = 0.02) (Figure 2.5).

2.3.4. Population differentiation – adaptive loci

The detection of outlier loci using all sampling locations identified 134 potential adaptive SNPs, showing the same North-South population structuring as detected when using neutral loci (Figure 2.2 & Figure 2.6A). The second principal component (PC2) showed a clear separation of the Kapiti Coast sampling location from the rest of the southern cluster. Subsequent outlier detection between sampling locations that were overlapping showed a more fine-scale population structure (Figure 2.6B-E). The identification of 104 potential outlier loci in the northern cluster detected a more fine-scale population structure separating both East Cape and Gisborne from the three sample locations in SNA1 (i.e. Northland, Hauraki, and Bay of Plenty) (Figure 2.6B). Outlier detection between the three SNA1 locations identified 40 SNPs (Figure 2.6D). Fine-scale population structure in the southern cluster was detected using 72 putative outlier SNPs, separating Hawke's Bay, Kapiti Coast, and Karamea Bight into distinct genetic clusters (Figure 2.6C). Finally, the identification of 120 putative outlier loci

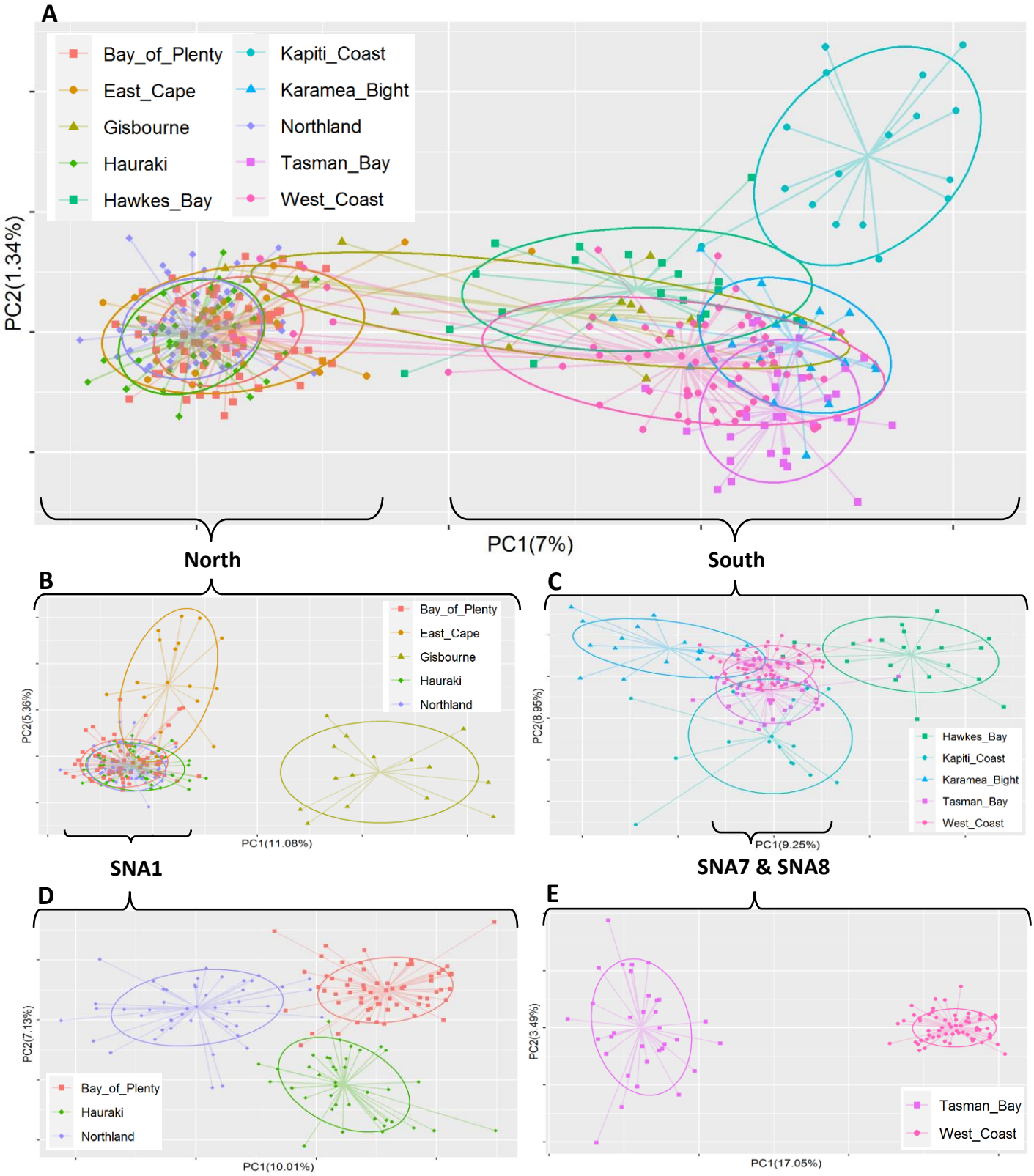

Figure 2.6: PCA showing fine-scale population differentiation between sampling location using outlier loci. A) All sampling locations, qval = 0.01 and 134 SNPs, B) northern genetic cluster, qval = 0.1 and 104 SNPs, C) Southern genetic cluster, qval = 0.1 and 72 SNPs, D) management area SNA. qval = 0.5 and 40 SNPs 1, E) genetically similar sampling areas from management areas SNA7 and SNA8, qval = 0.01 and 120 SNPs.

between Tasman Bay and the West Coast separated the two sampling locations (Figure 2.6E).

2.4. Discussion

This study assessed the population genetic structure of snapper around New Zealand to quantify genetic variation, how it partitions across the sampling range and to identify genetically distinct groups. Using the complete SNP dataset, the results showed that snapper consists of two distinct genetic clusters around New Zealand based on the neutral markers. Outlier analyses indicated that structure is more refined when applying putative adaptive SNPs, which possibly reflect adaptive differences between demographically independent stocks. Sampled locations displayed low but significant population differentiation, something that is commonly reported for marine species as a result of their large effective population sizes. While previous studies had reported evidence of similar patterns (Bernal-Ramírez et al., 2003; Smith, 1979), the higher statistical resolution provided by genome-wide markers substantiates the previous findings of genetic differentiation, and indicates new clusters that were previously not identified. The detection of isolation by distance is also a new insight and suggests that there are few barriers to gene flow, and genetic dissimilarities are driven by the physical distance between populations. The discussion below will focus on what is driving the observed genetic clusters and the implications for fisheries management.

2.4.1 Genetic variation

The average heterozygosity in snapper was 0.210 (0.206-0.215) after quality filtering, and 0.196 (0.192-0.199) after filtering for neutrality (Table 2.4). The difference between the two estimates suggests that there is less variation present in neutrally evolving loci compared to the complete pool of genetic variants. This is likely caused by the large proportion of heterozygote genotypes that were removed due to allelic imbalance (~ 24% of high-quality SNPs). This study found no significant differences in genetic variation between populations (Table 2.4), suggesting no significant loss of genetic variation in a single population. It is important to note that nationwide reductions in genetic diversity could not be detected this way. A previous study reported that the Tasman Bay population experienced a significant loss of genetic variation (Hauser et al., 2002). The loss of genetic variation was thought to be caused by increased genetic drift in a population that suffered population declines, possibly due to fishing. Unfortunately, estimates of genetic variation were incomparable between studies due

to the choice of genetic marker. Compared to genomic data, microsatellites represent only a fraction of the genetic variation, assumed to provide a good representation of the genetic variation within a population. Microsatellites are often selected based on variability, selecting markers with many alleles, which can mean estimates of genetic variability can be inflated (Brandström & Ellegren, 2008; Väli et al., 2008). Previous studies on snapper using microsatellites reported levels of heterozygosity between 0.521 - 0.938 (Hauser et al., 2002), 0.650 – 0.743 (Ashton, 2013), 0.54 - 0.78 (Le Port et al., 2017), while heterozygosity of the neutral SNP data was on average 0.196 (Table 2.3). The high variability could make microsatellites a more sensitive marker to reductions in genetic variation, making them a good marker for early detection of genetic erosion within a population. However, genome-wide estimates of genetic diversity might be more relevant for fisheries management. It is also worth noting that genetic variability can be maintained in heavily exploited populations (Therkildsen et al., 2010). For example, there was no evidence that genetic variation was lost in populations of the heavily exploited North Atlantic cod (Pinsky et al., 2021).

2.4.2. Population structure using neutral SNPs

The AMOVA showed most of the neutral genetic diversity is observed within individuals (94.95%) (Table 2.5), suggesting high levels of gene flow. The AMOVA based on putative adaptive SNPs showed that only 77.25% of genetic variation is present within individuals (Table 2.6). This suggests that while neutral variation is exchanged between stocks, adaptive variation is resisting the homogenization of genetic variation due to local adaptation. Faststructure failed to detect separate populations. This approach for detecting population structure is known to underperform in situations where there are low levels of population differentiation (i.e. large population sizes and a high number of migrants between populations) (Waples & Gaggiotti, 2006). The Mantel test suggests genetic differentiation is correlated with geographic distance (r^2 = 0.75, p-value = 0.002). However, this significant correlation disappeared within the southern region (r^2 = 0.22, p-value = 0.28), which is likely due to the reduction in statistical power and high genetic similarity between the West Coast and Hawke's Bay.

Two distinct genetic clusters were identified based on the PCA and DAPC analyses (Figure 2.2 & Figure 2.3). Genetic disjunctions appear to be present at the top

of North Island (Cape Reinga) and East Coast of the North Island (Mahia peninsula) (Figure 2.1). The northern Western genetic disjunction could be caused by a lack of migration of fish across the current of the Tasman Front as it splits down either side of Cape Reinga (Bull et al., 2018; Papa et al., 2020), promoting southward migration along both sides of the North Island. The South Eastern genetic disjunction is located around the Mahia peninsula. Here, individuals of both genetic stocks are observed around Gisborne with no evidence of gene flow between the two genetic stocks. Gene flow between both genetic clusters is likely occurring in Hawke's Bay. This fits with the observation that there is a known breeding ground of snapper in Hawke's Bay, and not in Gisborne (Parsons et al., 2014).

The level of genetic differentiation found between populations was low (Figure 2.4), but similar to other fish with large population sizes including Atlantic cod (*Gadus morhua*) (Knutsen et al., 2011), blue whiting (*Micromesistius australis*) (McKeown et al., 2017), and herring (*Clupea harengus L.*) (Jørgensen et al., 2005). The heatmap showing pairwise F_{ST} estimates between sampling locations showed that West Coast and Hawke's Bay individuals are genetically similar (Figure 2.4). The standing hypothesis was that there is high connectivity between the locations via the D'Urville and Wairarapa Coast current (Bernal-Ramírez et al., 2003). This high level of connectivity between the two populations on opposite sides of the North Island is striking. Especially since the Kapiti Coast population, which is situated between the two locations does not share the same level of genetic similarity (Figure 2.4). One possible explanation is that local adaptation is preventing Kapiti Coast individuals from mixing with individuals from the other two locations. The colder waters surrounding the Cook Strait (including Kapiti Coast) could result in unfavourable reproductive conditions for individuals from both the West Coast and Hawke's Bay that are better adapted to spawning in warmer waters. Ocean temperature is a potential selective driver (Donelson et al., 2010), reducing the reproductive output of migrant snapper from the North.

Another explanation for the genetic similarity between the West Coast and Hawke's Bay is caused by gene flow within the same genetic cluster (i.e. northern cluster) (Figure 2.7). When the northern and southern clusters are separated, pairwise F_{ST} estimates no longer show clustering of the West Coast with Hawke's Bay (Figure

2.4B & C). This suggests that the two populations were grouped because they share genetic similarities with the northern cluster. Individuals that are genetically more similar to the northern cluster were sampled on the West Coast (southern cluster) (Figure 2.2 & Figure 2.3), showing movement around Cape Reinga (Figure 2.7). Similarly, the intermediate placing of Hawke's Bay individuals between the two genetic clusters in both the PCA and DAPC suggests this is where the genetic mixing occurs (Figure 2.2 & Figure 2.3). These observations support the hypothesis that the genetic similarity is caused by gene flow within the same external genetic cluster.

2.4.3. Fine-scale population structure using putative adaptive SNPs

Fine-scale genetic structure was detected through the identification of putative adaptive loci through outlier analyses (Figure 2.6). PCA plots showing clustering of all sampling locations were similar between neutral and putative adaptive loci. (Figure 2.2 & Figure 2.6*Figure 2.6*A). Within the southern cluster, Kapiti Coast separated from the other locations on the second axis (PC2). Previous studies have suggested the presence of an additional stock present in the Marlborough Sounds (Figure 2.7) (Parsons et al., 2014). Kapiti Coast, which is geographically close to the Marlborough Sounds, could be genetically related to this so far unsampled stock. Sequencing of individuals from Marlborough Sounds would be required to test this hypothesis. More fine-scale population structure is detected when tests for outlier loci were performed separately for the genetic clusters.

Fine-scale genetic structure in the northern cluster is primarily explained by geographic distance. Gisborne, which is geographically most isolated, shows the strongest segregation from the other locations (Figure 2.6B). East Cape is also segregating from the three populations in management area SNA1 (Figure 2.1). A high acceptance tolerance for false discovery rates was required to identify enough putative outlier loci that could separate sampling locations (10%). This is because low genetic differences and smaller sample sizes reduce the statistical power to confidently identify outliers. Using stricter boundaries for false discovery rates showed a similar yet less distinct pattern (Figure S2.7). East Cape is currently located in the same management areas as Gisborne and Hawke's Bay (SNA2), but shows more genetic similarity with management area SNA1. The three locations in SNA1 were genetically very similar

(Figure 2.6D). The presence of multiple stocks that reflect the sampled locations have been suggested based on year-class strength and growth rates (Parsons et al., 2014; Walsh, 2003; Walsh et al., 2011), and corresponds with the clustering that is observed using putative outlier loci. However, putative adaptive loci between these areas are

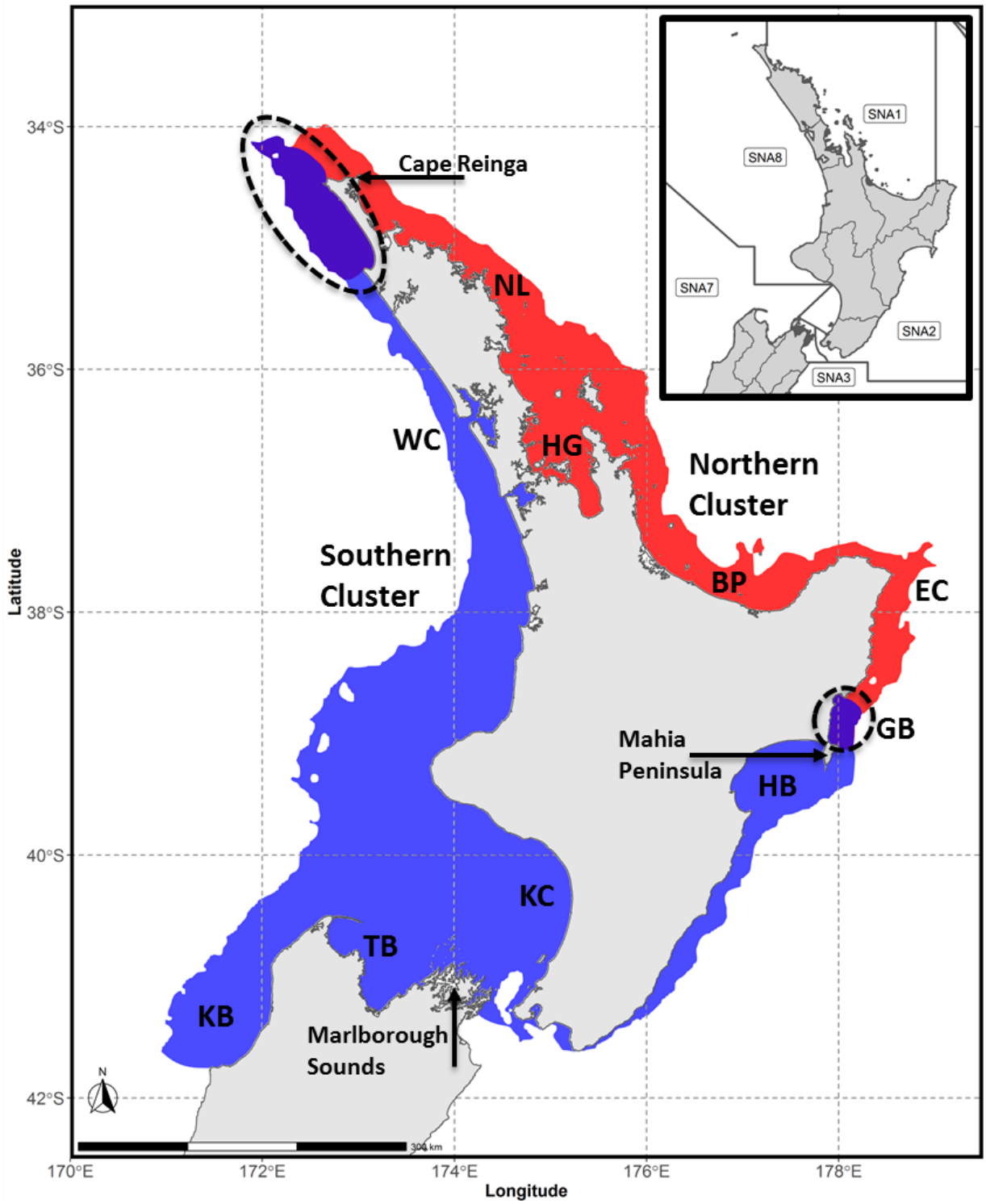

Figure 2.7: Spatial distribution of the North and South genetic cluster, including two areas of mixing (purple areas) that provide a possible explanation for the close genetic similarity between the West Coast (WC) and Hawke's Bay (HB). The boundaries of each cluster indicate the 200 meter bathymetry which considered the maximum depth at which snapper occurs. Top-right panel shows the boundaries of the snapper management areas.

highly speculative and should be interpreted with extreme care. The analyses required a false discovery rate boundary of 50% before enough divergent SNPs could be identified to observe genetic clustering (Table 2.2). Random assignment of samples to sampling locations and re-identifying putative outlier loci resulted in near-identical clustering (Figure S2.8).

In the southern cluster, Kapiti Coast, Karamea Bight, and Hawke's Bay started separating into separate clusters (Figure 2.6C). The separation of Kapiti Coast from the other locations remains consistent with the hypotheses that Kapiti Coast is possibly related to an unsampled population in the Marlborough Sounds (Parsons et al., 2014). Also here, a lower tolerance for false discovery rates showed similar yet less distinct clustering (Figure S2.9). Hawke's Bay has a known spawning ground that likely experiences gene flow with the northern cluster, causing this location to be genetically distinct from the other population in the southern cluster. Karamea Bight individuals are considered to be at the southern boundary of their range, which is determined by sea surface temperature (Brooks, 2019). Karamea Bight is unlikely to be demographically independent because there are no known spawning grounds in this area. Natural selection on traits related to adaptation to colder waters could drive genetic differentiation between the other populations (see Chapter 5). Outlier loci between Tasman Bay and the West Coast were identified with high confidence using a false discovery rate of 1% (Figure 2.6E). Interestingly, these locations appear to be genetically the most similar in the southern cluster while they are located in separate management areas (SNA7 and SNA8 respectively), and recognized as demographically independent stocks (Fisheries New Zealand, 2018). The clustering is consistent pairwise F_{ST} estimates based on neutral loci (Figure 2.4), suggesting gene flow is present between the two stocks.

2.4.4. Implications for fisheries management

Sustainable fisheries management includes the preservation of genetic variation (Bernatchez et al., 2017). A previous study has reported the loss of genetic variation in snapper (Hauser et al., 2002). Here, the lack of significant differences in heterozygosity among populations suggests no loss of variation in any particular population. Hauser et al. (2002) detected the loss of genetic variation using microsatellites which tend to be

selected based on the number of alleles. Genetic markers with a high number of alleles may be more sensitive to reductions in variation, allowing loss of variation to be detected sooner (Brandström & Ellegren, 2008). An important question is what this implies for the genetic viability of the population. Based on the results of this study, Tasman Bay has comparable levels of genetic diversity compared to other locations. Early detection of loss of genetic variation is important because it provides an opportunity to act proactively, and prevent the loss functional genetic variation. Genomic tools will provide an important role in future fisheries management as it provides a higher resolution of the genetic variation in fished stocks (Bernatchez et al., 2017). It is also showing promising results at being able to identify fine-scale population structure in species with low levels of genetic differentiation.

Fish have a high potential for dispersal, resulting in homogenization of genetic variation over large distances (Figure 2.4). However, demographically independent stocks are thought to be present within these areas based on year-class strength and growth rates (Parsons et al., 2014; Walsh, 2003; Walsh et al., 2011). The use of adaptive loci provided a powerful tool for detecting structure when genetic differentiation is low (Vaux et al., 2021). It is unclear whether these populations are demographically independent. Our results show that a small number of genetic loci are resisting the homogenizing effect of gene flow, allowing us to differentiate between individuals from different regions (Figure 2.7). This provides fisheries management with the tool to genetically differentiate between previously identified stocks. It does not show whether gene flow between these regions is low enough for demographic independence to be established. Neutral genetic data showed that the level of connectivity between recognized stocks is high enough for genetic differentiation to be limited (Figure 2.4). However, homogenization of genetic variation occurs at much lower rates of dispersal compared to demographic independence (Lowe et al., 2017). Additional information regarding migration rates could help answers this question in the future (Allendorf et al., 2010; Manel et al., 2005). Genomics should play an important role in fisheries management. However, it's is important to recognize that genomics is one tool and should be combined with other monitoring methods. So far, genomics is not able to provide information on metrics such as year-class strength and growth rates (Parsons

et al., 2014; Walsh, 2003; Walsh et al., 2011), which are vital to monitoring the health of fished species.

2.4.5. Future research and conclusion

The identification of structural variation in snapper may provide additional resolution and statistical power to detect genetic differences between different stocks. In other species, structural variants have provided much more detailed population differentiation compared to SNPs (Dorant et al., 2020). The amount of structural variation in snapper is known to outweigh the number of SNPs threefold (Catanach et al., 2019). Structural variation also plays an important function in adaptation to local environments (Hoban et al., 2016), both copy number variation (Dorant et al., 2020; Perry et al., 2007), and inversions (Berg et al., 2016; Star et al., 2017). In snapper, structural variants may help provide more statistical power to delineate boundaries between reproductively isolated stocks. For example, the identification of separate stocks in management area SNA1 where genetic differentiation between proposed reproductively isolated stocks is low. These structural variants may also provide information on what traits are being selected for.

There other analyses that may provide valuable tools for monitoring stock sizes and movement between stocks. For example, close-kin-mark-recapture which uses the level of relatedness between individuals to estimated population sizes (Bravington et al., 2016), mixed-stock-analyses which tests for the presence of multiple stocks that overlap in their spatial distribution (McKinney et al., 2017), and assignment tests which genetically determines the source population of an individual (Berry et al., 2004). While these approaches require large sample sizes, whole-genome data sets can be used to develop SNP panels that would provide relative low-cost genotyping of a large number of individuals. These methods could potentially replace industry-standard mark-recapture methods that can be heavily influenced by the survival rate of tagged individuals or re-catch rates.

This study showed the power of genomics for the detection of fine-scale population structure, and its potential for fisheries management. Neutral loci helped reinforce previous observations regarding genetic disjunctions around Cape Reinga and

Mahia peninsula. Low levels of genetic differentiation between the West Coast and Tasman Bay suggests that Tasman Bay is not as genetically isolated as previously suggested. It provided new insights that help explain why populations on either side of New Zealand show low levels of genetic differentiation. The identification of putative adaptive SNPs between sampling locations was able to detect genetically distinct clusters that correlate with previously proposed demographically independent stocks. These results showed that genomics could provide value information for monitoring one of New Zealand's largest fisheries.

2.6. Supplementary information chapter 2

2.6.1. Supplementary Tables

2.6.2. Supplementary Figures

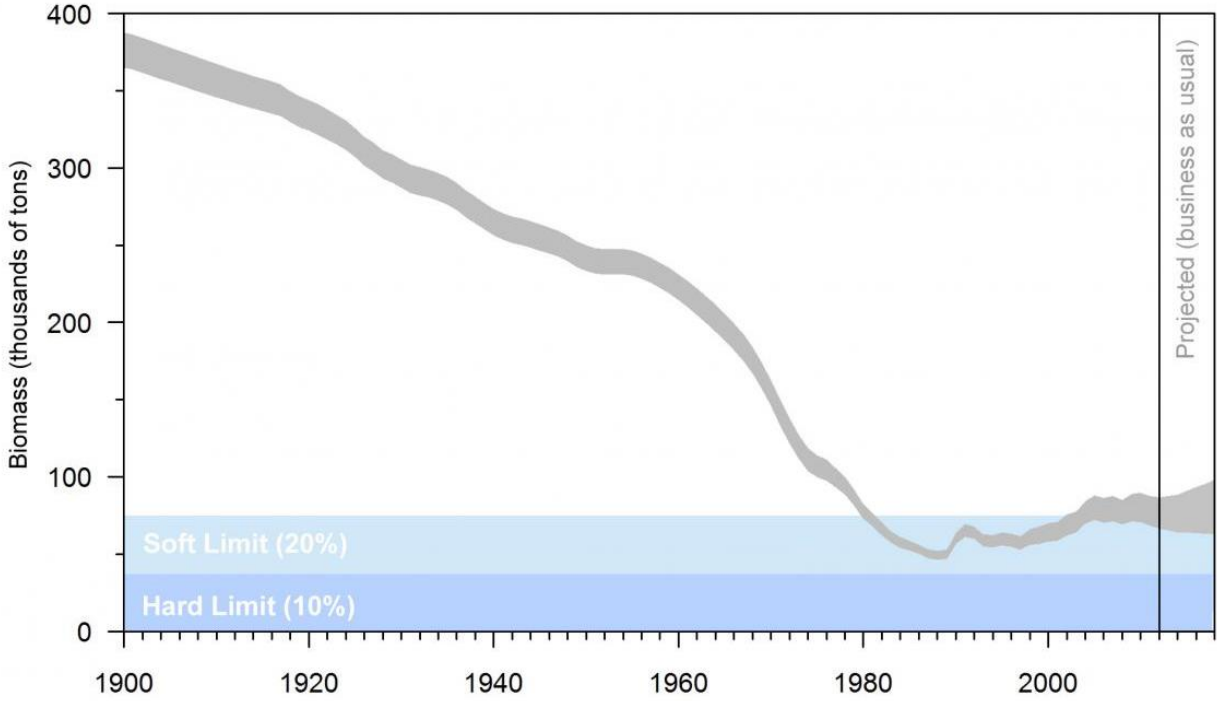

Figure S2.1: population abundance of snapper in management area SNA8 in thousands of tonnes. Source: NIWA

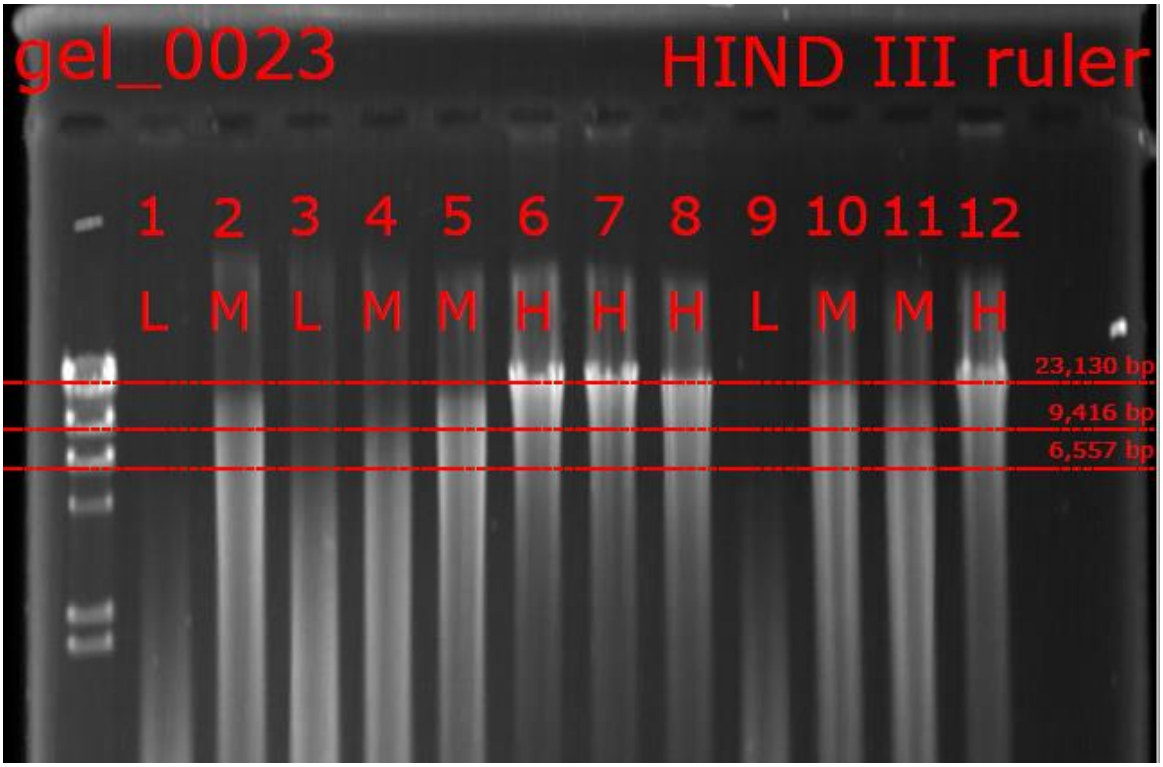

Figure S2.2: example of gel electrophorese showing DNA extraction with various quality scores. A high molecular weight DNA ladder (i.e. HindIII) was used to reference DNA fragment sizes.

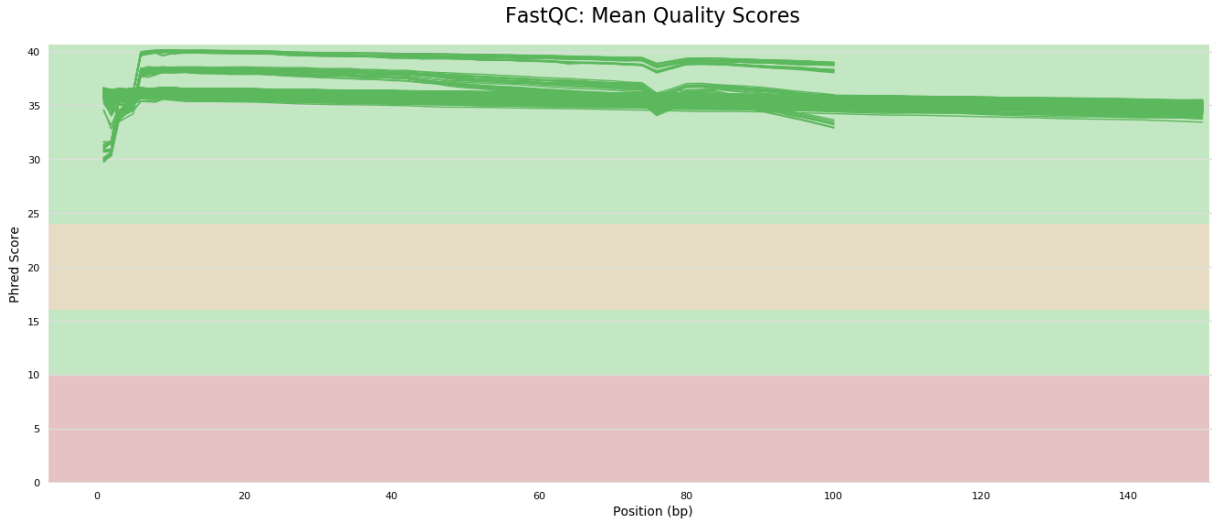

Figure S2.3: Mean quality scores per position from all modern samples. Read files were analysed using fastqc and summarizes with multiQC for visualisation. Samples from the 12 individuals provided by PFR were sequenced with 100 bp fragments (Catanach et al., 2019), while all subsequent re-sequencing was performed with 150 bp fragments.

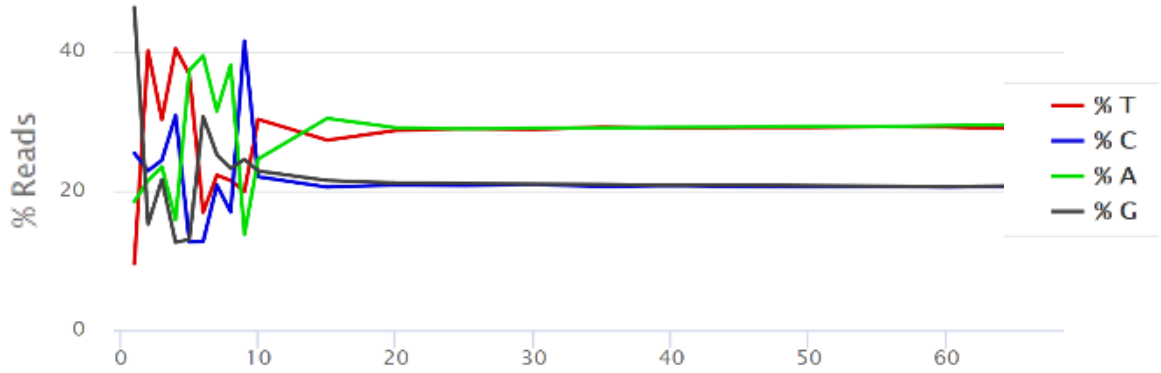

Figure S2.4: Example of the per base sequence content from reads generated on the NovaSeq following the Nextera flex library prep protocol. This figure has been cropped to only show the relevant section of the read.

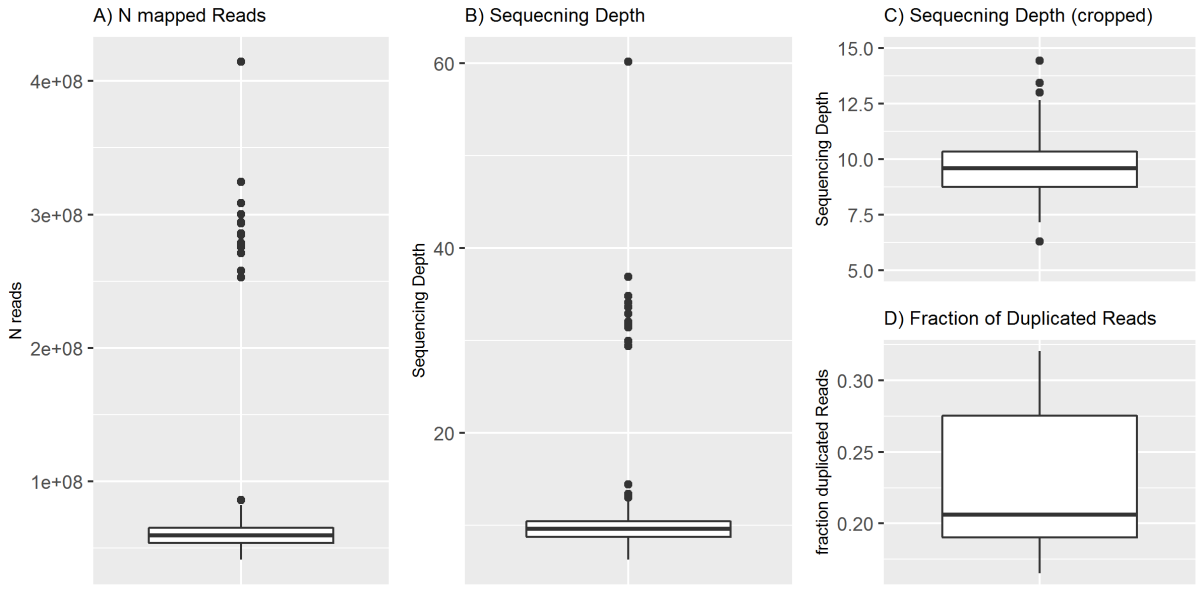

Figure S2.5: : A) number of mapped reads per individual, B&C) sequencing depth per individual, D) fraction of duplicated reads per individual

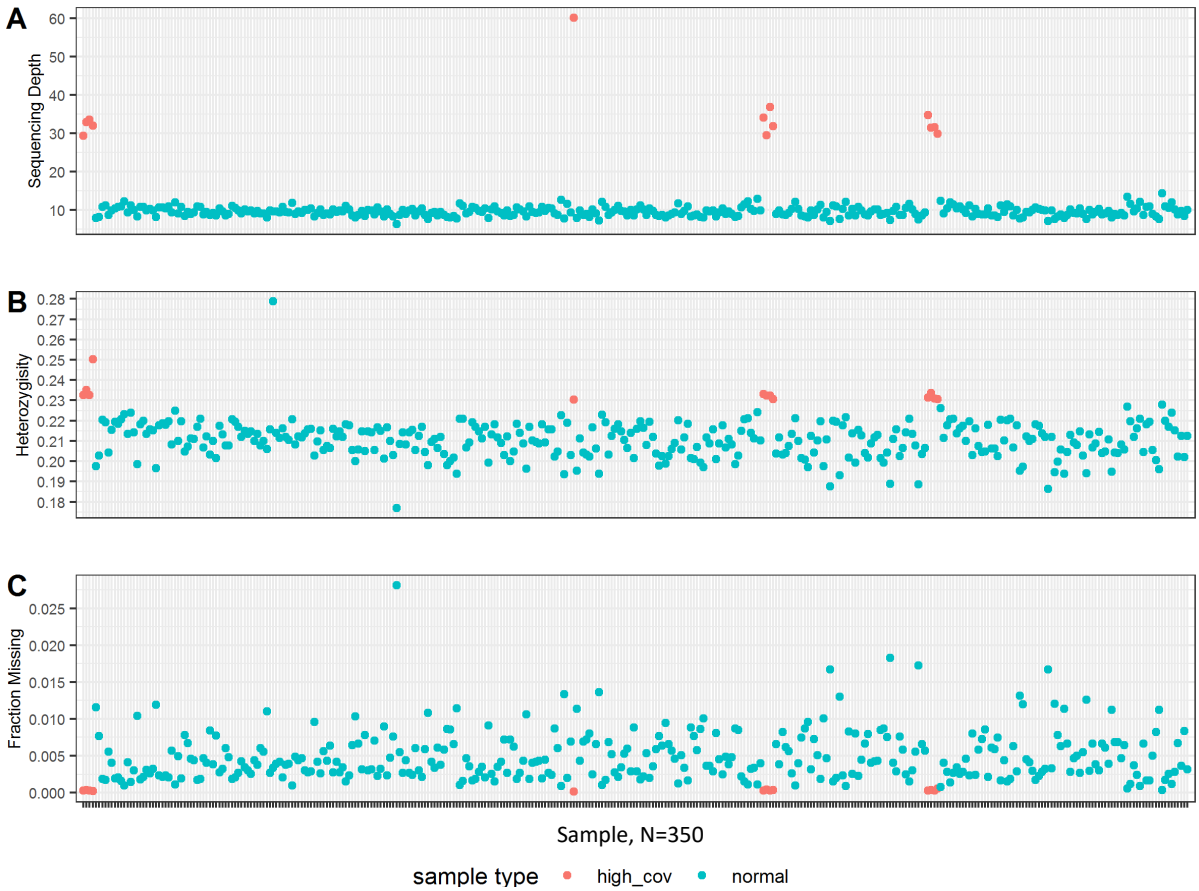

Figure S2.6: Point plots showing A) Sequencing depth, B) Heterozygosity, and C) fraction of missing genotypes per individual. Colours indicate the 2 different sequencing classes: High_cov samples (orange) are samples with were sequencing at a greater depth compared to the majority of samples. Normal samples (blue) indicate the bulk of samples that were sequenced at an average of 10.2x.

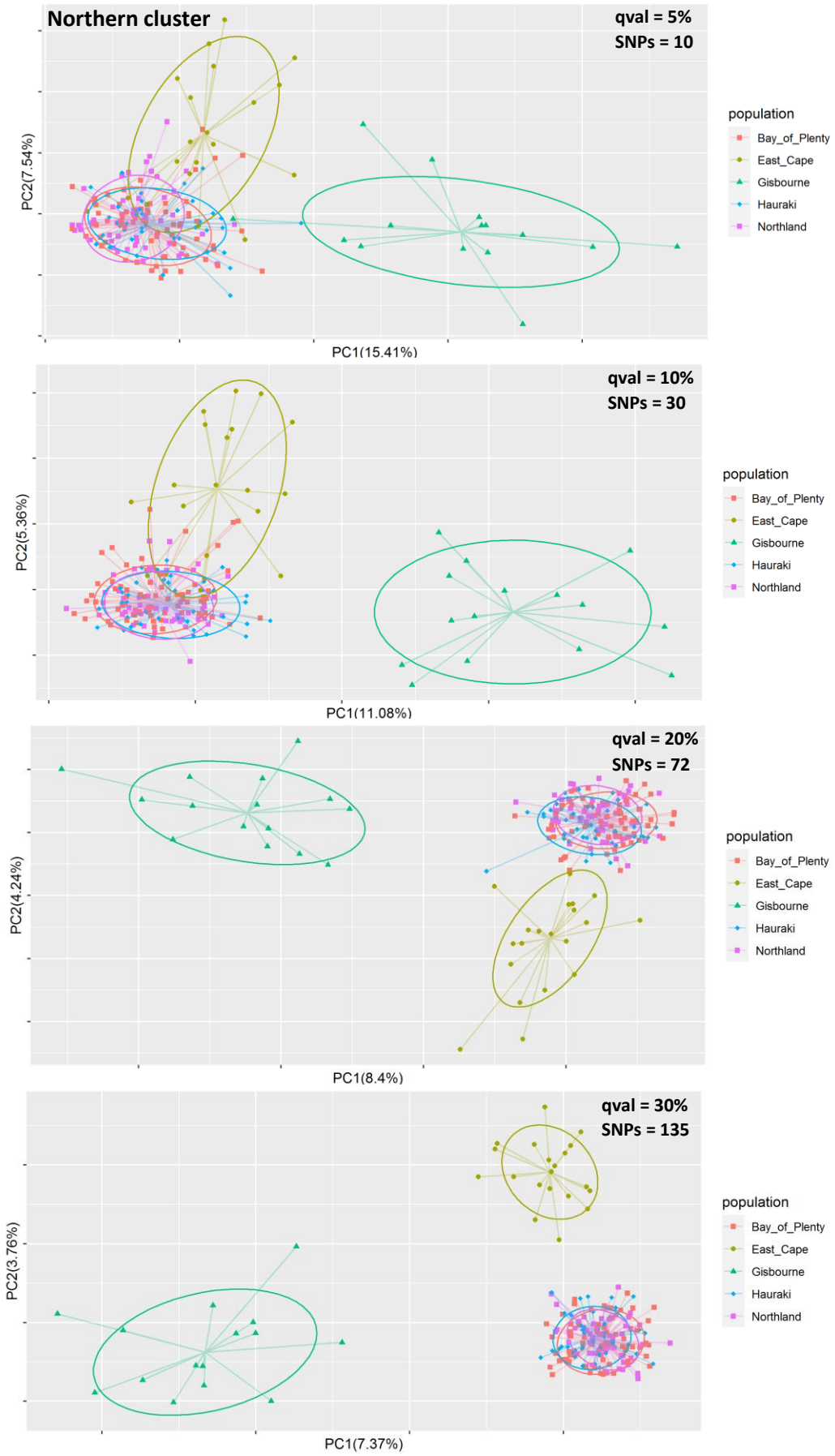

Figure S2.7: PCAs showing the effect of different false discovery rate boundaries (qval), and the number of putative adaptive SNPs and genetic clustering in the northern genetic cluster.

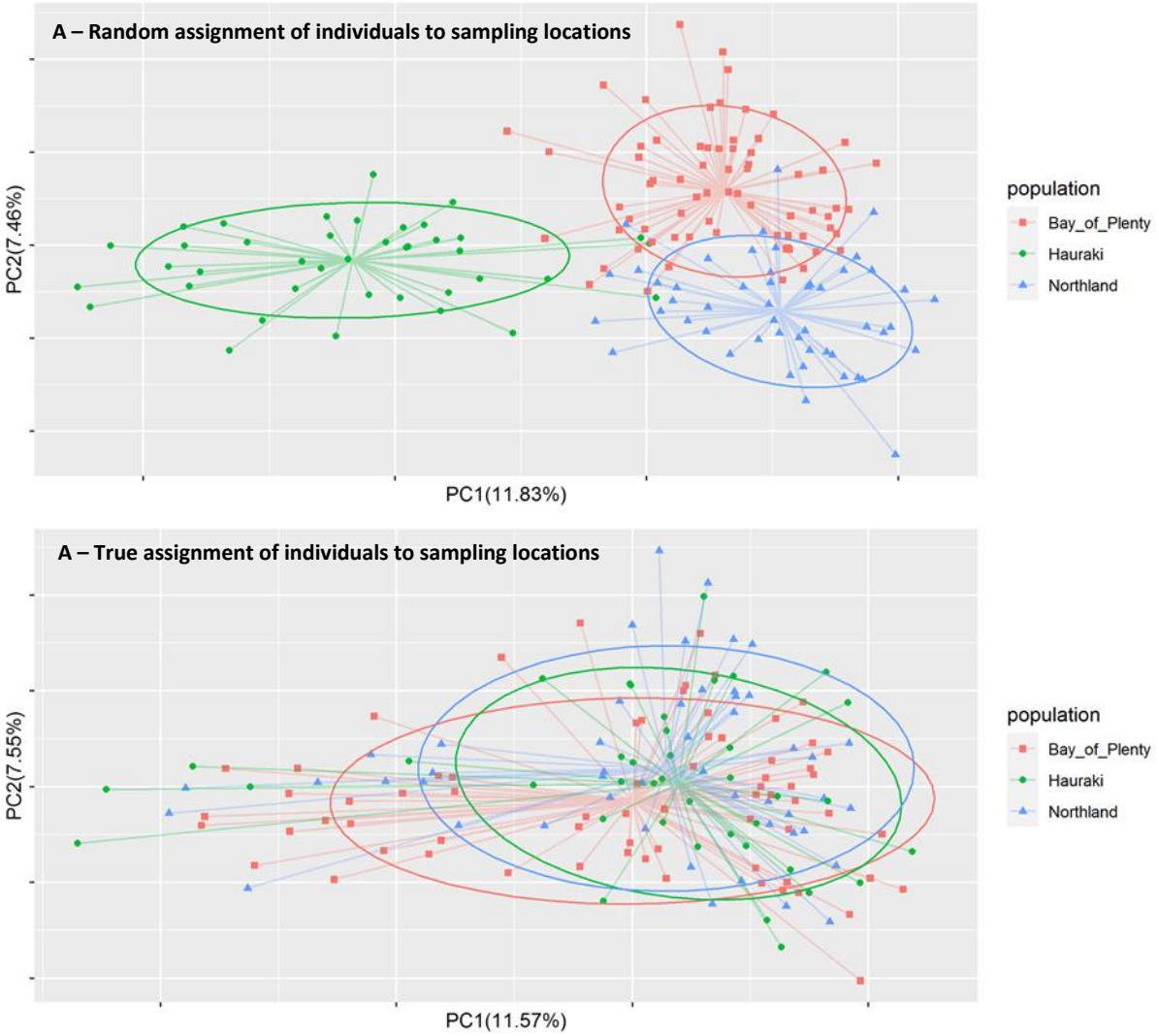

Figure S2.8: PCAs showing the effect of random assignment of individuals when identifying putative adaptive SNPs with high false discovery rates (50%). A) Random assignment of individuals to sampling locations. B) true assignment of individuals to sampling locations.

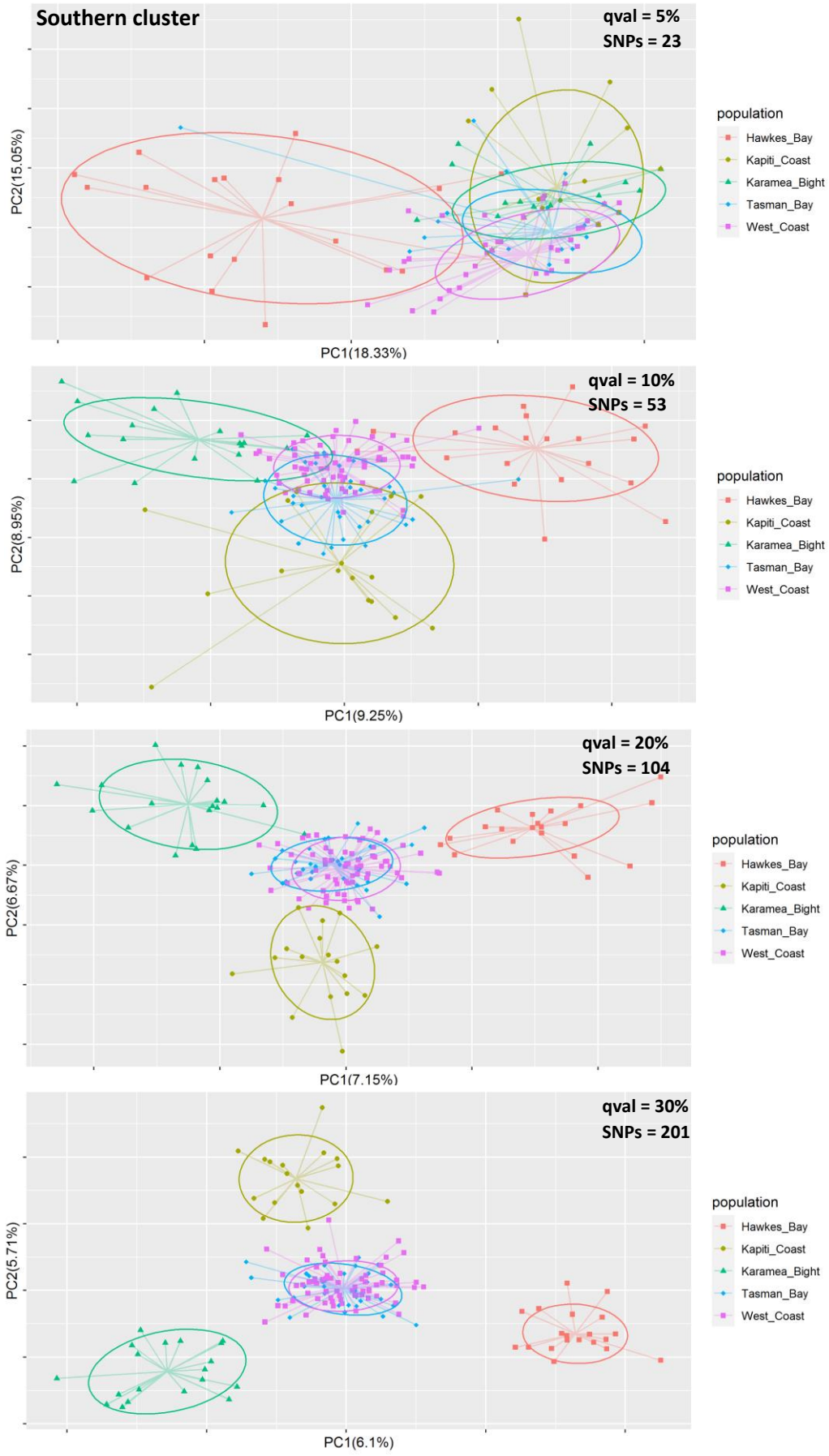

Figure S2.9: PCAs showing the effect of different false discovery rate boundaries (qval), and the number of putative adaptive SNPs and genetic clustering in the northern genetic cluster.

Chapter 3. The influence of glacial cycles on the Australasian snapper (*Chrysophrys auratus*) in New Zealand: phylogeographic and demographic history using modern and ancient DNA sequences

3.0. Abstract

Glacial cycles play important roles in determining the phylogeographic structure of terrestrial species, nonetheless relatively little is known about their impacts on the distribution of marine biota. In this study, mitochondrial and nuclear genomes from modern and ancient samples were used to assess the demographic and phylogeographic history of the Australasian snapper. Twenty-six mitochondrial genomes and one nuclear genome were obtained for samples up to 750-year-old. Ancient sequences were combined and compared with 350 modern mitochondrial genomes, and 12 high-coverage (30x) modern nuclear genomes. Mitochondrial haplotype diversity was high (0.968-0.982), similar to other fish species known to have high effective population sizes. An approximately 650,000-year-old split between two distinct mitochondrial clades resulted in a high level of nucleotide diversity (0.0045-0.0051). These old lineages suggest the species was separated into multiple populations during previous glacial cycles, creating geographically isolated populations and phylogeographic structuring. More recent lineages also provide evidence for isolated populations during the last glacial maximum. An exponential population increase observed around 20,000 years ago provided strong support for a post-glacial expansion, likely facilitated by increased ocean temperatures and rising sea levels. The exponential population increase around 10,000 years ago detected in the ancient nuclear genome is consistent with a post-glacial expansion. The population expansion was not detected in the modern nuclear genomes. The demographic history of snapper suggests Pleistocene glacial cycles played an important role in their demographic history. This study adds to our understanding of how geological changes have influenced the evolution of New Zealand's biodiversity.

3.1. Introduction

Phylogeography is the study of the geographic distribution of genealogical lineages, and provides key insights about past demographic events such as population fragmentation, expansion, and contraction. Geological events such as continental drift, mountain formation, changes in climate, sea currents, and glacial cycles have been shown to play important roles in the evolution and geographic structuring of species (Teske et al., 2011; Wallis & Trewick, 2009). Specifically, the substantial impact of glacial cycles on the phylogeographic structure and population demography has been observed in both terrestrial and marine settings (Hewitt, 2000). In terrestrial ecosystems, cooling in some areas has led to the formation of glaciers, disrupting gene flow between populations in a diverse range of species. Long-term extrinsic barriers to gene flow typically produce deep divergences within gene tree in non-recombining genetic markers, such as the mitochondrial DNA (mtDNA) genome, and similar patterns are often shared among taxa. For example, common phylogeographic patterns between terrestrial species in North America were attributed to populations isolated in refugia present during the last glacial maximum (Shafer et al., 2010). Similarly, red deer in Europe were found to display deep mitochondrial lineages, suggested to have separated in allopatric populations during the last glacial maximum (Skog et al., 2009). In marine ecosystems, species experienced range contractions to lower latitudes due to sea ice extensions, and lower ocean temperatures (Hewitt, 2000). In addition, as a result of widespread glaciation sea levels dropped by approximately 120 meters (Dlabola et al., 2015; Peltier & Fairbanks, 2006), significantly changing coastal marine habitats, current patterns, and blocking dispersal and migration pathways (Bowen et al., 2016). For example, reduced ocean sea levels during the Pleistocene glacial periods are thought to have created a land bridge between Taiwan and China, creating two distinct mitochondrial lineages in Chinese four-eyed sleeper (*Bostrychus sinensis*) (Ding et al., 2018; Qiu et al., 2016). Sea level changes in the same area were also connected to the divergence of two species of shore birds (X. Wang et al., 2019). Moreover, sea-level change in the Qiongzhou Strait during glacial cycles has been linked to phylogeographic structuring in *Lepturacanthus savala* (Gu et al., 2021), *Monodonta labio* (Zhao et al., 2017), *Gobiopterus lacustris* (Wang et al., 2018), and *Pampus argenteus* (Sun et al., 2019).

These studies show sea-level changes during glacial cycles play a dominant role in the genetic structuring of coastal species.

Studies on the effects of the last glacial maximum on phylogeography have largely been focussed on northern Hemisphere species, as the effects were most prominent in this region. In the southern Hemisphere, glaciers did not cover the terrestrial habitat to the extent they did in the northern Hemisphere (Hewitt, 2000). In New Zealand, repeated glaciations during the Pleistocene made many areas uninhabitable but there were still pockets of habitable refugia. These sustained isolated populations and facilitated a high degree of phylogeographic structuring in kiwi (Weir et al., 2016), mycophagous beetle (Marske et al., 2009), and alpine insects (McCulloch et al., 2010), to name a few. In contrast, increased dispersal was facilitated through reduced sea levels which created land bridges between the three main islands during the last glacial maximum (i.e. North Island, South Island, and Stewart Island) (Wallis, 2019). An increase in shared haplotypes was observed between populations of *Kikihia* spp (cicadas) and *Galaxias* spp (freshwater fish) found on Stewart Island and the South Island (Marshall et al., 2009; Waters et al., 2001).

While the effects of the last glacial cycle in New Zealand's terrestrial habitats are well studied, the impact on marine species is less well studied. Available habitat for coastal species was likely reduced due to the low sea levels and relatively small continental shelf. In addition, deeper waters separate New Zealand from Australia, restricting range contractions to lower latitudes for coastal species (Barrows et al., 2007). Ocean temperatures were on average 6°C lower compared to the current day, suggesting species distributions would have been shifted northward. This would have impacted marine species inhabiting New Zealand waters. For examples, phylogenetic structuring was reported in two species of whelk (*Cominella virgata* & *C. maculosa*) (Dohner et al., 2018; Fleming et al., 2018). The phylogeographic structure was possibly established during the last glacial maximum when the species resided in refugia. Also, Phylogeographic structuring in kelp around New Zealand has been reported (Fraser et al., 2009), but is unclear whether glacial cycles drove this genetic differentiation.

Australasian snapper (*Chrysophrys auratus*) is a coastal marine fish species that is common in the waters around the North Island, and top of the South Island of New

Zealand (Parsons et al., 2014) (Figure 3.1). The species also occurs in Australian waters but shows clear phylogenetic separation suggesting minimal gene flow between the two regions (Tabata & Taniguchi, 2000). *Chrysophrys auratus* and its closest relative, *Pagrus major* (red seabream/真鯛 madai), were reported to have split between two and six million years ago (Tabata & Taniguchi, 2000). Previous studies (including the results reported in Chapter 2 of this thesis) have observed weak but significant population structure using nuclear DNA sequences of *C. auratus*, while no significant population structure was detected using mitochondrial markers (Bernal-Ramírez et al., 2003; Smith, 1979; Smith et al., 1978). This suggests that in recent history, populations have been relatively well connected to maintain genetic similarity. It has been well-documented that commercial harvesting caused drastic population declines of snapper, which resulted in stock biomass dropping below the soft limit (i.e. 20% of their original biomass) (Fisheries New Zealand, 2018), however, little is known about the demographic history of the species prior to human arrival in New Zealand. Snapper are commonly found in water up to 50 meters deep, and observed down to a maximum depth of 200 meters. Their range appears to be restricted by water temperature at their southern boundary. During the cooler periods of the last glacial cycle, lower sea levels would have limited the amount of suitable habitat, and the colder water temperature reduced the extent of their southern distribution to a smaller overall northern area. This suggests that snapper in New Zealand would have experienced a significant level of disruption to gene flow and their distributional range. A demographic impact that could have changed their levels of genetic diversity and shaped their phylogeographic structure.

This study uses mitochondrial and nuclear data obtained from modern and ancient snapper to assess the phylogeographic and demographic history of snapper. The application of ancient DNA has primarily focussed on terrestrial vertebrates as conditions for preservation are more favourable in cold and dry environments. Porous bones and harsh conditions for preservation causes fish DNA to degrade much faster. Fish that have been harvested by humans and deposited on land under favourable conditions have been used for genetic research. Studies that have utilised ancient DNA to research fish have almost exclusively used small mitochondrial markers (Oosting et al., 2019). However, recent studies show great potential for the retrieval of ancient DNA

from fish bones (Ferrari et al., 2021; Oosting et al., 2019), suitable for obtaining entire genomes from specimens. This study used massive-parallel-sequencing of both modern and ancient samples to obtain complete mitochondrial and nuclear genomes. This data was then used to assess the demographic history of Australasian snapper in New Zealand. The following questions will be addressed: 1) is there evidence of phylogeographic structure, 2) when did *C. auratus* and its closets relative (*Pagrus major*) diverge, 3) how did snapper abundance in New Zealand change through time, and 4) what geological events could have played important roles in the demographic history of snapper?

3.2. Methods

This chapter includes samples from 350 modern, and 26 ancient snapper from 10 populations around New Zealand (Figure 3.1, Table 3.1). In addition, data from 12

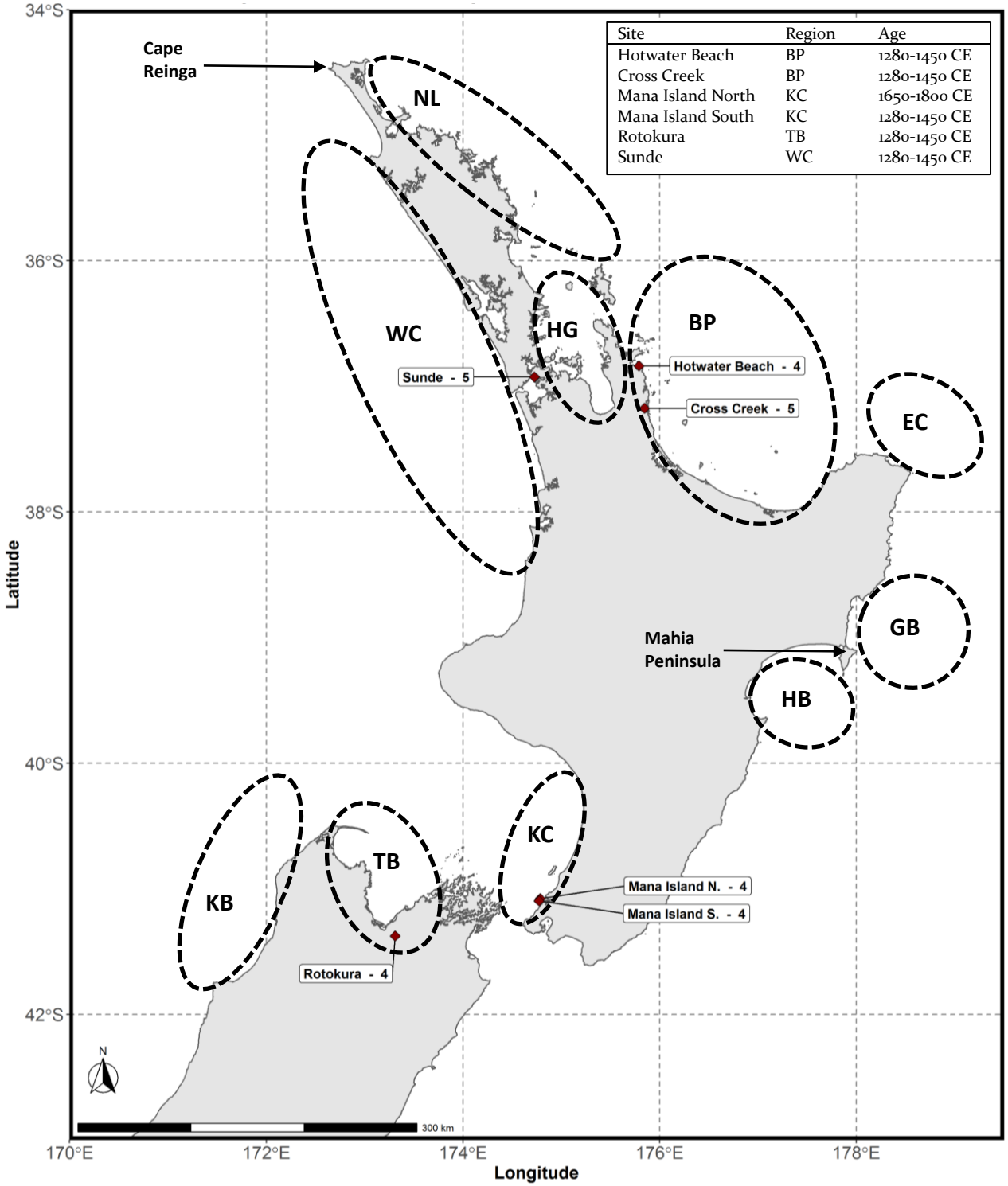

Figure 3.1: Map showing the sampling distributions of both modern and ancient samples. Red diamonds indicate sites where ancient mitochondrial genomes were obtained, numbers behind site names indicate number of successfully sequenced mitochondrial genomes. See Figure S3.1 for a map showing all ancient sampling sites and Table S3.1 for location information. The table on the top right shows the estimated age of the ancient sampling sites, and samples regions to which they are allocated.

modern and one ancient nuclear genomes have been included in the data. The sampling and sequencing of all modern samples are described in Chapter 2.

3.2.1. Sampling of ancient DNA samples

Ancient snapper dentaries were collected from the Museum of New Zealand Te Papa Tongarewa, Wellington, New Zealand. In total, 124 snapper dentaries from 18 different sites were sampled (Figure S3.1 & Table S3.1). Sampling was restricted to snapper dentaries as they are relatively abundant and easy to identify (Figure 3.2). At each site only a left or right dentary was sampled, depending on which one was most abundant, to avoid sampling the sample individual twice. Sampling sites were classified as either early (1280-1450 CE) or late (1650-1800 CE) Māori occupation (Figure S3.1) (Smith & James-Lee, 2010). In all cases the relevant Māori iwi were consulted with, and the project goals and methods were discussed. Ngāti Kuia and Ngāi Tai Ki Tāmaki supported the research, while Ngāti Toa, Ngāti Hei and Ngāti Koata will be kept up to date on the progress of the research despite not responding to requests for consultation. Samples were transferred directly from Te Papa to a dedicated ancient DNA laboratory at the Otago Palaeogenetics Laboratory, Department of Zoology, University of Otago, led by Nic Rawlence, to maintain physical separation between the modern and ancient DNA samples, which minimized the risk of cross-contamination (Knapp et al., 2012).

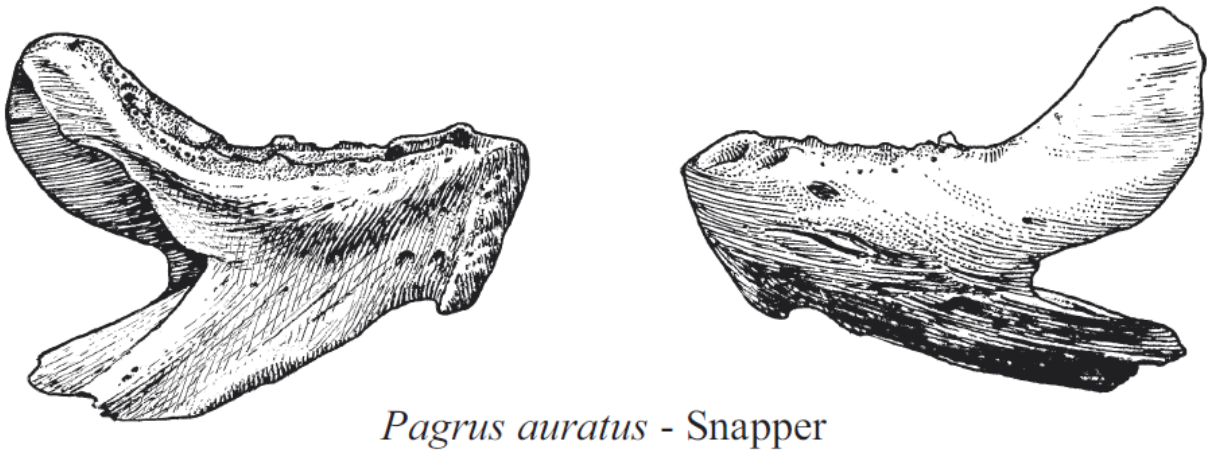

Figure 3.2: Sketch of left dentary used for identification of ancient snapper (*Chrysophrys auratus*) samples. *Pagrus auratus* refers to old nomenclature prior the assignment of *C. auratus* for New Zealand and Australian snapper (Parsons et al., 2014), source image: Leach (1997).

3.2.2. Ancient DNA extraction

The tooth row from snapper dentaries' was cut out using a Dremel tool and carbon fibre cutting disk. Care was taken to not destroy the morphometric landmarks. Each tooth row was ground to a fine powder using a sterile mortar and pestle. DNA was extracted from the powdered bone samples using the method described in Dabney et al. (2013) and Boessenkool et al. (2017). Ancient DNA work was performed in dedicated labs using strict protocols (Cooper & Poinar, 2000; Knapp et al., 2012), and the possibility of undetected cross-contamination was minimized using negative extractions and blank PCRs. The quantity of DNA obtained from the extraction procedure was determined using an Agilent 2100 Bioanalyzer following the manufacturer's instructions. Analytical profiles of samples that had a normal distribution of fragment sizes (approximately between 30-50 bp) suggested they contained sufficient quantity of DNA for PCR amplification and had the expected degradation profiles. These samples were shipped using continuous cold-store transportation to an ancient DNA laboratory in Oslo, Norway, for library preparation and sequencing. Samples that had a profile of much shorter fragment size distribution were deemed too degraded for sequencing and set aside. Samples that had fragmentation distribution profiles over 200 bp were most likely contaminated and contained DNA from modern microorganisms or humans, and were excluded from this study.

3.2.3. Ancient DNA sequencing

DNA sequencing and library preparation were performed in the laboratory of Bastiaan Star and Sanne Boessenkool at the University of Oslo (UiO), Norway. A dedicated ancient DNA facility was used that has never had any samples of modern DNA (Cooper & Poinar, 2000). Double-stranded sequencing libraries were constructed according to Ferrari et al. (2021). Fragments were double-indexed (7bp indexes) after blunt-end repair and amplified using PCR. The quality of the libraries were checked using the Agilent 5200 fragment analyser using the high sensitivity (HS) assay (DNF-473-1000), following manufacturer's instructions. Here, fragment size distribution of approximately 200 bp were considered appropriate for DNA sequencing. This matched the anticipated fragment lengths of sample DNA, indices and adapter sequences. Libraries were pooled with a minimum concentration of 2.0 ng/µl, and the DNA

sequences determined on a single lane on Illumina Hi-Seq4000. Sequences were deplexed allowing for zero mismatches in the index, yielding approximately 10 million reads per sample. This exploratory sequencing was done to determine what percentage of reads within each sample originated from snapper (endogenous DNA content). Endogenous DNA content was determined by mapping sequence reads to the snapper genome, calculating the percentage of mappable reads to assess sample quality and to determine whether more reads would need to be generated to obtain the mitochondrial genome or nuclear genome with sufficient coverage. Library concentrations were adjusted to generate the desired number of reads that allowed the retrieval of mitochondrial or nuclear genomes with an average fold coverage depth of 5X and 10X, respectively.

3.2.4. Data alignment of ancient samples

Reads were aligned to both the reference nuclear and mitochondrial genomes to obtain an estimate of the endogenous DNA content for each sample (Ashton, 2013; Catanach et al., 2019). Reads were mapped to the nuclear and mitochondrial genome in the same analyses to prevent false alignment of reads to either genome due to the absence of the correct genomic regions. Mapping of the reads to the references was done using the bam_pipeline in paleomix v1.2.13.3 (Schubert et al., 2014). Below is a description of the steps included in the bam_pipeline. First, Adapter removal v2.2.3 was used for trimming off the Illumina adapters.

Adapter 1:
GATCGGAAGAGCACACGTCTGAACTCCAGTCACNNNNNNNATCTCGTATGCCGTCTTCTGCTTG

Adapter 2:
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCATT

A 33% mismatch rate (--mm 3) and a minimum read length of 25 bp were set for sequence comparisons. Ambiguous nucleotides (N's) and low-quality bases were trimmed of all 5' and 3' end. In contrast to reads generated from the modern DNA samples, mapped reads pairs were collapsed and merged to a single read for the ancient

DNA samples. This was done because the fragment sizes were so small that there was no insert size and paired reads with overlapping sequences. MapDamage v2.0.9 was used to detect post-mortem DNA damage patterns which causes an excess of cytosine to thymine (C-to-T) misincorporations at 5′ ends, and complementary guanine to adenine (G-to–A) misincorporations at 3′-termini (Jonsson et al., 2013). Burrows-Wheeler Aligner (BWA) v0.7.15 (bwa-mem algorithm) was used to map reads to both the nuclear and mitochondrial reference genomes (see section 2.1.4. for details on nuclear and mitochondrial genomes). Unmapped reads and reads with a mapping quality of <25 were removed. PCR duplicates generated in amplification steps during library construction were removed using MarkDuplicates, Picard tools v2.18.20 (Broad Institute, 2020). Finally, the pipeline used the indel realigner (GATK v4.0.8.1) for local realignment (McKenna et al., 2010).

3.2.5. Alignment of modern sequences and outgroup

The mitochondrial and nuclear alignment of 350 modern samples were described in Chapter 2, section 2.2.4. Ten whole-genome-resequencing data from wild-caught red seabream/真鯛 madai (*P. major*) from Japan were used as an outgroup (Nam et al., 2019). The sequences from Australasian snapper from New Zealand were aligned to the sequences from the nuclear (Shin et al., 2018) and mitochondrial genomes of *P. major* (Miya et al., 2001). Parameters identical to those used for the modern snapper samples were used in the paleomix pipeline to generate the alignment files for *P. major* (see section 2.2.4.).

3.2.6. Assembling mitochondrial genomes

Mitochondrial genomes of both ancient and modern samples were assembled using the same pipeline. GATK v4.1.4.1 was used to call variants for all mitochondrial genomes as the joint calling algorithm incorporates variant information from all individuals when calling genotypes for each individual (Poplin et al., 2018). This is particularly useful when handling assemblies from ancient DNA samples which often have low individual coverage. Only ancient samples that had an average coverage of >5X were selected for generating complete mitochondrial assemblies.

First, genomic Variant Call Format (g.vcf) files were generated from the alignment file outputs of paleomix. This was done using the GATK's HaplotypeCaller, using the snapper mitochondrial genome as a reference, setting ploidy to 1 (-ploidy 1) and the reference confidence mode set the gvcf (-ERC GVCF). A g.vcf file is generated for each individual and used in the next steps for joined genotyping. All g.vcf files were subsequently merged using GATK's CombineGVCFs function (default settings). GATKs best practices advise using their DBImport function to merge g.vcf. However, there is no specific option for haploid sequences. The older function is recommended which allows the user to specify the ploidy of the sample.

Genotypes were called using GenotypeGVCFs using the following settings, -G StandardAnnotation, -ploidy 1, -new-qual. Subsequently, the vcf file was filtered using bcftools v1.9, and vcftools v0.1.16 in bcftools sites we filtered on possible strand bias (FS<60.0, SOR<4), mapping quality (MQ>30), and quality by depth (QD>2.0). In vcftools, sites were filtered on minimum genotype quality (--minGQ 15) and minimum sequencing depth (--minDP 3).

Fasta sequences for each individual were generated using GATK's FastaAlternateReferenceMaker. First, the genotype information of each sample was split into individual vcf files. Then, nucleotide sites with missing genotypes were replaced by N using a custom script (provided by Bastiaan Star, Oslo University). Finally, genotype information and the mitochondrial reference genome were used in combination to reconstruct the full mitochondrial genome of each sample. All FASTA files from snapper (*C. auratus*) were concatenated into one alignment. The sequences of *P. major* were aligned by eye to those of *C. auratus*. Basic summary statistics on the mitochondrial genome data were generated in R (Packages: ape, pegas, and haplotypes)(R Core Team, 2013).

3.2.7. Mitochondrial genome variation

The number of unique haplotypes was calculated using the R-package haplotypes (Aktas, 2015; R Core Team, 2013). The R-package ape v.5.4-1 was used to read alignment files and calculate the number of segregating sites (Paradis & Schliep, 2019). The R-package pegas v.0.14 was used to estimate nucleotide diversity (pairwise.deletion =

True), haplotype diversity (variance = True), Tajima'sD, and genetic diversity (theta) based on the number of segregating sites (Paradis, 2010). The number of missing sites was inferred using the R-package seqinR (Charif & Lobry, 2007). An hierarchical AMOVA (1000 permutations) was performed using the R-package apex (Jombart et al., 2017), testing for significant genetic differences between sample locations (Figure 3.1).

Nei's D_A for nucleotide divergence between New Zealand and Japan was estimated using the R package strataG and based only on the mitochondrial control region sequence. This enabled the divergence rates reported by Tabata and Taniguchi (2000) to be used. The previous divergence time was estimated between two and six million years. This was based on a sequence divergence of 3.48% (Nei's D_A), and the assumption that the mitochondrial control substitution rate was between 0.6 and 1.4% per million years. A critical mistake by Tabata and Taniguchi (2000) was using the time estimate based on D_A and the mutation rate as a divergence time. However, the estimate should take into account the fact that the two species have diverged, so this time estimate needs to be divided by two to get the time when these species diverged, assuming both species diverged at the same rate. This implies that the two species diverged between one and three million years ago.

3.2.8. Partitioning and substitution model selection

The mitochondrial genome was annotated in Geneious by searching for homologous sequences. The *C. auratus* annotation was compared for consistency to the annotation of *P. major*, obtained from the Mitochondrial Genome Database of Fish, MitoFish (Sato et al., 2018). The mitochondrial genome was split into the individual protein-coding genes, 12S rRNA, 16S rRNA, control region (or D-loop), and the 22 tRNA sequences. Intergenic regions were added to the tRNA sequences. Protein coding genes were subsequently partitioned into the 1st, 2nd, and 3rd codon positions, and concatenated to obtain 1st, 2nd, and 3rd codon alignments. Genes with reversed reading frames (from '3 to '5, e.g. ND6) were grouped as if the 3rd codon was the first. PartitionFinder2 identified substitution models for each isolated region, and were subsequently merged to obtain sequence partitions with unique substitution models. Additional parameters for each substitution model parameters were estimated for each

"sequence partition" using Jmodeltest2 (Darriba et al., 2012). Model selection was evaluated using AIC, AICc, and BIC.

3.2.9. Genealogy and demographic history

A maximum-likelihood tree was estimated in IQTREE (Nguyen et al., 2015) and 100 bootstrap replicates were used to assess support for each node. The bootstrap consensus tree was used to produce a haplotype genealogy in Fitchi (Matschiner, 2016), to visualize the relation between haplotypes. Samples were grouped into the two genetic clusters identified in Chapter 2, ancient samples and outgroup. Default parameters were used when running Fitchi, except the aesthetics value for node size (-m 0.3) and specifying the haploid nature of the data (--haploid). The output of Fitchi was used to group samples based on their mitochondrial haplotype, and grouped samples into clades.

The temporal stability of the relative proportion between different clades was tested. This was done to test if certain clades had a potential selective advantage. Large sample size differences between the modern and ancient samples required subsampling of the modern data set to produce comparable data sets. Fisher's exact test was used to test for a significant change to haplotype distribution. Subsampling was bootstrapped 1,000 times to obtain confidence intervals and biases from uneven samples sizes. The modern data set was randomly subsampled down to the same number of observation as the ancient DNA data set (i.e. 26). Calculations were performed using a custom Rscript (R Core Team, 2013).

Phylogenetic analyses were done using BEAST2 (Bouckaert et al., 2014). Data partitioning and mutation models were based on the results obtained from PartitionFinder2 and Jmodeltest2 (Table S3.5). A strict clock rate was used for all partitions, however, a different clock rate of $5.0e^{-8}$ per site per year ($site^{-1}/year^{-1}$) was used for the control region (Bowen et al., 2006). This clock rate for the control region is higher compared to the rest of the mitochondrial genome because it is thought to be selectively neutral and less conserved. This results in mutations accumulating at a higher rate in the control region. A clock rate of $3.28e^{-9}$ $site^{-1}/year^{-1}$ was used for all other regions of the mitochondrial genome. This latter estimate was based on the average

clock rate of nine fish species presented in Gillooly et al. (2005). Clock rates presented in both studies were based on estimates of nucleotide divergence (D_A), and are specific to the respective genetic regions used in this thesis. The substitution rate was allowed to vary between Codon1+RNA, Codon2, and Codon3 partitions. Base frequencies were obtained empirically from the data.

Three models (i.e. Yule, Coalescent Exponential Population, and Coalescent Bayesian Skyline) were used. Models were chosen based on the data set that was analysed. The Yule model was used to analyse the data set which included an outgroup as this model is better suited for genealogies that include multiple species. The Coalescent Exponential Population, and Coalescent Bayesian Skyline models were run on data sets which included a single species (i.e. *C. auratus*). Tip dates indicating the age of each sample were applied in the Coalescent Exponential Population, and Coalescent Bayesian Skyline model. The Yule model was run using the modern+ancient+outgroup data set. The Coalescent Exponential Population, and Coalescent Bayesian Skyline model were run on the modern and modern+ancient data set. The Coalescent Bayesian Skyline was run allowing for 3 population size changes. This relatively small number of population size changes was required to obtain effective sample sizes over 200. Final runs for each model consisted of two independent chains (each with a randomly generated seed), with a chain length of 50 million steps for the Yule and Coalescent Exponential Population, and 200 million steps for the Coalescent Bayesian Skyline. Each chain had a burnin of 5 million (10%), and data points saved one out of every 500 iterations. Subsequently, another 10% of each run was discarded before runs were concatenated. Trace profiles were visually checked for convergence using Tracer v1.7 (Rambaut et al., 2018). Trees estimated in BEAST2 were summarized in TreeAnnotator v2.6.2 (Bouckaert et al., 2014), generating a consensus tree. A custom script was used to extract heights (age) for each node, including 95% confidence intervals. Bayesian skyline plots were generated using Tracer v1.7 (Bouckaert et al., 2014), and visualised using a custom R script.

Pairwise Sequentially Markovian Coalescent (PSMC) was used to assess population size changes through time from single nuclear whole-genome alignments (Li & Durbin, 2011). Effective population size estimates over time were estimated for one

ancient nuclear, and the 12 high-coverage (30x) modern nuclear genomes described in Chapter 2. Sequence alignments of the modern nuclear genomes (described in section 2.2.4), and single ancient nuclear genome (described in section 3.2.4) were filtered following the instructions provided with the PSMC distribution. First, consensus sequences for each linkage group and individuals were obtained by first calling SNPs from alignment files using bcftools: bcftools mpileup –Q 30 –q 30 –r -f | bcftools call –c. Subsequently, SNPs were converted to sequence using vcfutils: vcfutils.pl vcf2fq –d 5 –D 34 –Q30. Polymorphic sites with a sequencing depth between 5 and 34X were retained. Linkage groups were merged into a single sequences file for each individual. Effective population size (N_e) was inferred over 54 free atomic time intervals (PSMC –p 4 + 25*2 + 2 + 4 + 6). Output was scaled by using a generation time of 4 years (Parsons et al., 2014), and a mutation rate of $5.5e^{-9}$. This mutation rate was chosen based on nuclear gene mutation rates used for other fish species (Table S3.2). PSMC plots were generated in R using a custom script from Liu and Hansen (2017).

3.3. Results

3.3.1. Ancient DNA sequencing

Of the 90 dentaries collected for ancient DNA extraction, 62 extractions showed promising bioanalyzer profiles with a peak in DNA concentration between 30 and 200 base pairs. This result suggested DNA was successfully extracted from the samples. After exploratory sequencing of these 62 samples, endogenous DNA levels of selected samples ranged between 0.003% and 58% (Figure 3.3 & Table S3.3). No sequences were obtained from three locations for which exploratory sequencing was performed (i.e. The Glen, Foxton, and Kokohuia). The remaining locations showed variation across sequenced samples, highlighting the high level of variability in sample preservation there was among samples from the same location. Based on the initial sequencing results, 25 samples were selected for deeper-coverage sequencing using a single lane of the Illumina Hi-Seq. This enabled the entire mitochondrial genome sequence to be determined at approximately 5x coverage (Figure 3.4).

Endogenous DNA levels of these samples ranged between 1 and 30% (Table S3.3). One individual obtained from Hotwater Beach was exceptionally well preserved with an endogenous DNA content of 58% and was chosen to be sequenced individually on a single lane to obtain the entire nuclear genome. The large number of reads required to

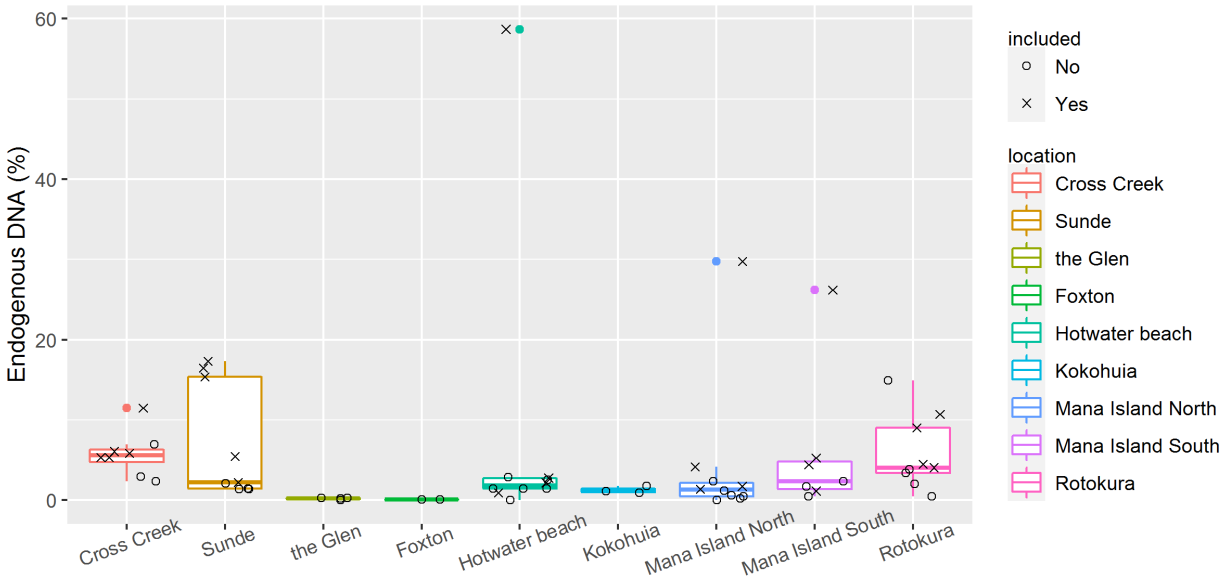

Figure 3.3: Boxplots showing the levels of endogenous DNA content (%) per site. Black crosses show individual samples that successfully yielded whole mitochondrial genomes. Black circles indicate samples that did not yield whole mitochondrial genomes. Crosses and circles are randomly plotted along the width of the boxplot. The number of samples per area from left to right: 8, 9, 4, 2, 10, 3, 10, 10, 9. For a cropped version see Figure S3.9.

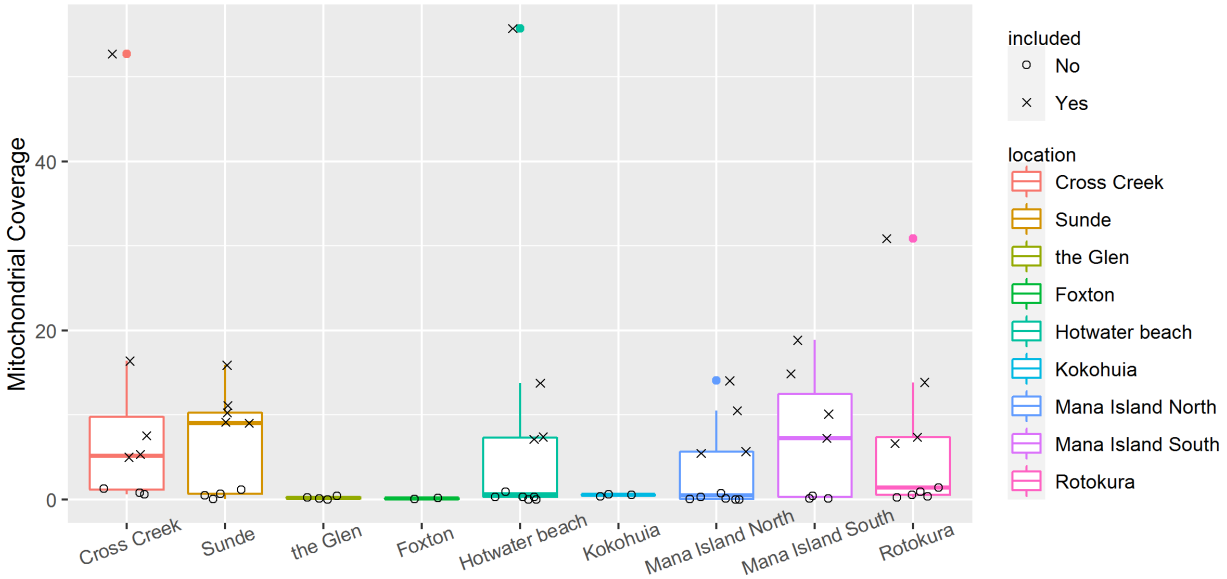

Figure 3.4: Boxplots showing the mitochondrial coverage of sequenced samples per site. Black crosses show individual samples that successfully yielded whole mitochondrial genomes. Black circles indicate samples that did not yield whole mitochondrial genomes. Crosses and circles are randomly plotted along the width of the boxplot. The number of samples per area from left to right: 8, 9, 4, 2, 10, 3, 10, 10, 9. For a cropped version see Figure S3.10.

obtain the nuclear genomes implied that the mitochondrial genome was able to be assembled as well. The single ancient nuclear genome was sequenced at a depth of 13x. Successfully sequenced mitochondrial genomes ranged in sequence depth from 5.02x and 55.7x (Figure 3.4). MapDamage profiles for ancient samples showed no significant signs of excess misincorporations at either end (Figure S3.11), suggesting only low levels of post-mortem depurination.

Out of the 26 whole mitochondrial genomes, the majority (22) of the successfully sequenced ancient DNA samples were obtained from middens associated with early Māori occupation (1280-1450) (Table S3.3). This provided a temporal sampling that maximised the time points between the modern and ancient samples. Samples associated with Late Māori occupation (1650-1800) were obtained from Mana Island North. Together with the samples obtained with Mana Island South, this geographic location is the only site with a temporal sampling with three time points (early, late, and modern).

3.3.2. Annotation and mutation models

PartitionFinder2 identified four distinct partitions that followed different distinct substitution models (Table S3.4), based on the mitochondrial annotation (Figure S3.2). First codon positions were grouped with ribosomal and transfer-RNA sequences, while 2nd codon, 3rd codon and D-loop had their mutation model. The best fitting mutation models for each partition was selected based on BIC as it overall the strongest support (Table S3.5). The different information criteria did not consistently select the same mutation model for each partition. BIC was used as the most reliable statistic, and the criterion to select mutation models for subsequent analyses. After alignment, all sequences (including outgroup) consisted of 16724 base pairs. Insertions and deletions were not present in this data set as BEASTs treats these as missing data, and multiple packages do not accept sequence alignments that contain sequences that vary in length. For this reason, one monomorphic site was removed from an intergenic region in snapper to visually align the mitochondrial genomes of both species. Finally, an additional three invariant sites were trimmed off at end of the control region so genomes of both *C. auratus* and *P. major* were of the same length.

3.3.3. Mitogenomic variation

The overrepresentation of mitochondrial reads in the libraries of the modern samples provided mitochondrial alignments at very high depth. On average, mitochondrial genomes were sequenced at an average depth of 286x, with a minimum of 151x and a maximum of 1704x. Similar, the single ancient DNA sample for which the nuclear genome was obtained at 55.7x coverage. The complete mitogenomic data set consists of 376 sequences, which were comprised of 350 modern and 26 ancient DNA samples. Average coverage for ancient mitochondrial genomes was as low as 5x, resulting in a higher number of nucleotides that could not be resolved and a higher proportion of missing data (Table 3.1).

Table 3.1: Information for 10 sampled populations, both modern and ancient samples

Populations	Code	Cluster	N	Np	Nh	Nm	π	h	θ	T_D
Northland	NL	North	50	244	31	25	0.00516	0.942	54.47	1.614
Hauraki Gulf	HG		40	247	33	21	0.00424	0.978	58.07	0.514
Bay of Plenty	BP		70 (79)	267 (271)	51 (57)	24 (251)	0.00524 (0.00509)	0.966 (0.973)	55.41 (54.85)	1.517 (-0.771)
East Cape	EC		20	229	17	22	0.00510	0.979	64.55	0.902
Gisborne	GB		15	206	12	0	0.00551	0.962	63.35	2.011*
Hawke's Bay	HB		20	222	13	28	0.00545	0.932	62.58	1.412
West Coast	WC	South	66 (71)	261 (263)	54 (56)	29 (89)	0.00532 (0.00527)	0.988 (0.990)	54.84 (61.23)	1.679 (1.527)
Kapiti Coast	KC		17 (25)	209 (228)	15 (19)	33 (114)	0.00493 (0.00489)	0.978 (0.967)	61.82 (54.38)	1.082 (0.531)
Tasman Bay	TB		33 (37)	226 (227)	26 (30)	30 (194)	0.00505 (0.00491)	0.975 (0.980)	55.69 (60.38)	1.516 (-0.351)
Karamea Bight	KB		19	214	16	89	0.00546	0.977	54.42	1.527
Modern	-	-	350	466	220	35	0.00510	0.968	72.43	0.214
Ancient	-	-	26	325	18	339	0.00448	0.982	56.08	-2.039*
total	-	-	376	472	233	339	0.00505	0.972	72.55	-1.984*

Notes: N = number of individuals, Np = number of polymorphic sites, Nm = number of missing sites, Nh = number of unique haplotypes, π = nucleotide diversity, h = haplotype diversity, T_D = Tajima's D, θ = genetic diversity (Waterson). Values in brackets show the estimates including the ancient samples collected at that site, *p<0.05

In total, 472 polymorphic sites resulted in 233 unique haplotypes (Table 3.1). The overall haplotype diversity of the pooled samples was high (0.972). Among populations, it ranged between 0.932 and 0.990. The high diversity of haplotypes was also seen in the rarefaction curve (Figure S3.3). Nucleotide diversity (0.00424 – 0.00551) and theta (54.38 – 64.55) was consistent across populations (Table 3.1). There were no significant changes to the diversity estimates with the inclusion of the ancient samples, except for Tajima's *D* (Table 3.1). The high abundance of nucleotide positions that were excluded from the analyses due to unresolved bases (N's) in the ancient samples resulted in biased estimates of Tajima's D and considered unreliable. Tajima's D estimates obtained from modern samples showed a lack of rare alleles, consistent with a population bottleneck. The hierarchical AMOVO showed that there is no significant difference between sampled populations, with genetic variance between populations was less than 0.01%, and thus practically all variation was present within populations (*p* 0.764).

3.3.4. Ancient DNA nuclear genome sequence

The ancient nuclear genome was obtained from Hotwater Beach with an endogenous DNA content of 58%. In total, $3.73e^8$ paired-end reads were generated from this sample which were collapsed to single reads with an average sequence length of 56 and 50 bp, nuclear and mitochondrial reads respectively. The genome was sequenced at a depth of 13x. After quality filtering, the level of observed heterozygosity was 0.208, which similar to the observed levels of heterozygosity estimated from the modern samples (Chapter 2, Table 2.2). The percentage of missing sites was 3.5%, which is approximately seven times more when compared to the modern samples (0.47%).

3.3.5. Phylogeographic structure

The haplotype network showed that both modern and ancient mitochondrial genomes clustered in two major mtDNA clades. These could then be subdivided into six minor clades (Figure 3.5). The two major clades had a sequence divergence (D_A) of 0.0084. The two clades were also found in the ancient DNA samples. There was no change in the relative abundance of haplotypes between clades over time, *p* 0.723 and *p* 0.737 for the major and minor clades respectively (Figure 3.5). The mitochondrial control region showed a 4.08% sequences divergence (Nei's D_A) between modern *C.*

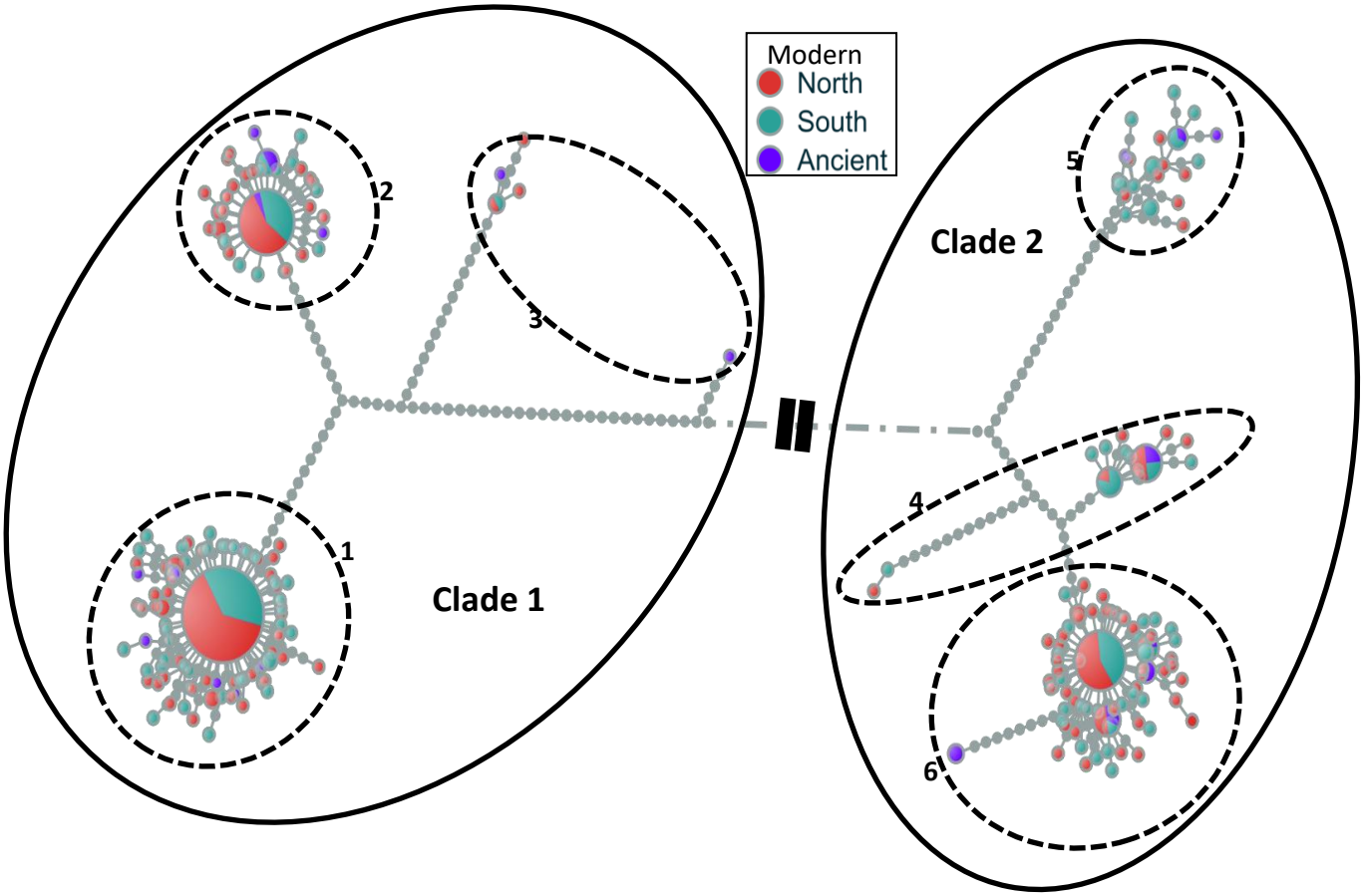

Figure 3.5: Haplotype network generated in fitchi, based on maximum likelihood tree produced in IQTREE. Clade 1 and clade 2 indicate the major clades. Colours indicate sample type, North, Centre, and South indicate modern sequences from one of the genetic clusters identified in chapter, and Ancient indicates ancient sequences. Each grey circle represents a single nucleotide change. The black bars indicates a section between the major clades representing 108 nucleotide changes. See Figure 3.6 for an unmodified version of the figure that shows the distance between the two major clades.

auratus and *P. major* (Figure S3.5). The AMOVA suggested no population structure based on mitochondrial haplotypes (p = 0.764). This also reflected in the distribution of the major and minor clades (Figure S3.7).

Posterior distributions for all parameters estimated with BEAST2 had effective sample sizes over 200, and trace profiles all reached stable states. All of the estimated parameters were normally distributed under the different models. Substitution rates were shown to be variable between partitions that were run using the same clock rate (3.28^{-9} site^{-1}/year^{-1}) (Figure S3.5). The codon1_RNA partition had a mutation rate approximately 7.5 times faster than the Codon2 partition. While the codon3 partition had a mutation rate approximately four times faster than the codon1_RNA partition (and 20 times faster than the codon2 partition). The 95% CI estimated for the growth

rate under the exponential expansion model did not include 0.0 in any of the three data sets (mean = 3.7238E-6, CI95% = $1.3998E^{-6}$ – $6.2791E^{-6}$), confirming that the population experienced a population expansion.

Trees produced in TreeAnnotator for the Yule and population expansion model were well resolved (Figure 3.6 & Table 3.2). For both trees, branches over 40,000 years old were highly resolved (posterior > 0.99). Younger branches in the trees were often poorly resolved from each other, with support below 0.5. A large radiation of new lineages is observed approximately ~20,000 years ago (T_4). This time estimate is consistent with the height of the last glacial maximum (Clark et al., 2009). Nodes T_2 and T_3 were estimated at ~150,000 years ago. Based on the exponential expansion model, the two major clades were estimated to have separated approximately 650,000 years ago (T_1). The Yule model that included the outgroup (*P. major*) estimated the clades that represent the two species that diverged 727,000 years ago (T_0, 95%CI 591,000 - 865,000). Runs that included the ancient genomes consistently placed nodes within the tree further back on time increase (Figure 3.6 & Table 3.2).

Table 3.2: node ages estimated using BEAST2

Data set	**Model**	**Time estimates in x 10,000 years**				
		T_0	**T_1**	**T_2**	**T_3**	**T_4***
Modern	Exponential	-	63.6 *47.6-80.6*	14.0 *9.6-18.8*	14.3 *10.0-19.0*	1.9 *0.7-3.1*
Modern + Ancient	Exponential	-	66.3 *49.0-84.0*	16.8 *11.8-22.2*	15.9 *11.1-20.5*	2.1 *0.8-3.5*
Modern + Ancient + Japan	Yule	72.7 *59.1-86.5*	54.8 *43.8-66.2*	13.9 *10.3-17.6*	13.1 *9.7-16.6*	2.6 *1.4-3.8*

Notes: time estimates are in x 10,000 years. The time estimates indicate the node ages in Figure 3.6. Values in italic indicate 95%HPD for each point estimate. *T_4 is the point average node age across all nodes younger than 30,000 years. Exponential indicates the data set was run using the exponential population expansion model, and Yule indicates that the Yule model was used.

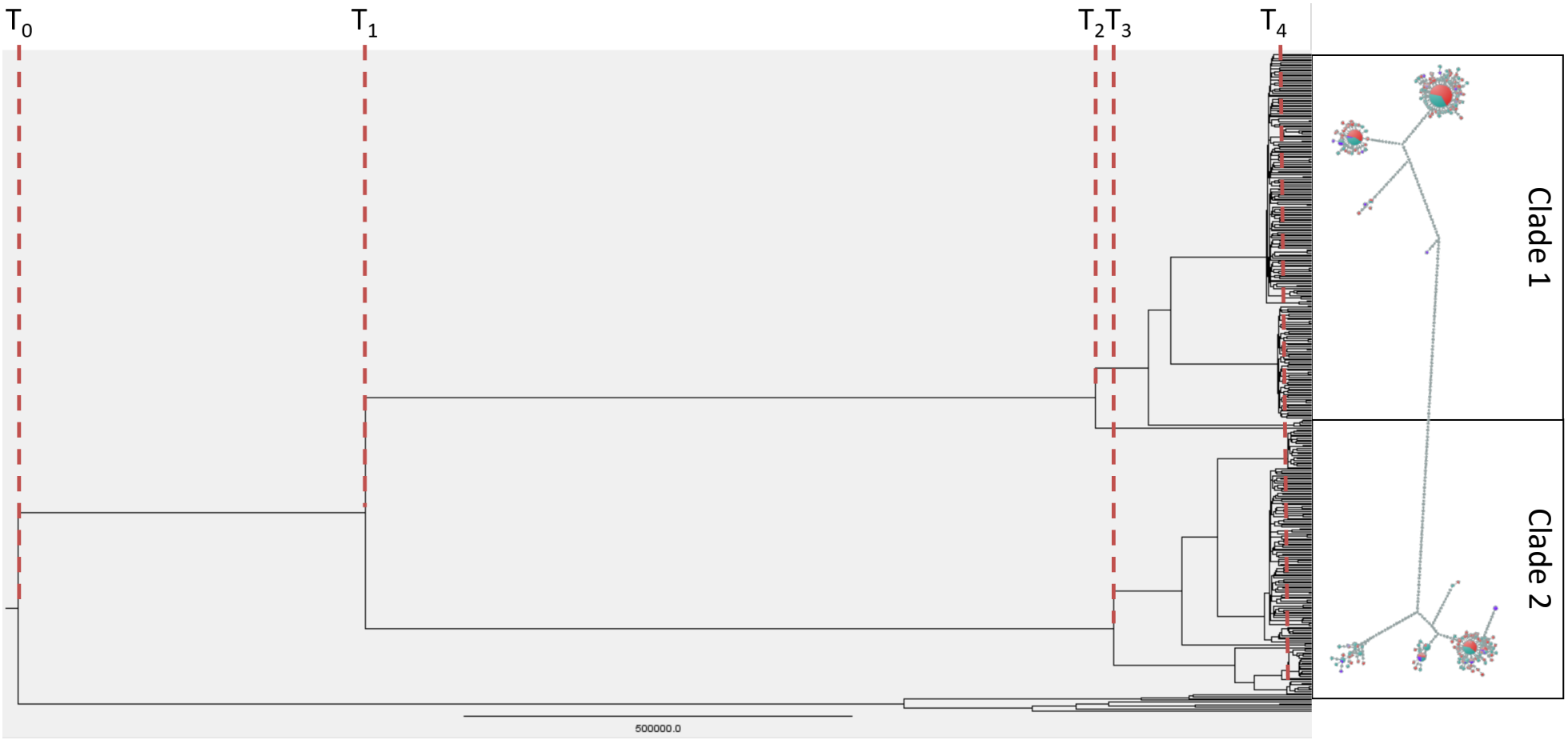

Figure 3.6: Phylogenetic tree produced with TreeAnnotator. See Table 3.2 for time estimates for each node. This figure is produced using the Yule model using modern, ancient, and outgroup samples. Although trees that were generated under the other models were exactly the same, the older nodes were consistently resolved, and thus shown in a single figure. T_0 was not estimated for models without an outgroup. The haplotype network is the unmodified version of Figure 3.5.

3.3.6. Demographic history

Bayesian skyline plots show an exponential increase in population size approximately 10,000 years ago based on modern data, but this same pattern is not detected when only the ancient DNA samples are used (Figure 3.7). The finding of an exponential increase is also supported by the exclusion of zero from 95% credible interval of exponential growth parameter of the model (mean = $3.7238E^{-6}$, CI95% = $1.3998E^{-6}$ – $6.2791E^{-6}$). The same signal of expansion was observed in the ancient nuclear genome using the PSMC analyses (Figure 3.8), with a recent effective population size of $4.5e^{8}$. Here, the exponential expansion was dated to approximately 20,000 years ago. The large exponential expansion was not observed from the PSMC analyses of modern samples. One modern sample (CA115004) showed a recent effective population size a full factor higher ($5.4e^{5}$) compared to all other modern samples ($1.8E^{4}$ – $4.8E^{4}$). There is at most a 750-year difference between the modern samples and the ancient sample where the discrepancy between the population size trends could have been generated. These population size changes should be interpreted with extreme care as estimates younger than 20,000 years are considered unreliable (Li & Durbin, 2011; Mather et al., 2020).

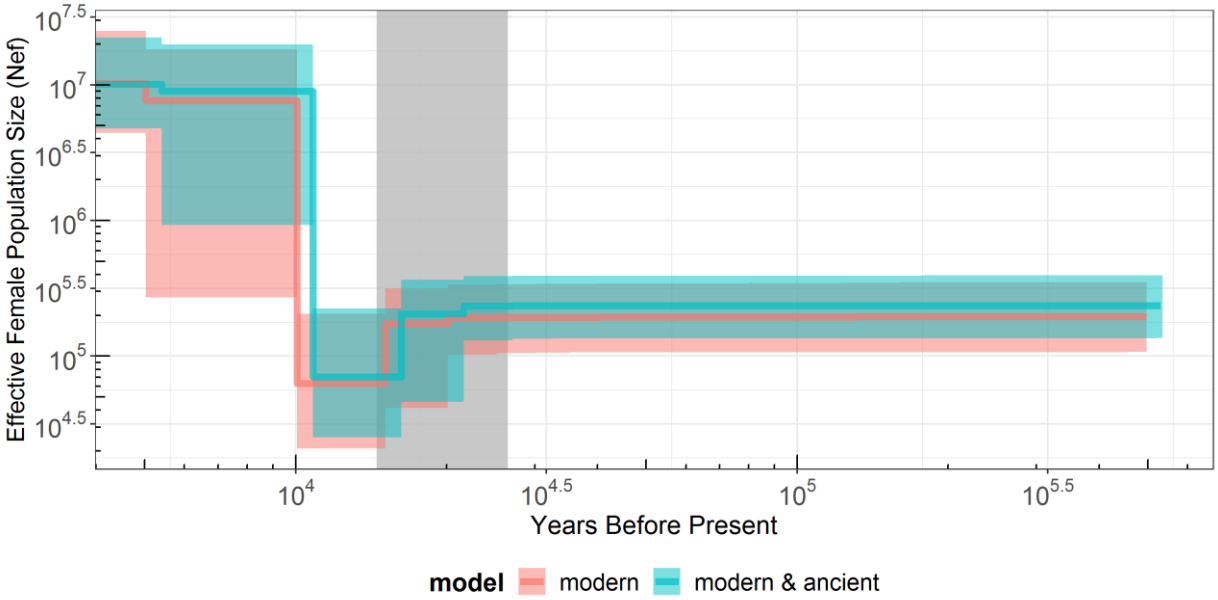

Figure 3.7: Bayesian skyline plot based on the full mitochondrial genome of snapper. Solid lines shows the mean estimated effective population size, shaded areas show 95% credible interval. The shaded areas in grey shows the last glacial maximum between 14,500 and 26,500 years ago.

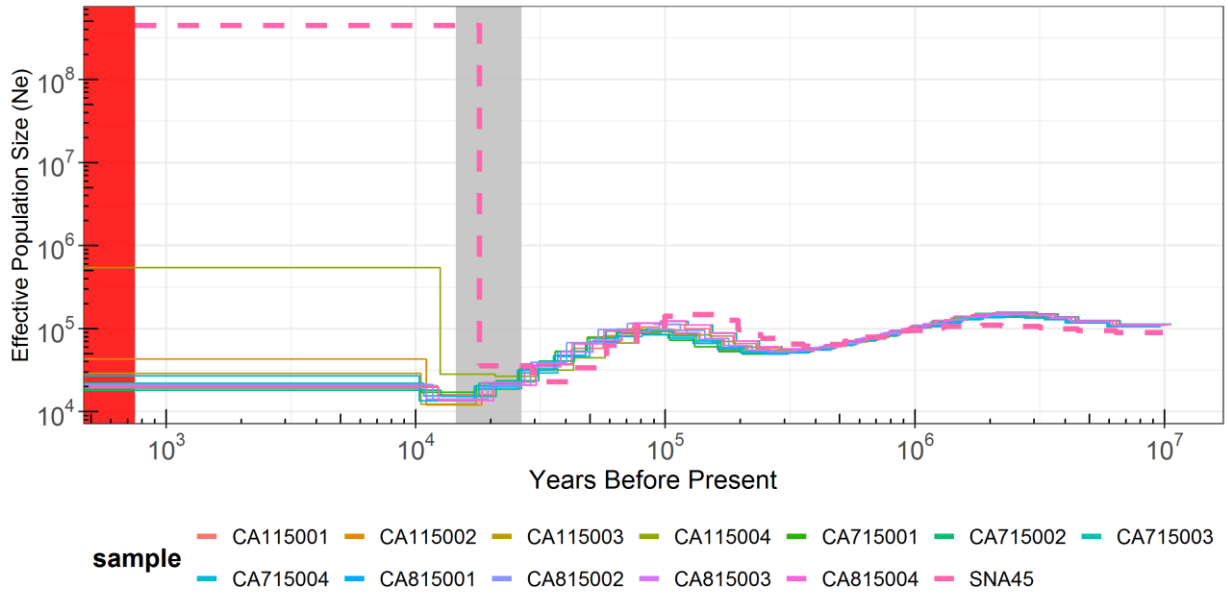

Figure 3.8: Effective population size change over time using PSMC. Solid lines show 12 modern individuals, dashed line shows one ancient individual. The shaded area in red shows the 750 years between modern and ancient samples. . The shaded areas in grey shows the last glacial maximum between 14,500 and 26,500 years ago.

3.4. Discussion

3.4.1. Mitogenomic variation and phylogeographic structure

The analyses of snapper mitochondrial genomes show that they have a high level of nucleotide diversity when compared to other fish (π = 0.005) (Table 3.1). Although direct comparisons are limited, because of the low number of studies available that have reported whole mitochondrial genome nucleotide diversity. One similar study on Atlantic cod reported a nucleotide diversity of 0.002 (Martinez-Garcia et al., 2021; in review). The relative high nucleotide diversity in snapper compared to other species was also observed using mitochondrial control region data (Figure S3.8) (Papa et al., 2021). High levels of genetic diversity also suggest populations have been large for a long period. Levels of variation were maintained despite large reductions in abundance due to intense fishing effort. However, if bottlenecks are short of duration genetic drift won't be able to remove genetic variation. Low levels of genetic drift are also inferred by the absence of temporal change in the diversity or types of haplotypes of either clade over the last 750 years (p-value 0.723).

Tajima's D estimates were heavily influenced by the high number of sites with missing sites, and resulted in an overestimation of the number of rare variants (Table 3.1). This resulted in negative estimates of Tajima's *D* suggesting either a recent selective

sweep, or a period of population size expansion. Other diversity estimates were only minimally affected by missing data in the DNA sequences from the bone samples, and differences between estimates are likely caused by a small bias introduced by missing data, and differences in genetic diversity between the modern and ancient samples (Table 3.1).

The hierarchical AMOVO showed there was no significant genetic differences between sampled sites (p-value: 0.764), which is supported by the distribution of haplotypes around New Zealand (Figure 3.1). This is in contrast to the fine-scale population structure detected using whole-genome data (Chapter 2), but consistent with previous studies that assessed population genetic structure using mitochondrial markers (Bernal-Ramírez et al., 2003).

The high nucleotide diversity found for snapper is primarily the result of the deep split between two major mtDNA clades that occurred around 650,000 years ago (Figure 3.5). Similar distinct genetic differences were not detected in the nuclear data (Chapter 2), which means the mtDNA lineages do not represent separate populations or cryptic species. Here, recombination has broken up the signal of the separate lineages. This shows the value of mitochondrial genomes for observing long-term population dynamics. The observed phylogeographic structure is consistent with recent admixture of previously separated populations (Avise et al., 1987). Similar patterns can be observed in the Indo-Pacific region where glacial cycles caused sea levels to drop by 130 meters separating the two oceans (Bowen et al., 2016). Species that were separated on either side of the Indo-Pacific barrier, diverged into populations containing distinct mtDNA lineages. Currently, a mixture of these lineages has been reported in regions where secondary contact has most likely occurred (Craig et al., 2007; Duncan et al., 2006; Gaither et al., 2011). Similarly, snapper may have experienced a barrier to gene flow that led to the two distinct mtDNA lineages. The time estimate of the split between the lineages is consistent with the separation of two lineages in Chinese four-eyed sleeper (*Bostrychus sinensis*), which was 437,343 years ago (CI 241,000–671,000) (Qiu et al., 2016). Here, reduced sea levels during the Pleistocene glacial cycles are thought to have created a land bridge between Taiwan and China (Railsback et al., 2015), creating two distinct mitochondrial lineages. In snapper, it is unclear how populations got spatially

separated. Both mitochondrial clades have been detected in snapper populations in Australia using mitochondrial control region data (unpublished data, Figure S3.6). One possibility is that the clades diverged on opposite sides of the Tasman Sea and later merged again during secondary contact. Another possibility is that the two clades diverged in New Zealand and later dispersed to Australia. Another possibility is that the two clades diverged in Australia. If so, a possible barrier could have been a land bridge between Australia and Tasmania during the last glacial maximum (LGM) (Williams et al., 2018). However, as the age of the split does not significantly differ from the time estimate when *C. auratus* split from *P. major* diverged, it is possible that the two clades diverged in the Indo-Pacific. The Indo-Pacific is known for facilitating phylogeographic structuring during glacial periods (Hewitt, 2000).

Over the last 140,000 – 170,000 years, six lineages diverged from each other creating distinct clades (Figure 3.6 & Table 3.2), represented by the six minor clades in haplotype genealogy (Figure 3.5). It is possible that snapper experienced multiple range expansions and contractions throughout Pleistocene glacial-interglacial cycles as the available habitat and sea temperatures fluctuated over this period (Barrows et al., 2007). In these situations, populations are often restricted to isolated pockets of suitable habitat. The disruption to gene flow over an extended period of time can result in disjunct populations that form very different compositions of mtDNA diversity. Similar processes have been shown to facilitate phylogeographic structuring in New Zealand native terrestrial species (Wallis & Trewick, 2009). Fluctuating sea temperatures over the last 140,000 years could have caused range contractions in snapper (Barrows et al., 2007), creating isolated populations that generated genetically distinct lineages. Also, changes in sea level have likely played an important factor in the fragmentation of populations (Dlabola et al., 2015; Dohner et al., 2018; Fleming et al., 2018). While population fragmentation facilitated by geological processes over the last 140,000 years could have generated the six genetic lineages, it should be noted that genetic differentiation in large populations can occur in the absence of geographic isolation (Hogner et al., 2012; Webb et al., 2011).

3.4.2. Divergence time estimate between *C. auratus* and *P. major*

The time to the last known common ancestor between *C. auratus* and *P. major* was estimated using the Yule prior to be 727,000 (95%CI 591,000-865,000) years before present (Figure 3.6 & Table 3.2). This estimate is in contrast with a previous estimate ranging between one and three million years inferred from Tabata and Taniguchi (2000). When using the same substitution rate from Tabata and Taniguchi (2000) (0.6 and 1.4% per million years), and D_A obtained using mitochondrial control region sequences from this study (4.08%), this study also finds a divergence time between 1.5 - 3.4 million years. However, when applying the substitution rate used in the BEAST2 analyses (5.0% per million years), and the D_A obtained using mitochondrial control region sequences from this study (4.08%), a divergence time of 816,000 years is obtained. This last estimate falls within the 95% credible confidence interval of the Yule model: 727,000 (95%CI 591,000-865,000). This shows that the discrepancy in divergence time estimate between Tabata et al, and this study is primarily caused by the difference in substitution rate. The substitution rate used in this study is 3.5 to 8.3 times faster compared to Tabata and Taniguchi (2000). Verifying substitution rates is extremely difficult and estimates of divergence time should always be used with caution. The mutation rate for the control regions used in this study is based on Bowen et al. (2006). In this study the authors compared the substitution rates from 10 different fish species which ranged between 2.2% and 100% per million years, with the majority of estimates ranging between 2.2% and 10%. Because there are no estimates of substitution rates for snapper or close relative, we argue that using an averaged substitution rate of 5% per million years based on the comparative analyses from Bowen et al. (2006) is justified. An argument for choosing a faster mutation rate when looking at recent demographic changes is that divergence rates ignore mutations that were not fixed or removed from the population. This can lead to an underestimating of the mutation rate on evolutionary short time scales. Thus, under the assumption that the substitution rate applied in this study is in the upper range of possible substitution rates for snapper, the two species have a minimum divergence time estimate of 727,000 years (591,000-865,000).

3.4.3. Exponential population expansion after the last glacial period

Demographic analyses of population size changes over time show an exponential population expansion that correlates with the timing of the LGM, estimated to have been between 19,000 and 29,000 years ago (Blunier & Brook, 2001; Clark et al., 2009; Pahnke et al., 2003). The non-zero exponential growth rate (mean = $3.7238e^{-6}$, CI95% = $1.3998e^{-6}$ – $6.2791e^{-6}$) shows the exponential model is statistically different from a constant population size. Increasing population size results in weaker genetic drift allowing more variation to be retained. This pattern is seen with the increase in mitochondrial lineages between 7,000 and 35,000 years ago, similar to the change in population size seen in the Bayesian skyline plots (Figure 3.7), and the PSMC analyses on the ancient genome (Figure 3.7).

The skyline plots estimated in BEAST show an exponential increase in population size around 10,000 years ago, close to Pleistocene-Holocene Boundary 11,650 years ago (Figure 3.7) (Walker et al., 2009). No changes in population size were detected before the last glacial maximum (LGM). These early histories of constant population sizes are likely caused by the loss of genetic variation through drift during the LGM when populations were likely fractured and much smaller (Grant et al., 2012). Using the nuclear DNA sequence data, only the analysis based on one ancient genome was able to detect the exponential population increase approximately 20,000 years ago (Figure 3.7). Population size changes that occurred over the last 20,000 years are considered unreliable and should be interpreted with extreme care (Li & Durbin, 2011; Mather et al., 2020). Additional analyses are required to test for more recent population size changes using Sequential Markovian Coalescence frameworks like PSMC.

One interpretation of the PSMC results is that genetic diversity has been lost over the last 750 years. Genetic diversity is lost when a populations experience a decline in abundance and the effect of genetic drift increases, removing genetic diversity. It is unclear what environmental factors would have contributed to the loss of genetic variation, however, intense fishing pressure is one likely candidate (Fisheries New Zealand, 2018). Comparable levels of heterozygosity between the ancient and modern genomes suggest genetic diversity has not been lost. Drastic population declines due to

intense fishing also occurred relatively recent (since the 1950s), which may not have been enough time for genetic drift to remove variation from the population. A recent study on Atlantic cod (*Gadus morhua*) has shown that there is no empirical evidence for substantial loss of genetic variation due to intense fishing effort (Pinsky et al., 2021).

Two environmental changes likely played important roles in the expansion of snapper after the last glacial cycle. First, rising sea levels would have extended the available habitat on the continental shelf. With sea levels having risen over 120 meters over the last 20,000 years (Dlabola et al., 2015; Peltier & Fairbanks, 2006), a large proportion of the continental shelf would have become submerged. Snapper most commonly occurs in waters up to 50 meters, with 200 meters considered as their maximum depth (Parsons et al., 2014). With the edge of the continental shelf running relatively close to New Zealand shores, it implies that suitable habitat during the LGM for snapper would have been limited and most likely fragmented. This can be shown using a rough estimation of the suitable habitat using the current bathymetric line of 120 meters around New Zealand (Figure S3.12). The reduction and fragmentation of habitat during glacial cycles have also been linked to the genetic structuring and demographic changes in leafy sea dragons (*Phycodurus eques*, *Syngnathidae*) along the South Australian coast (Dolby, 2021; Stiller et al., 2020). Second, the post-LGM increase in average sea temperature around New Zealand by approximately 6°C (rising from ~10°C to ~16°C) would have expanded the available habitat for snapper southwards (Barrows et al., 2007). With sea surface temperatures expected to rise with 0.8-2.5 degrees in New Zealand by 2100 (Law et al., 2018), snapper will likely continue to be able to expand their distribution down the coast of the South Island. This is supported by ecological niche modelling of snapper distribution under different climate scenarios (Brooks, 2019).

3.4.4. Variation in the timing of population size estimates

The discrepancy in the timing of the exponential expansion of snapper between the mitochondrial and nuclear data can likely be attributed to inaccuracies in the mutation rates (Figure 3.7 & Figure 3.7). Mutation rates are difficult to estimate and very small inaccuracies can already shift the time estimates by thousands of years. While it is not possible to attribute the population expansion to the environment conditions at

the end of the last glacial cycle with a high degree of confidence, it is the most common climate event that happened during this period, which has been linked to population expansions for a range of marine taxa (Jenkins et al., 2018).

The estimate of the most recent effective population size based on the mitochondrial data was much lower compared to the PSMC results based on the ancient nuclear genome. Lower estimates compared to those from nuclear DNA sequences are expected from mitochondrial markers because the maternally inherited haploid genome has a quarter the N_e when compared to a diploid locus on the nuclear genome. Even after accounting for the difference in effective population size between nuclear ad mitochondrial markers, estimates still differed by an order of magnitude. Second, several assumptions are commonly violated when using genetic variation to estimate the effective population sizes. Waples and Yokota (2007) showed that overlapping generations can lead to an overestimation of effective population size.

It is even more difficult to extrapolate census population sizes (N_c) from effective population size estimates. Snapper are serial broadcast spawners (Parsons et al., 2014). Some reproducing individuals contribute a disproportional number of offspring each season. These reproductive sweepstakes can produce a significant difference between the effective and census populations sizes (N_e/N_c) with ratios as small as 10^{-3} (Waples, 2016). Such heavily skewed reproductive output was observed on captive snapper (Ashton et al., 2019a). In contrast, a recent study reported a N_e/N_c ratio of 0.33 in snapper in the Hauraki Gulf (Jones et al., 2019), suggesting that the reproductive sweepstakes in the species are negligible. It is possible that after bottlenecks a large proportion of the population can reproduce due to lack of competition. If so, current-day N_e/N_c ratios inferred from recovering populations will likely not accurately reflect N_e/N_c ratios from pre-exploited populations. In summary, it is impossible to provide accurate census population size estimates from pre-exploited populations based on effective population size estimates.

3.4.5. Considerations for future research & conclusion

Poor sample preservation was the main factor that limited the successful recovery of DNA sequences from the preserved bone material. The brittle and porous

nature of fish bones, combined with harsh preservation conditions in coastal areas would damage and degrade DNA (Oosting et al., 2019). The low sample sizes made it difficult to determine any common attributes of sites that may provide better preservation conditions samples. Other studies have also reported difficulty in predicting whether a site can provide the right conditions for yielding high levels of endogenous DNA (Ferrari et al., 2021; Keighley et al., 2021). Cold and dry environments have been shown to give the best results (Bollongino et al., 2008; Ferrari et al., 2021). Local knowledge about common practices used by Māori to prepare food could potentially be a useful approach to studying sample variation among sites (Leach, 2006). Fish exposed to high heat would have been a major source of degradation compared to cultural practices where fish were consumed raw (Oosting et al., 2019). The unusual bone sample from Hotwater Beach with an endogenous DNA content of 58% and enabled a full genome at a depth of 13x to be obtained, yet it is unclear what factors could have contributed to the exceptional quality of this sample.

The lower average coverage of the ancient mitochondrial genomes compared to modern mitochondrial genomes resulted in more nucleotide positions with missing nucleotides (Table 3.1). This can be an issue if downstream analyses entirely remove nucleotide positions with missing data. This could be observed in Tajima's D estimates. Here, diversity estimates were underestimated because many polymorphic sites were removed due to high levels of missing nucleotides in ancient samples. This reduced the average number of pairwise differences between haplotypes, creating the appearance of an excess in rare alleles. There was less concern for biased results caused by missing data in BEAST2 analyses. A previous study found that missing data has little effect on phylogenetic inference beyond the loss in statistical power in BEAST2 (Roure et al., 2013). This is also supported by the comparable results between BEAST2 runs including modern sequences only, and runs including both modern and ancient sequences (Figure 3.6 & Table 3.2).

The addition of more ancient mitochondrial and nuclear genomes will provide a better representation of the genetic variation in snapper prior to large declines due to fishing (Oosting et al., 2019). It is likely that the fishing induced bottleneck removed lineages that can provide a more detailed picture of the demographic history of snapper.

It would also allow for direct comparisons of genetic diversity between ancient and modern populations which will provide information on the possible impacts humans have had on snapper. Population size changes based on Sequential Markovian Coalescence have been limited to analyses based on single individuals (PSMC). Additional analyses that utilize the joined signal of nuclear genomes are suspected to provide more power in detecting recent demographic changes. Based on the same statistical framework, programs such as MSMC, smc++, and MAGIC (Schiffels & Durbin, 2014; Terhorst et al., 2017; Weissman & Hallatschek, 2017), provide multi-sample approaches, giving more power to detect more recent population size changes (Mather et al., 2020).

The findings reported in this chapter of snapper demographic history suggests Pleistocene glacial cycles played an important role in their population history. Distinct phylogenetic lineages potentially reflect once isolated populations caused by range contraction, creating barriers to gene flow. The end of the last glacial cycle resulted in an exponential population expansion, likely facilitated by increased ocean temperatures and suitable habitat. Current climate change is likely to continue the expansion of suitable snapper habitat, as is supported by ecological niche modelling (Brooks, 2019). Much is known about how glacial periods fragmented terrestrial habitats in New Zealand and facilities evolutionary change (Wallis & Trewick, 2009). This study adds to our understanding of how geological changes have influenced the evolution of New Zealand's biodiversity.

3.6. Supplementary materials

3.6.1 Supplementary Tables

Table S3.1: Sampling information: ancient samples per location.

location	Map name	Period	Sampled	Extracted	Sequenced	Full Mito	Latitude	Longitude
Aupouri	Aupouri	Late	4	0	0	0	-34.7512	173.0488
Breaksea Sound	Breaksea Sound	Late	1	0	0	0	-45.5771	166.7096
Cooper island	Cooper Island	Early	1	0	0	0	-45.7317	166.8417
Cross Creek	Cross Creek	Early	10	8	8	5	-37.1738	175.8451
Davidson undefended site	Davidson undf.	Early	8	0	0	0	-36.7774	175.1128
Foxton	Foxton	Early	10	2	2	0	-40.486	175.2268
Hararonga West	Hararonga W.	Early	9	0	0	0	-36.2436	175.4669
Hotwater Beach	Hotwater Beach	Early	10	10	10	4	-36.8365	175.7893
Kokohuia	Kokohuia	Late	10	3	3	0	-35.5262	173.3921
Mana Island North	Mana Island N.	Late	10	10	10	4	-41.0788	174.7849
Mana Island South	Mana Island S.	Early	11	7	7	4	-41.0963	174.7736
Panau Banks	Panau Banks	Late	2	1	1	0	-43.8273	172.9834
Papatowai	Papatowai	Early	1	0	0	0	-46.4458	169.7813
Parewanui	Parewanui	Late	1	0	0	0	-40.0219	175.095
Rotokura	Rotokura	Early	10	8	8	4	-41.376	173.3097
Sunde	Sunde	Early	10	9	9	5	-36.925	174.7265
te Ika a Maru	te Ika a Maru	Late	6	0	0	0	-41.1741	174.748
the Glen	the Glen	Early	10	2	2	0	-41.288	173.7474
Total	-	-	124	60	60	26	-	-

Notes: Period indicates from what time period each originates, Early = 1280-1450 CE, Late = 1650-1800 CE

Table S3.2: nuclear mutation rates reported in other species.

Organism	Family	reference	Mutation rate ($site^{-1}/year^{-1}$)
Mudskipper	Gobiidae	You et al. (2014)	$3.51e^{-9}$
Malawi cichlid	Cichlidae	Malinsky et al. (2018)	$1.17e^{-9*}$
Icefishes	Channichthyidae	Kim et al. (2019)	$3.28e^{-9}$
haddock	Gadidae	Tørresen et al. (2018)	$5.0e^{-9}$
Cod	Gadidae	Tørresen et al. (2018)	$6.0e^{-9}$
Tanaka's snailfish	Liparidae	K. Wang et al. (2019)	$6.77e^{-9}$

Notes: *assuming a generation time of 3 years (Hey et al., 2004)

Table S3.3: Information for all sequenced individuals.

Sample ID	Sampling Site	Period	FMA (SNA)	Genetic Cluster	Paired Reads	Unique Reads	Endogenous DNA (%)	Mitochondrial		Mitogenome Obtained
								Hits	Coverage	
SNA10	Cross Creek	Early	SNA1	North	38678612	1981633	5.9	5518	16.40	Yes
SNA11	Cross Creek	Early	SNA1	North	2880413	60825	2.4	243	0.65	-
SNA12	Cross Creek	Early	SNA1	North	5689836	151428	2.9	495	1.30	-
SNA13	Cross Creek	Early	SNA1	North	4496032	7976	0.2	61	0.12	-
SNA14	Cross Creek	Early	SNA1	North	3386255	204210	7.0	282	0.78	-
SNA15	Cross Creek	Early	SNA1	North	7301215	400549	6.1	2014	5.37	Yes
SNA15	Cross Creek	Early	SNA1	North	18659647	916027	5.3	1762	5.02	Yes
SNA16	Cross Creek	Early	SNA1	North	12419040	1296939	11.5	2755	7.55	Yes
SNA9	Cross Creek	Early	SNA1	North	133066938	6029486	5.4	18467	52.70	Yes
SNA30	Foxton	Early	SNA8	South	3679519	2435	0.1	10	0.03	-
SNA33	Foxton	Early	SNA8	South	2747002	2240	0.1	36	0.21	-
SNA44	Hotwater beach	Early	SNA1	North	24773132	624139	2.8	2196	7.41	Yes
SNA45*	Hotwater beach	Early	SNA1	North	372908223	199346350	58.7	23408	55.75	Yes
SNA46	Hotwater beach	Early	SNA1	North	2010313	51505	2.9	103	0.32	-
SNA47	Hotwater beach	Early	SNA1	North	3297524	41321	1.5	114	0.29	-
SNA48	Hotwater beach	Early	SNA1	North	24701547	194895	0.9	3060	7.12	Yes
SNA49	Hotwater beach	Early	SNA1	North	30428183	578989	2.2	4576	13.77	Yes
SNA50	Hotwater beach	Early	SNA1	North	6499055	80635	1.4	372	0.89	-
SNA51	Hotwater beach	Early	SNA1	North	4262241	418	0.0	2	0.005	-
SNA52	Hotwater beach	Early	SNA1	North	92655	2273	2.5	1	0.003	-
SNA53	Hotwater beach	Early	SNA1	North	2888986	34776	1.4	118	0.30	-

SNA56	Kokohuia	Late	SNA8	Centre	3766975	59217	1.8	203	0.61	-
SNA61	Kokohuia	Late	SNA8	Centre	4605548	38795	0.9	138	0.37	-
SNA63	Kokohuia	Late	SNA8	Centre	7755064	76759	1.1	201	0.57	-
SNA64	Mana Island North	Late	SNA8	South	1891275	19673	1.2	112	0.31	-
SNA65	Mana Island North	Late	SNA8	South	13655288	429555	4.2	2363	5.70	Yes
SNA66	Mana Island North	Late	SNA8	South	1530197	9710	0.6	25	0.08	-
SNA67	Mana Island North	Late	SNA8	South	4579701	93716	2.3	260	0.73	-
SNA68	Mana Island North	Late	SNA8	South	3698487	7625	0.2	57	0.11	-
SNA69	Mana Island North	Late	SNA8	South	104639	485	0.5	1	0.003	-
SNA70	Mana Island North	Late	SNA8	South	3223360	896	0.0	6	0.01	-
SNA71	Mana Island North	Late	SNA8	South	14039751	177517	1.4	1675	5.47	Yes
SNA72	Mana Island North	Late	SNA8	South	5119625	1358703	29.8	4019	10.51	Yes
SNA73	Mana Island North	Late	SNA8	South	13849932	216624	1.8	4629	14.06	Yes
SNA76	Mana Island South	Early	SNA8	South	1532310	6077	0.5	53	0.10	-
SNA77	Mana Island South	Early	SNA8	South	7687576	1842305	26.2	3229	10.12	Yes
SNA79	Mana Island South	Early	SNA8	South	957430	14501	1.7	47	0.12	-
SNA80	Mana Island South	Early	SNA8	South	1749009	36786	2.4	180	0.45	-
SNA82	Mana Island South	Early	SNA8	South	24184799	995073	4.4	2563	7.24	Yes
SNA83	Mana Island South	Early	SNA8	South	57434975	599710	1.2	5138	14.87	Yes
SNA84	Mana Island South	Early	SNA8	South	42721889	2071830	5.3	6152	18.86	Yes
SNA89	Rotokura	Early	SNA7	South	11774399	940393	9.1	2124	6.64	Yes
SNA90	Rotokura	Early	SNA7	South	70228402	2632607	4.5	10961	30.87	Yes
SNA91	Rotokura	Early	SNA7	South	16339116	617027	4.1	2479	7.40	Yes
SNA92	Rotokura	Early	SNA7	South	3786028	113264	3.4	215	0.58	-
SNA93	Rotokura	Early	SNA7	South	11915518	209697	2.0	483	1.40	-
SNA95	Rotokura	Early	SNA7	South	5223706	171988	3.9	325	0.95	-
SNA96	Rotokura	Early	SNA7	South	471200	68975	14.9	111	0.36	-
SNA97	Rotokura	Early	SNA7	South	11193246	47054	0.5	94	0.22	-
SNA98	Rotokura	Early	SNA7	South	21407951	1965492	10.7	4580	13.86	Yes

SNA101	Sunde	Early	SNA8	Centre	2100282	23289	1.4	263	0.49	-
SNA102	Sunde	Early	SNA8	Centre	3986806	50786	1.4	219	0.65	-
SNA103	Sunde	Early	SNA8	Centre	26178801	543952	2.2	3219	10.28	Yes
SNA104	Sunde	Early	SNA8	Centre	721650	15167	2.1	31	0.09	-
SNA105	Sunde	Early	SNA8	Centre	7136564	85231	1.4	402	1.14	-
SNA106	Sunde	Early	SNA8	Centre	21933381	3150654	15.4	3357	11.12	Yes
SNA107	Sunde	Early	SNA8	Centre	6853490	1087667	17.3	2739	9.18	Yes
SNA108	Sunde	Early	SNA8	Centre	23142566	1191806	5.5	5085	15.90	Yes
SNA99	Sunde	Early	SNA8	Centre	24584789	3722763	16.5	2970	9.05	Yes
SNA116	the Glen	Early	SNA7	South	24888086	11692	0.0	56	0.23	-
SNA117	the Glen	Early	SNA7	South	1773880	4145	0.3	11	0.02	-
SNA124	the Glen	Early	SNA7	South	8904914	27175	0.3	130	0.43	-

Notes:*nuclear genome was obtained for this individual, Early = 1280-1450 CE, Late = 1650-1800 CE

Table S3.4: Substitution models and clock rates for partition used in all BEAST2 analyses.

Partition	Clock rate	Sub model	Kappa 1	Kapp2	Inv	G count	G shape
Codon1+RNA*		TN93+I	35.6662	14.5694	0.845	-	-
Codon2	3.28E-9	HKY	1.25	-	-	-	-
Codon3		TN93	77.0797	14.8292	-	-	-
Control region	5.0E-8	TN93+G	59.2829	25.9365	-	4	0.017

Notes: Sub model=substitution model, Inv=proportion of invariant sites, G count=Gamma category Count, G shape = Gamma shape. *RNA indicates mitochondrial sequences coding for the 12s, 16s and tRNA's

Table S3.5: Results Jomeltest2 – substitution models for each mitochondrial partition. Selected models are highlighted in green.

codon1_RNA										
	AIC			AICc			BIC			
Model	weight	cumWeight		Model	weight	cumWeight	Model	weight	cumWeight	
GTR+I	0.304023	0.304023		TrN+I	0.266745	0.266745	TrN+I	0.737838	0.737838	
TrN+I	0.18573	0.489753		GTR+I	0.239175	0.505919	TrN+G	0.233158	0.970996	TRN+I
GTR+I+G	0.149082	0.638835		TrN+I+G	0.105957	0.611876	TrN+I+G	0.010828	0.981824	
TrN+I+G	0.090138	0.728973		GTR+I+G	0.095899	0.707775	TIM1+I	0.008214	0.990037	
GTR+G	0.088944	0.817917		TrN+G	0.084292	0.792067	TrN	0.003691	0.993728	
codon2										
	AIC			AICc			BIC			
Model	weight	cumWeight		Model	weight	cumWeight	Model	weight	cumWeight	
TrN	0.224769	0.224769		HKY	0.287181	0.287181	HKY	0.887157	0.887157	
HKY	0.15271	0.37748		TrN	0.255719	0.5429	TrN	0.057452	0.944608	HKY
TrN+I	0.082743	0.460222		HKY+I	0.064152	0.607052	HKY+I	0.014413	0.959021	
TrN+G	0.082299	0.542522		TPM1uf	0.063915	0.670967	TPM1uf	0.014359	0.973381	
TIM1	0.082139	0.624661		HKY+G	0.063553	0.734519	HKY+G	0.014278	0.987659	

codon3									
AIC			AICc			BIC			
Model	weight	cumWeight	Model	weight	cumWeight	Model	weight	cumWeight	TRN
GTR	0.34653	0.34653	GTR	0.270706	0.270706	TrN	0.925838	0.925838	
GTR+I	0.176763	0.523292	TrN	0.224306	0.495011	TIM1	0.031696	0.957533	
GTR+G	0.168134	0.691426	TIM1	0.105486	0.600497	TrN+I	0.020699	0.978233	
GTR+I+G	0.064675	0.756101	GTR+I	0.083216	0.683713	TrN+G	0.019606	0.997838	
TrN	0.06321	0.81931	GTR+G	0.079154	0.762867	TIM1+I	0.000701	0.998539	

D-loop									
AIC			AICc			BIC			
Model	weight	cumWeight	Model	weight	cumWeight	Model	weight	cumWeight	TRN+I+G
GTR+I+G	0.515097	0.515097	HKY+I	0.890482	0.890482	TrN+I+G	0.72066	0.72066	
TrN+I+G	0.248232	0.763329	HKY+I+G	0.060158	0.950639	HKY+I+G	0.11684	0.8375	
TIM1+I+G	0.212573	0.975902	TrN+I	0.047361	0.998001	TrN+I	0.091986	0.929486	
GTR+I	0.007358	0.98326	TrN+I+G	0.001162	0.999163	TIM1+I+G	0.051746	0.981232	
TVM+I+G	0.007039	0.990299	TPM1uf+I	0.00082	0.999983	TIM1+I	0.007351	0.988582	

3.6.2. Supplementary Figures

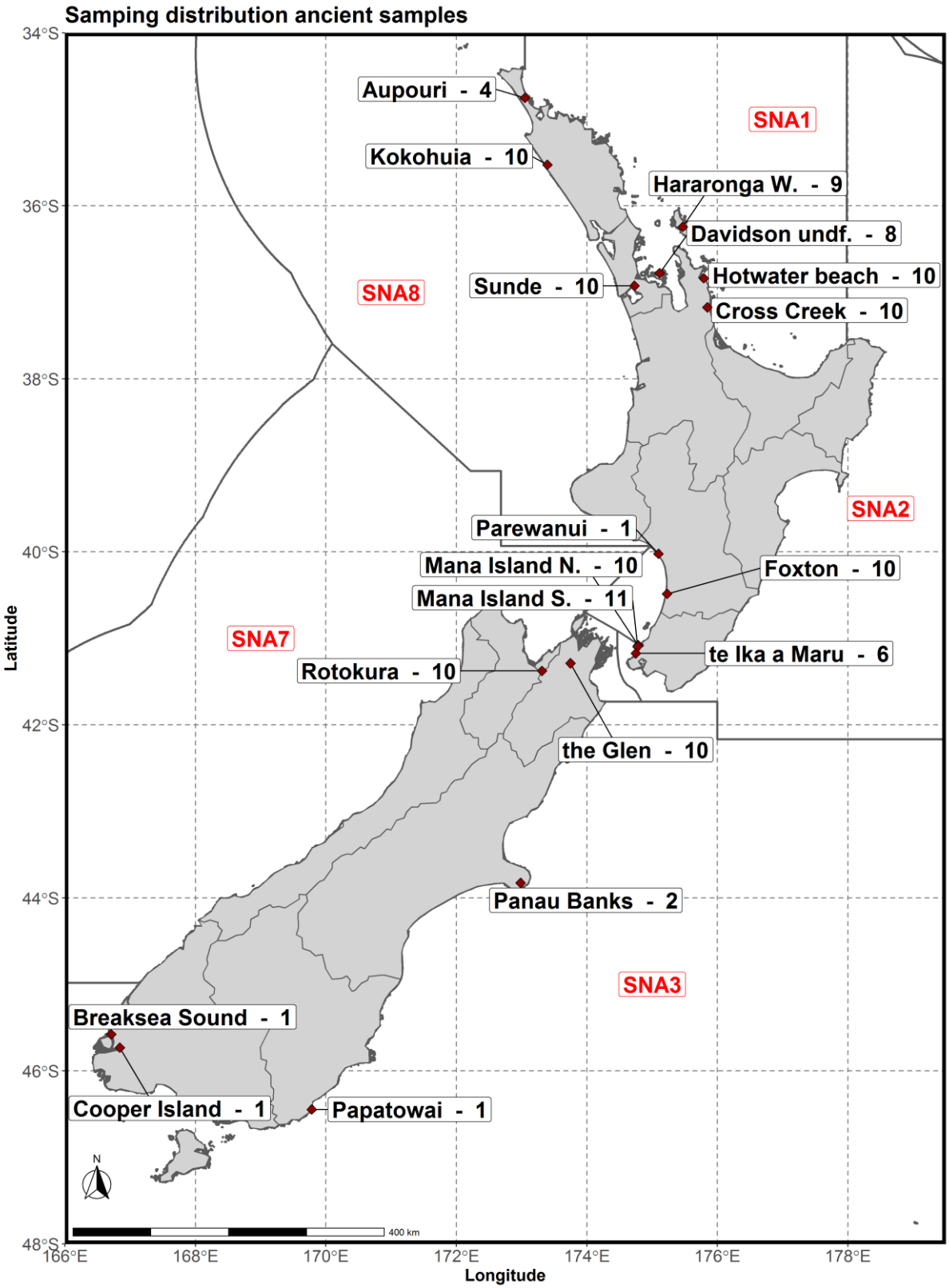

Figure S3.1: map showing the sites from which ancient samples were collected. The number behind the site name indicates the number of dentaries collected from that site. Note that not all sites were used for DNA extraction and sequences. For successful sites see Figure 3.1. SNA# indicate fisheries management areas.

Figure S3.2: Annotation of snapper mitochondrial genome.

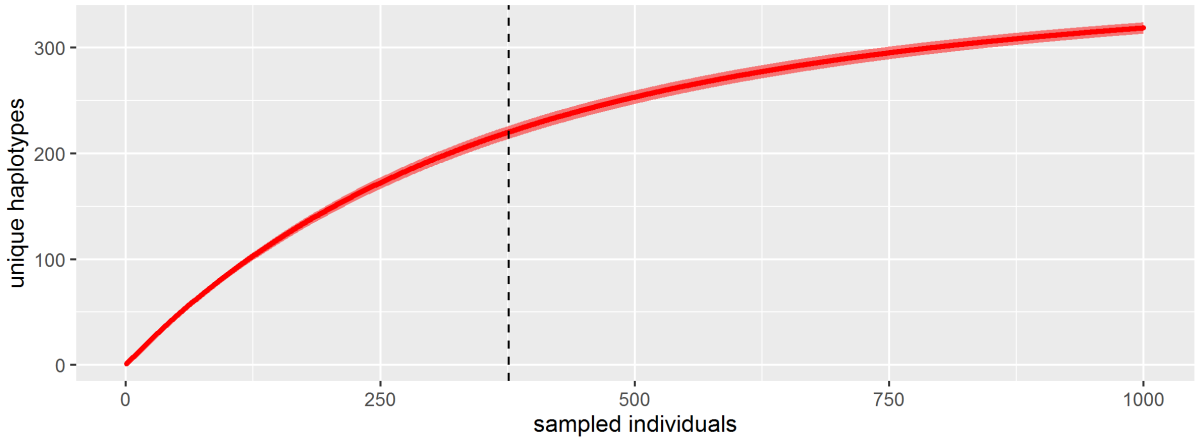

Figure S3.3: Rarefaction curve of mitochondrial haplotypes.

Figure S3.4: Haplotype network based on the mitochondrial control sequences of snapper and P. major.

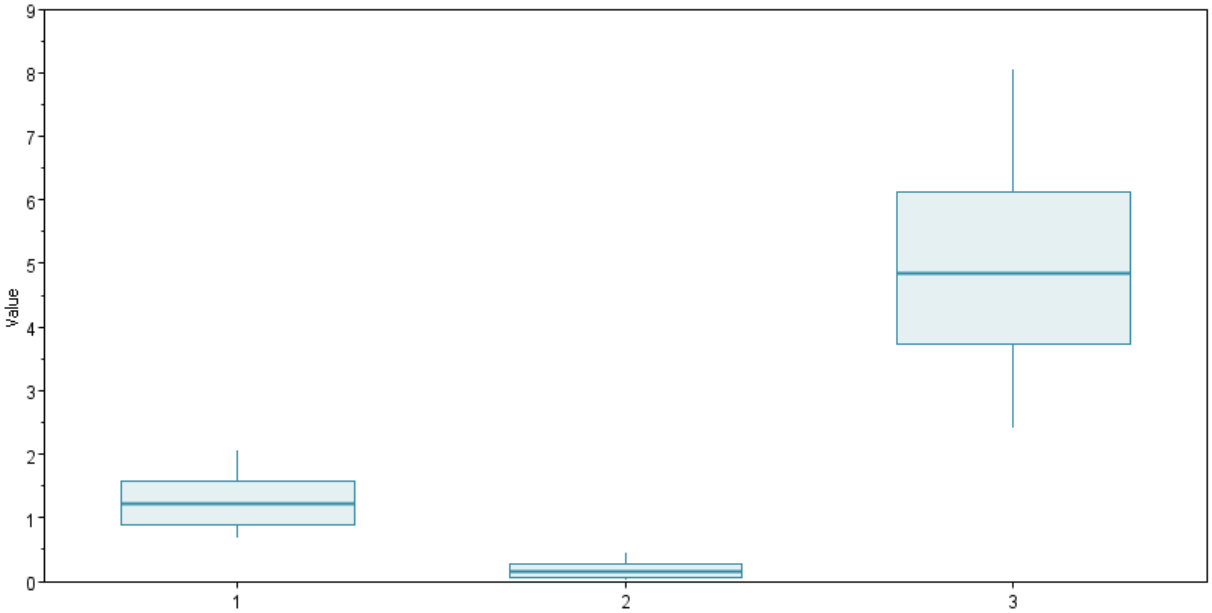

Figure S3.5: relative substitution rates of the 1e 2nd and 3rd codons.

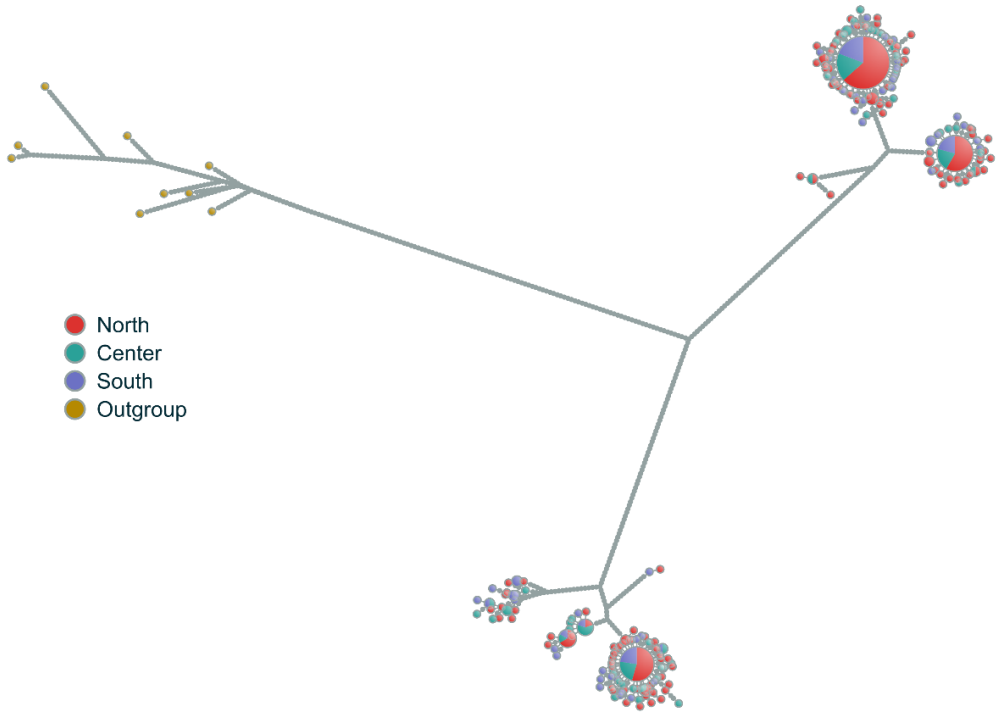

Figure S3.6: Haplotype genealogy showing *C. auratus* coloured by genetic cluster with *P. major* as an outgroup.

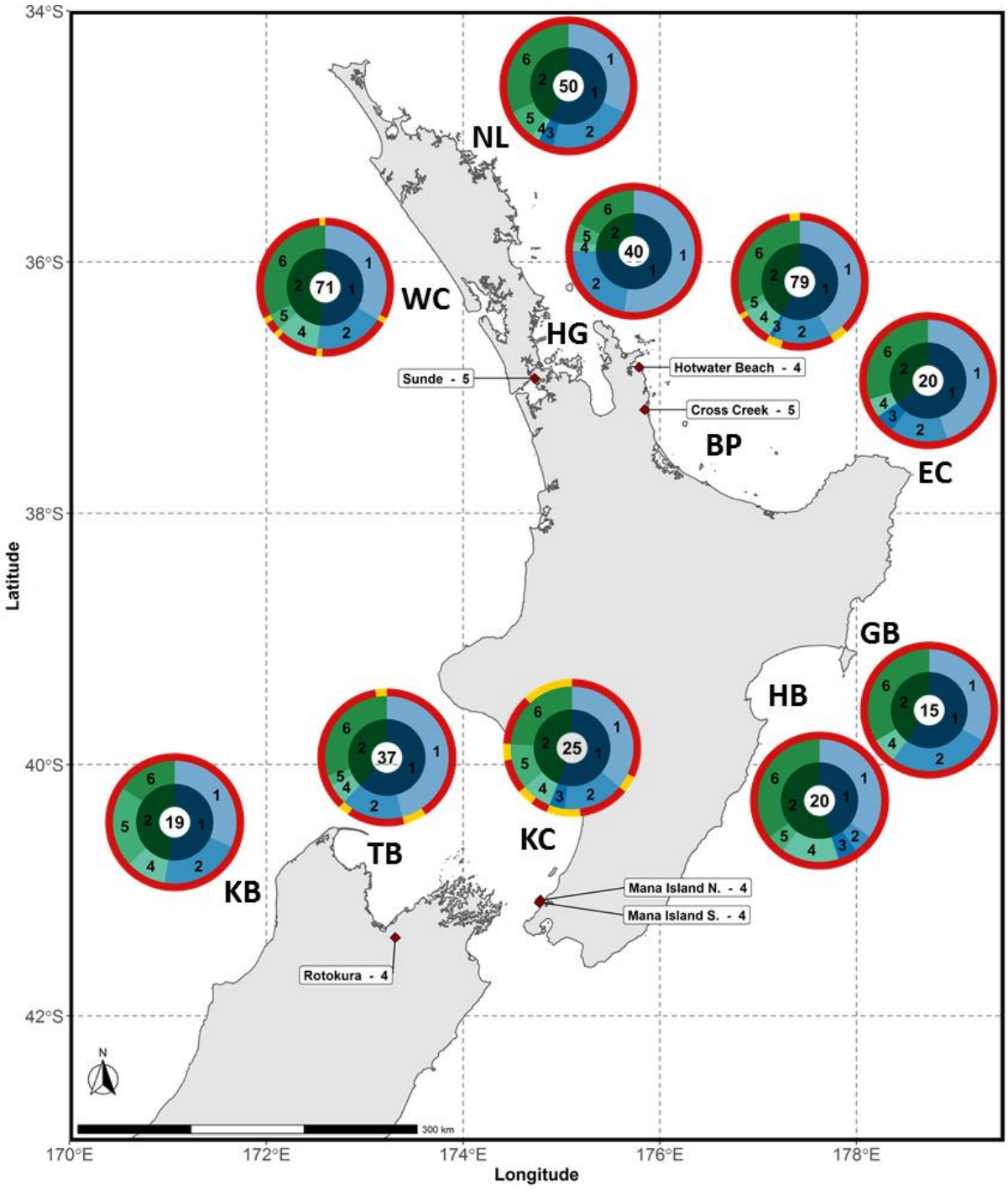

Figure S3.7: Map showing the sampling distribution of both modern and ancient samples. Red diamonds indicate sites where ancient samples were collected, numbers behind site names indicate number of successfully sequenced mitochondrial genomes. See Figure S3.1 for a map showing all ancient sampling sites and Table S3.1 for location information. Pie charts show, sample sizes for each population including ancient samples (centre), proportion of major (inner circle) and minor (middle circle) clades observed for each population (see Figure 3.5), and proportion of modern and ancient samples in red and yellow respectively (outer circle).

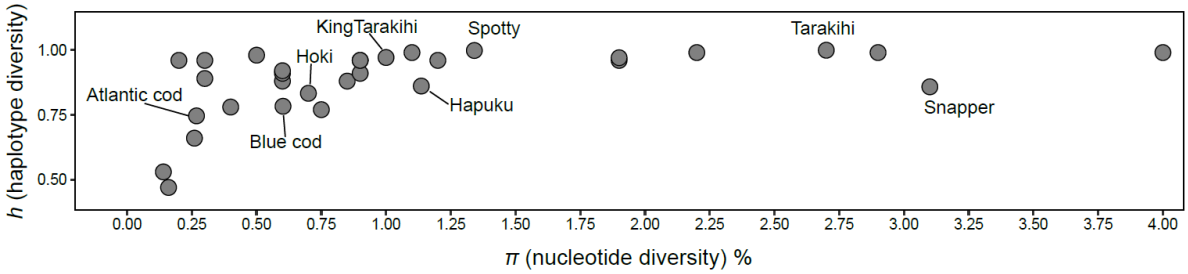

Figure S3.8: Haplotype (h) and nucleotide (π) diversity values of the mtDNA control region in a range of marine fish populations (adapted from Grant and Waples (2000)). Values from the New Zealand species and the Atlantic cod are labelled: Atlantic cod (Gadus morhua), blue cod (Parapercis colias), hāpuku (Polyprion oxygeneios), hoki (Macruronus novaezelandiae), snapper (Chrysophrys auratus), and spotty (Notolabrus celidotus). Source image: Papa et al. (2021).

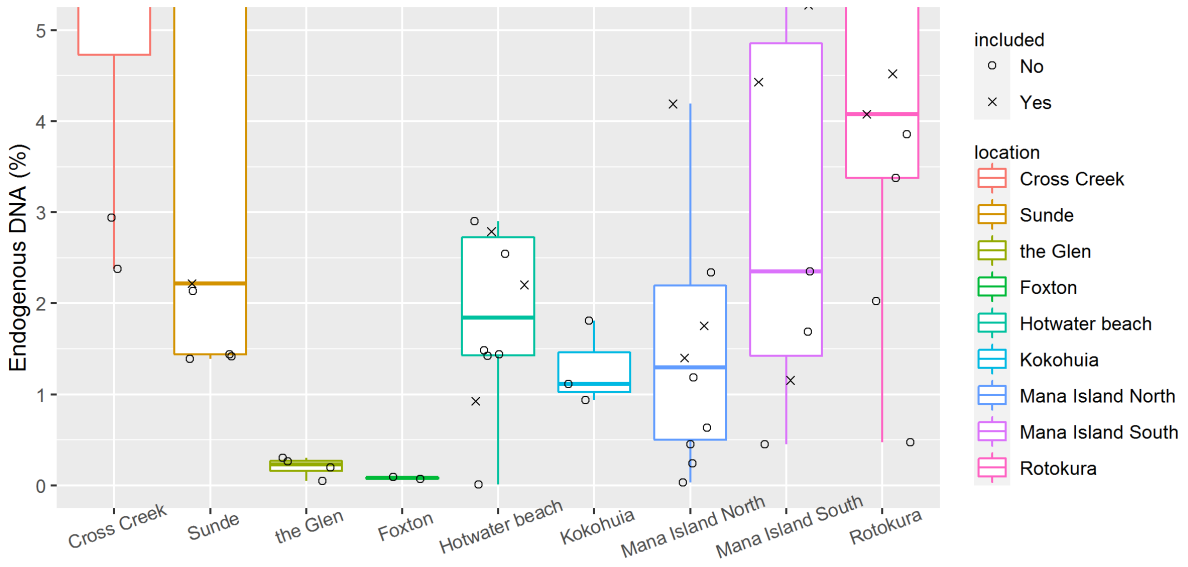

Figure S3.9: cropped version of Figure 3.3. Boxplots showing the levels of endogenous DNA content (%) per site. Black crosses show individual samples that successfully yielded whole mitochondrial genomes. Black circles indicate samples that did not yield whole mitochondrial genomes. Crosses and circles are randomly plotted along the width of the boxplot. The number of samples per area from left to right: 8, 9, 4, 2, 10, 3, 10, 10, 9.

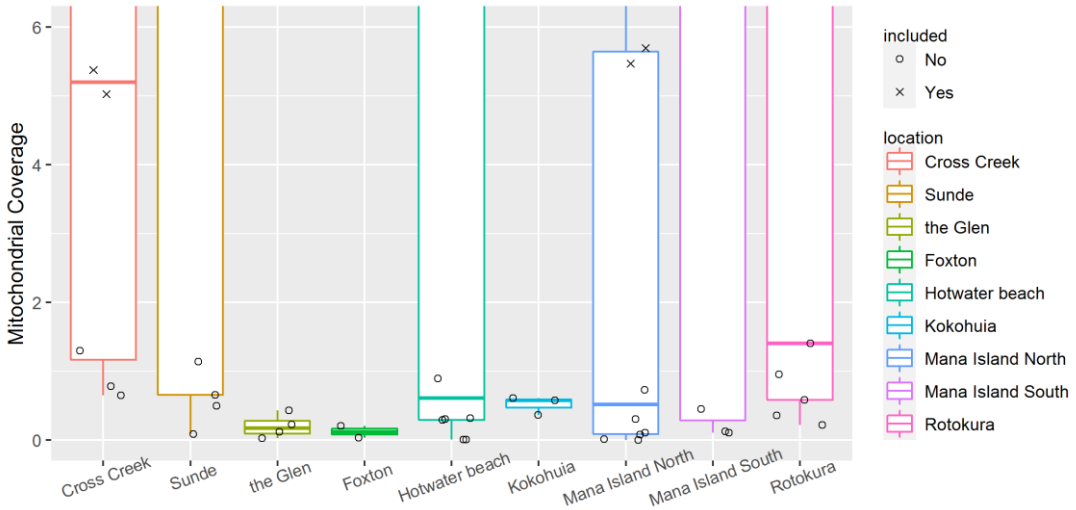

Figure S3.10: cropped version of Figure 3.4 Boxplots showing the mitochondrial coverage of sequenced samples per site. Black crosses show individual samples that successfully yielded whole mitochondrial genomes. Black circles indicate samples that did not yield whole mitochondrial genomes. Crosses and circles are randomly plotted along the width of the boxplot. The number of samples per area from left to right: 8, 9, 4, 2, 10, 3, 10, 10, 9.

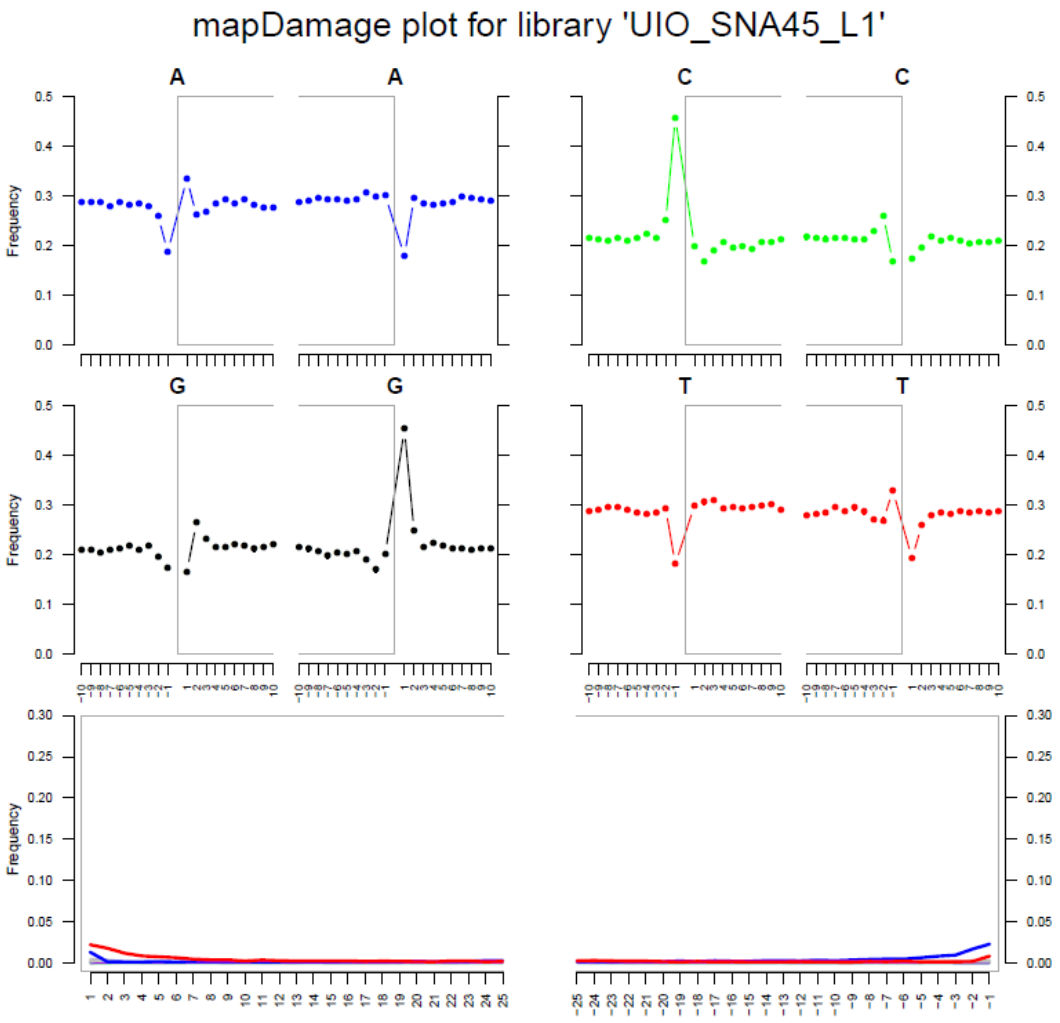

Figure S3.11: MapDamage pattern of the ancient nuclear genome, sample: SNA45. This damage pattern is representative for all samples that yielded complete mitochindrial genomes.

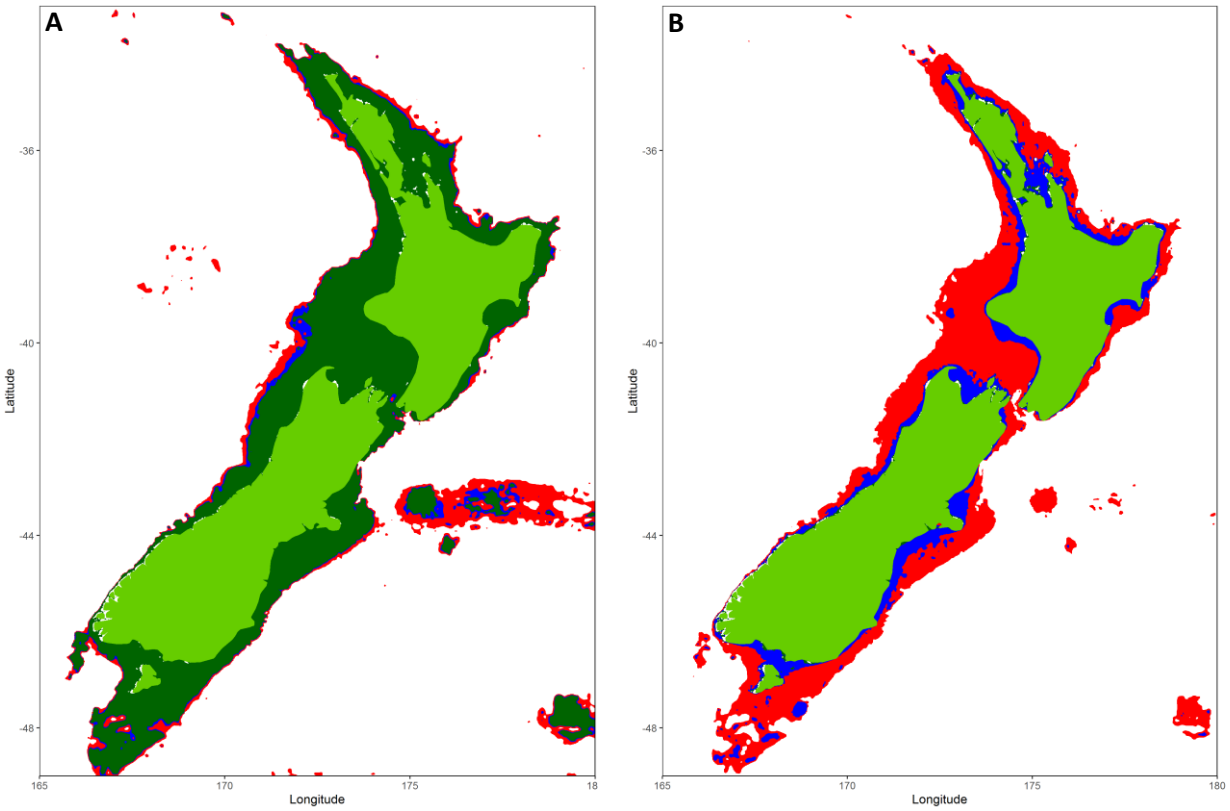

Figure S3.12: Rough extrapolation of available snapper habitat A) during last glacial maximum (LGM) with sea level 120m below current, and B) current available snapper habitat. Green shows the current terrestrial surface of New Zealand, dark green additional terrestrial surface that would have been above water during the LGM, blue is snapper habitat up to 50m (main distribution), red is snapper habitat up to 200m (max depth).

Chapter 4. Adaptive genetic variation in the Australian snapper (*Chrysophrys auratus*): natural selection or evidence of fishing-induced evolution?

4.0. Abstract

The identification of genes under selection has been central to better understanding what environmental conditions are driving evolutionary change. The ongoing advances in genomic methods now allow the identification of genes under selection throughout the genome of non-model species. Marine fish make perfect candidates for investigating selection. First, selection works effectively in species with large population sizes (e.g. fish). Second, fish populations are distributed across large heterogeneous habitats and exposed to a wide range of environmental conditions (including climate change and fishing). The Australasian snapper (*Chrysophrys auratus*) is one of New Zealand's largest inshore fisheries that has experienced heavy exploitation since the 1960s. Here we seek to identify adaptive genetic variation and to link it to genes to gain functional insights. Whole-genome sequences from snapper populations in New Zealand were used to scan for evidence of adaptation by 1) identifying selection occurring in a single population (local selection), and 2) occurring in all populations (national-wide selection). A selection scan identified 232 putative genomic regions containing 294 genes showing signs of selection. Two genomic regions were identified as strong outliers and both contained genes associated with angiogenesis, which has been linked to changes in maturation, and a pattern consistent with a recent or ongoing selective sweep. An adaptive response related to growth was also detected in a single genetic cluster, showing that there could be local selection on this trait. Finally, a selection scan contrasting two of the southernmost, and thus cold exposed populations, identified two genomic regions linked to cold stress. These results provide novel insights into what environmental factors are driving selection in snapper, and help identify the potential impacts of large-scale marine resource extraction.

4.1. Introduction

The identification of genes under selection has been central to a better understanding of how environmental conditions drive evolutionary change in species (Biswas & Akey, 2006). This has been difficult to study because traditional genetic approaches were generally limited to a small number of loci, making the identification of genes under selection difficult (Vitti et al., 2013). Moreover, many important phenotypic traits in nature are quantitative, and their typical polygenic basis requires dense whole-genome approaches to facilitate the detection (Mackay et al., 2009; Wellenreuther & Hansson, 2016).

The recent advancement of genomic methodologies enabled the genome-wide identification of genetic markers, allowing the study of all genes and their interactions. A large number of loci can now be sampled without prior knowledge of what traits they correspond to. The added benefit of having an annotated reference genome for non-model species (Lewin et al., 2018) is that it provides new opportunities to investigate interactions between loci, determine the relative positions to each other on the genome, and relate genes to functions that have been documented in other species. The use of population-level samples allows linkage disequilibrium to be estimated, which if high, is often indicative of a genomic region either currently or historically under selection (Vitti et al., 2013). Subsequent identification of genes located within regions under selection (via annotation or blasting against a database) (Stein, 2001), could enable inferences to be made about biological traits under selection, and associated environmental factors.

Regions of the genome under selection behave differently from the rest of the genome. In the absence of selection, all genetic loci are subject to genetic drift. This will eventually eliminate most alleles from the population, but a few will increase in frequency to reach or get close to fixation (Lande, 1976). When loci experience selection, specific alleles will propagate through the population at a faster rate when compared to those alleles subjected to drift alone. Because recombination is less efficient at breaking up physically linked genetic variants compared to their change in frequency under selection, there will often be a reduction in genetic variation surrounding the locus that selection is driving towards fixation (or close to fixation) in the population. A range of

methods have been developed to detect this pattern of a selective sweep, these include integrated haplotype scores (iHS) (Voight et al., 2006), identity-by-descent (IBD) analyses (Han & Abney, 2013), haplotype homozygosity (H12/G12) (Garud et al., 2015; Harris et al., 2018), or cross-population haplotype homozygosity (XP-EHH) (Sabeti et al., 2007). When new alleles begin to arise in the genomic region that experienced selection, the absence of any other alleles in the area will produce a pattern of overrepresentation of low frequency (rare) alleles. This surplus of rare alleles can be detected using frequency-based approaches, such as Tajima's D statistic. Another pattern that can be detected is when divergent selection occurs between two populations. In this case, adaptive alleles can be found by looking for unusual allele frequency differences (e.g. F_{ST}) between populations. Alleles under selection will resist the homogenising effects of gene flow between the two locations and stand out when the spectrum of allele frequencies is assessed.

Teleost fish are good candidates for studying natural selection in the wild. Selection is thought to act more efficiently in fish because they often maintain very large population sizes. In large populations genetic drift is weak, allowing alleles with small selection coefficients to propagate through the population ($s > 1/Ne$ to dominate evolution by random drift) (Li, 1997; Oleksiak, 2010). Rapid ecological divergence has been observed between pelagic and demersal spawning in European flounder when the Baltic Sea connected to the North Sea (Momigliano et al., 2017). In this case, the salinity gradient between the two seas appears to have favoured ecological niche differentiation, and the development of demersal spawning. The study found six genes near the flanking regions where evidence of selection was detected, including one gene (a *protocadherin fat1-like* gene) that has been associated with salinity stress in oysters (Zhang et al., 2016), and other teleosts (Boutet et al., 2006). Rapid niche expansion from pelagic to benthic niches in gangfisch (*Coregonus lavaretus macrophthalmus*) after ecosystem recovery has been associated with significant changes in body shape (Jacobs et al., 2019). The study found the number of gill rakes significantly increased and found nine genes under divergent selection associated with this phenotypic trait. This included a gene coding for *transforming growth factor beta* (*tgfβ*) *receptor*, which has also been associated with

ecological adaptation in salmonids (*Salvelinus alpinus*) and zebrafish (*Danio rerio*) (Ahi, 2016; Jacobs et al., 2018).

Marine species can be distributed over large spatial ranges and occupy a range of environmental gradients (Bernatchez, 2016; Cossins & Crawford, 2005). These conditions can be powerful drivers of natural selection (Schluter, 2001). A prominent environmental factor that is associated with adaptive variation is climate change (Hoffmann & Sgro, 2011). Increased ocean temperatures are causing changes in the distribution of species, ocean acidification, availability of suitable habitat, increased predation pressure, and resource availability (Koenigstein et al., 2016). In addition, more direct anthropogenic perturbations such as habitat destruction, pollution, and overfishing are additional environmental conditions (Halpern et al., 2015). Intense size-selective fishing has been reported to select for faster life-history traits, such as faster maturation and reduced growth rates (Enberg et al., 2012; Heino et al., 2015). Experiments have shown that size-selective removal of fish reduces fish size and detected associated changes in allele frequency (Therkildsen et al., 2019; van Wijk et al., 2013). However, possible confounding factors make it difficult to establish causal links between environmental conditions and genes that show a pattern of selection.

Australasian snapper (*Chrysophrys auratus*) is an ideal species for studying adaptive variation. In New Zealand, snapper occupies inshore waters along a latitudinal temperature gradient. Growth rates of captive snapper change in response to temperature (Wellenreuther et al., 2019a), and realised growth rates vary between stocks (Blackwell & Gilbert, 2008; Parsons et al., 2014; Walsh et al., 2011), suggesting that adaptive differences could be present. The boundary of their southern range is known to be related to water temperature (Parsons et al., 2014). Snapper also experience intense size-selective fishing (Fisheries New Zealand, 2018), and archaeological records show that the average size of snapper has decreased (Figure S4.1), making the species suitable to investigate for fishing-induced evolution.

This study focussed on the detection of genes under selection, link these to biological traits that were most likely the traits under selection, and discuss possible environmental drivers of evolutionary change. In this chapter, an analysis was conducted using 1) 350 genome sequences across New Zealand to conduct a genome-

wide scan for evidence of adaptive variation. 2) Genomic regions that were identified were searched for candidate genes related to growth (De-Santis & Jerry, 2007), and cold stress (Xu et al., 2018). 3) A bottom-up approach was used to detect genomic regions showing strong evidence of selection. Genes identified within these regions were linked to biological traits.

4.2. Methods

4.2.1. Data set

The data collection from the 350 individuals used in this study has been described in Chapter 2 (section 2.2.1.). To summarise, selection scans were performed using 6,034,882 high-quality SNPs, described in Chapter 2 (section 2.3.1). Samples were grouped into two populations based on the finding of the population differentiation from the analysis of neutral loci reported in Chapter 2 (i.e. North and South population), consisting of 195 and 155 samples respectively. An additional selection scan was performed to search for signs of divergent selection between Tasman Bay (n = 37) and Karamea Bight (n = 19).

4.2.2. Analyses

The approach in this chapter was to conduct a series of selection tests using a sliding-window approach, estimating multiple statistics over 5,000 base pair non-overlapping sections of the genome. Selection test statistics were estimated for individual sliding windows using the package popgenome in R (Pfeifer et al., 2014). Estimates for nucleotide diversity (π) and Tajima's D (T_D) were calculated for each population, and the statistics d_{XY} and F_{ST} for each population comparison. In addition, the difference in nucleotide diversity ($\Delta\pi$), and Tajima's D (ΔT_D) were calculated for each population comparison. G12 was estimated for each SNP to identify selective sweeps using unphased data (Garud et al., 2015; Harris et al., 2018). Subsequently, G12 for each sliding-window was determined by taking the maximum value observed among SNPs within that region (G12 max). Regions showing evidence for selection were identified using a bottom-up and top-down approach.

The bottom-up approach included the identification of outlier regions by taking the most extreme values from multiple statistics using quantile boundaries (Table 4.1). Different selection statistics were used to discriminate between two types of selection. Regions with outliers for both ΔT_D and F_{ST} across three consecutive sliding windows (i.e. 15,000 bp) showed significant evidence of selection in a specific region (local or divergent selection). Alternatively, regions with outliers for both T_D and G12 max across three consecutive sliding windows (i.e. 15,000 bp) showed significant evidence of selection in all regions (nation-wide selection). Regions that showed the strongest evidence of selection ($F_{ST} \geq 0.2$, or G12 max ≥ 0.2) were further investigated to identify possible genes and traits under selection.

Table 4.1: quantile boundaries for identifying selection

Type of selection	Quantile Boundary (%)			
	F_{ST}	ΔT_D	T_D	G12 max
Nation-wide selection			≤ 0.1	≥ 99.9
Local selection	≥ 99	≤0.5 , ≥ 99.5		

Notes: T_D : Tajima's *D*, ΔT_D: difference in Tajima's *D* between regions

In the top-down approach, candidate genes were searched for possible signs of selection. Thirteen candidate genes for teleost fish have previously been identified by De-Santis and Jerry (2007). In addition, 5 *growth arrest-specific proteins* (*gas*) were added as potential candidates (i.e. *gas1*, *gas2*, *gas6*, *gas7*, and *gas8*). One of these proteins (i.e. *gas2*) has previously been linked to low temperature stress-induced apoptosis in tilapia (Yang et al., 2017). Candidate genes were subsequently identified using the previously published annotation of the snapper genome (Catanach et al., 2019). Similar, a list of candidate genes related to cold stress were obtained from Xu et al. (2018). Genes from this list were assessed for evidence of selection because they could be associated with adaptation to colder waters. Sampled areas known to have different average sea temperatures such as the Tasman Bay and Karamea Bight were compared when using this test (Figure S4.2). Sea surface temperature data was obtained from Bio-ORACLE.org and imported into R using the sdmpredictors package for plotting (Bosch et al., 2017).

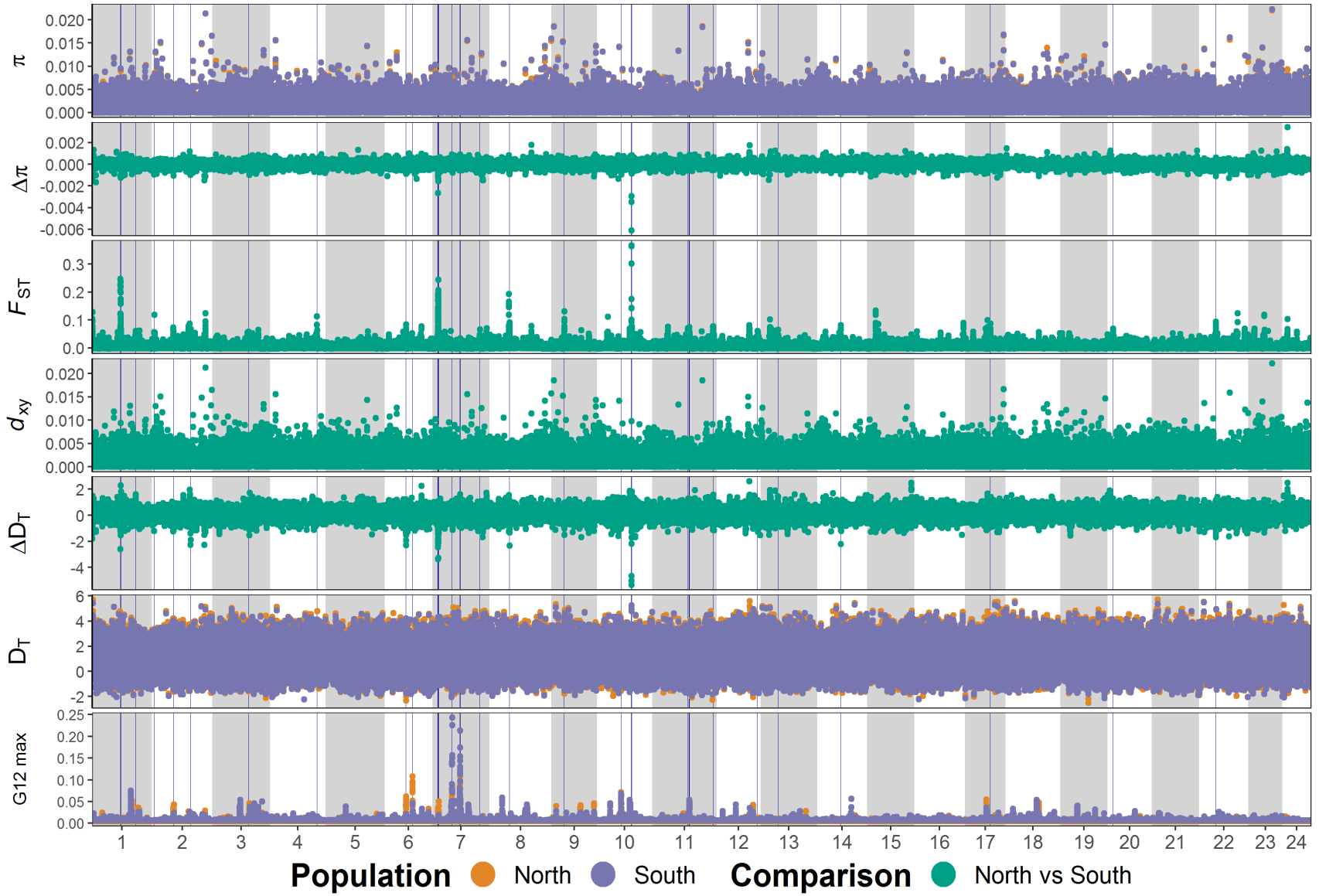

Figure 4.1: Selection scan including all three genetic clusters. Grey and white rectangles show different linkage groups. Vertical blue lines indicate genomic regions that showed significant evidence for selection for one or more selection statistics.

4.3. Results

4.3.1. Detection of local and nation-wide selection using genetic clusters including all sampling sites

Selection scans performed on the two genetic clusters identified 101 genomic regions that were significant (Figure 4.1). Within these regions, 253 genes that could be candidates for selection were found. Of these regions, 79 regions showed a pattern of divergent selection between genetic clusters (local selection). These regions either showed significant outliers for F_{ST} (99.5% quantile ≥ 0.04) and ΔD_T (0.05% quantile ≤ -0.77 & 99.5% quantile ≥ 0.97). In total, 1,434 sliding window frames showed significant divergence for one or more statistics, spanning 7.17 Mbp (0.98% of the genome). In addition, 22 regions showed evidence of selection in both genetic clusters (nation-wide selection). These regions either had outlier for either G12-max (99.9% quantile ≥ 0.054) or D_T (0.01% quantile ≤ -2.178).

The strongest signals of local selection (F_{ST} ≥ 0.2) were detected on linkage groups 1, 7, and 10 (Figure 4.2 & Table S4.1). On linkage group 1, a genomic region spanning 0.28 Mbp (16.9 – 17.3 Mbp) was identified containing three genes (Figure 4.2A). The gene located in the region of highest divergence was *gypsy retrotransposon integrase-like protein 1* (*gin1*). On linkage group 7, a genomic region spanning 0.54 Mbp (3.3 - 3.8 Mbp) was identified containing four known genes (Figure 4.2B). Two genes were located in the region of highest divergence: *myomegalin* (*pde4dip*), and *microtubule-associated serine/threonine-protein kinase 2* (*mast2*). The coding region for epidermal growth factor were also observed to show signs of divergent selection on linkage group 6 (data not shown). On linkage group 10, a genomic region spanning 0.35 Mbp (between 20.6 – 20.9 Mbp) was identified, containing eight genes (Figure 4.2C), of which two were of potential interest due to their molecular function and proximity to the adaptive peak. Two genes were located in the region of highest divergence: *hexokinase-2* (*hk2*), and *elin-2* (*erlin2*).

The strongest signals of nation-wide selection (G12 ≥ 0.2) were detected in two genomic regions on linkage group 7 (Figure 4.3 & Table S4.2). The first region on linkage

group 7 spanned 0.07 Mbp (11.79 – 11.87 Mbp) and contains two genes: *phospholipid-transporting atpase Ik* (*atp8a1*), and *e3 ubiquitin-protein ligase rnf213-alpha* (*rnf213a*) (Figure 4.3A). In zebrafish (*Danio rerio*), mutations in *rnf213a* have been shown to cause abnormal angiogenesis (Liu et al., 2011; Wen et al., 2016), suggesting this gene is involved in the vascular development in fish. The second region on linkage group 7 spanned 0.4Mbp (16.3 – 16.7Mbp), and codes for 14 different proteins (Figure 4.3B). Two notable genes in regions that showed the strongest signal for selective sweep are *myeloid-derived growth factor* (*mydgf*), and *megakaryocyte-associated tyrosine-protein kinase* (*matk*).

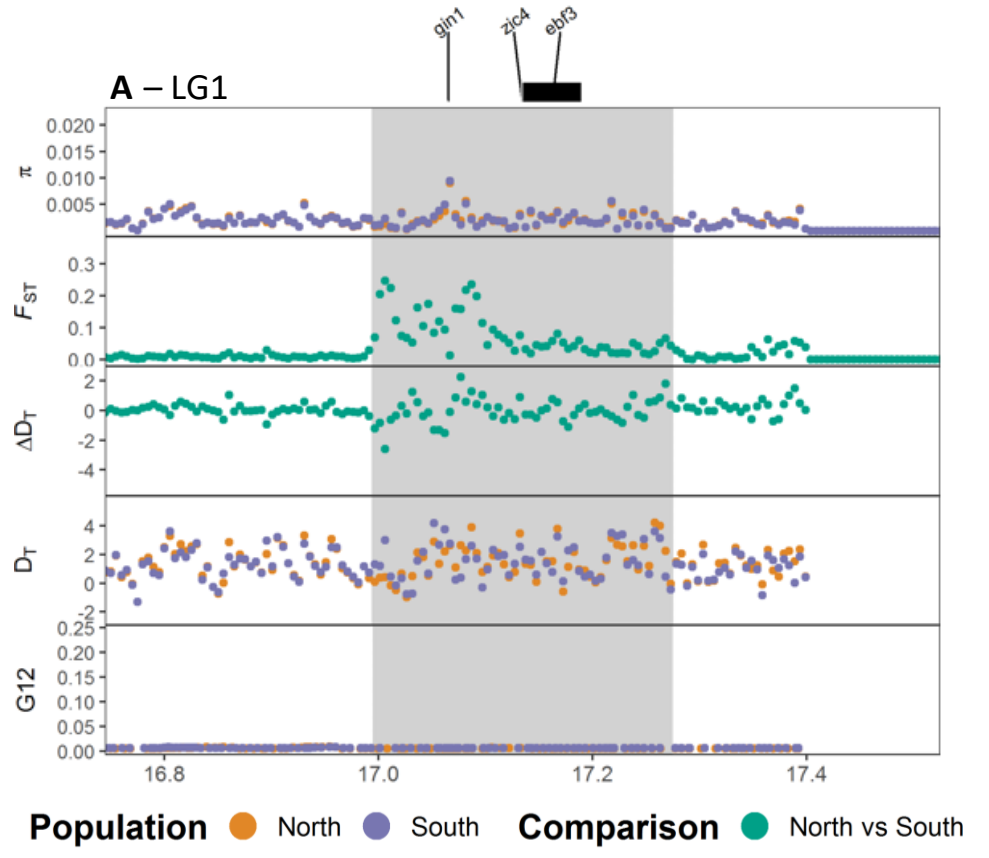

Figure 4.2: Three genomic regions showing strong signs of local selection (A-C). Grey areas indicate the genomic region containing outlier loci. X-axis shows the genomic location on the respective linkage group (LG) in Mbp. Genes located into the outlier are depicted at the top of the plots, full gene names can be found in Table S4.1.

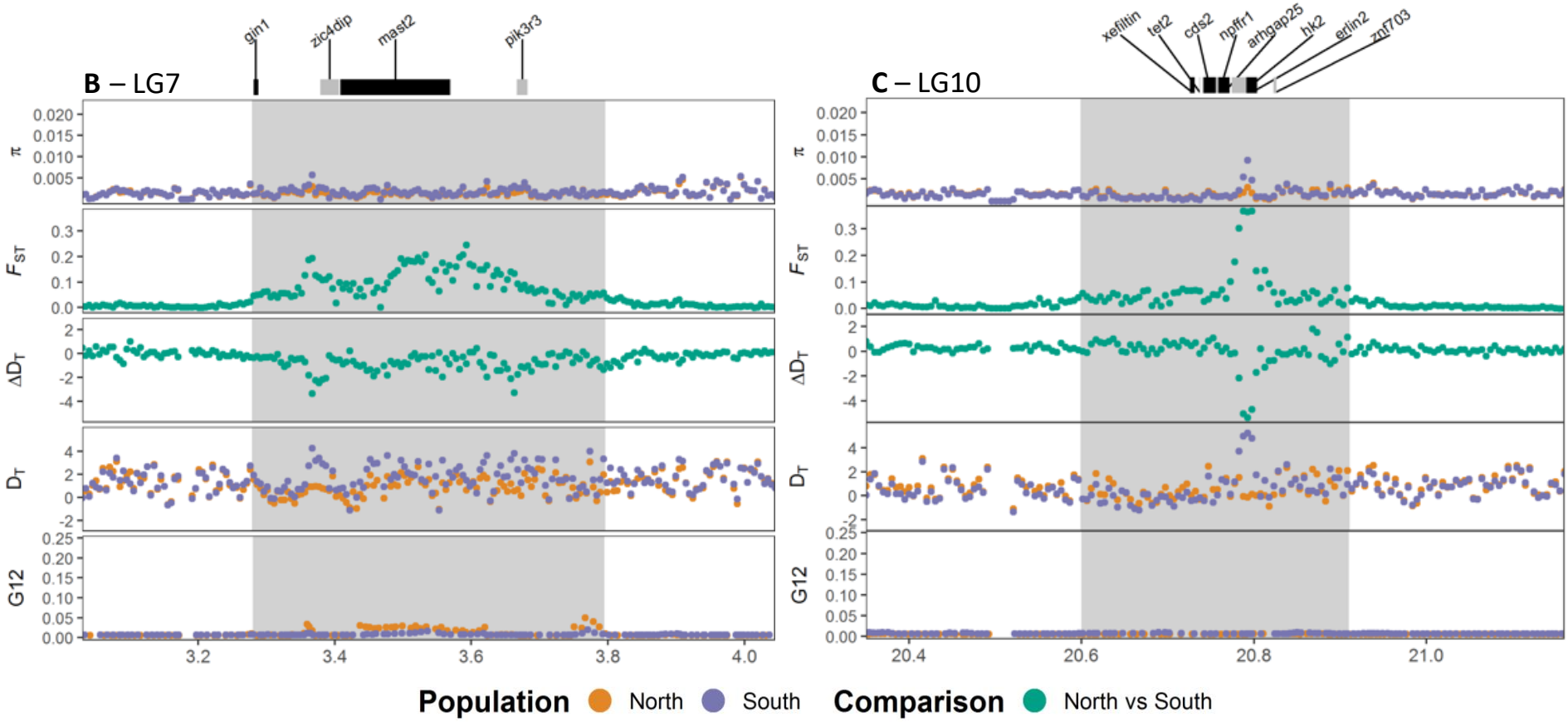

Figure 4.2 (continued): Three genomic regions showing strong signs of local selection (A-C). Grey areas indicate the genomic containing outlier loci. X-axis shows the genomic location on the respective linkage group (LG) in Mbp. Genes located in to the outlier are depicted at the top of the plots, full gene names can be found in Table S4.1.

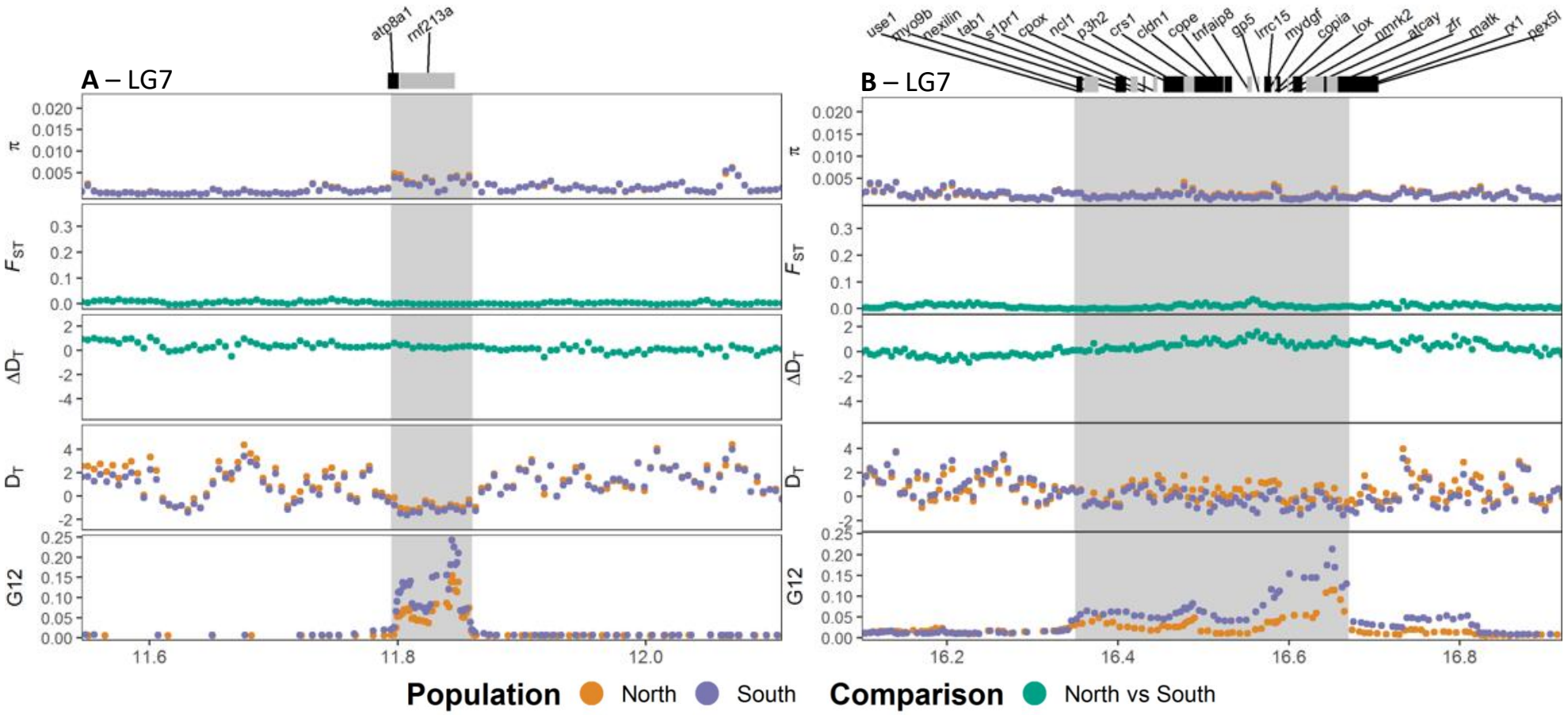

Figure 4.3: Two genomic region showing strong signs of nation-wide selection (A&B). Grey areas indicate the genomic region containing outlier loci. X-axis shows the genomic location on the respective linkage group (LG) in Mbp. Genes located in the outlier regions are depicted at the top of the plots, full gene names can be found in Table S4.2.

4.3.2. Divergent selection between Tasman Bay and Karamea Bight

The selection scan conducted between samples from the Tasman Bay and Karamea Bight populations identified 123 regions that were significant. Overall, 6.5 Mbp of the genome showed evidence of selection based on one or more of the statistical tests (0.9% of the genome). Out of the 296 genes identified, 197 genes showed a pattern consistent with divergent selection between the two locations. These regions contained significant outliers for F_{ST} (99.5% quantile ≥ 0.07), ΔD_T (0.05% quantile ≤ -1.636, and 99.5% quantile ≥ 1.298), or both. One showed a high level of population differentiation (F_{ST} ≥ 0.2) between Karamea Bight and Tasman Bay (Figure S4.4). This region contained the coding region for the gene *mast2*. This was also identified as a divergent region between genetic clusters (Figure 4.2B).

Two genes that were differentially expressed in a cold-stress experiment were found in regions showing evidence for divergent selection (Xu et al., 2018). On linkage group 7 the gene coding for *neurogenic locus notch homolog protein 2* (*notch2*) (Figure 4.4A) was detected, and on linkage group 22 *calcium/calmodulin-dependent protein kinase type II subunit gamma* (*camk2g*) (Figure 4.4B), a subunit of the camk2 candidate gene, was identified.

Furthermore, two genomic regions showed evidence of selective sweeps in the Karamea Bight population (G12 ≥ 0.2). One region was also detected in the selection scan that included all sampled sites (Figure 4.5A), which contained the *rnf213a* gene. An additional region on linkage group 10 (14.4 - 14.5Mbp) was also of interest because a selective sweep (G12 max) was detected in the Karamea Bight populations (Figure 4.5B). This region had already been identified as a region of interest for the South genetic cluster that includes both of these populations. The region contains the coding region for *kinase suppressor of Ras2* (*ksr2*).

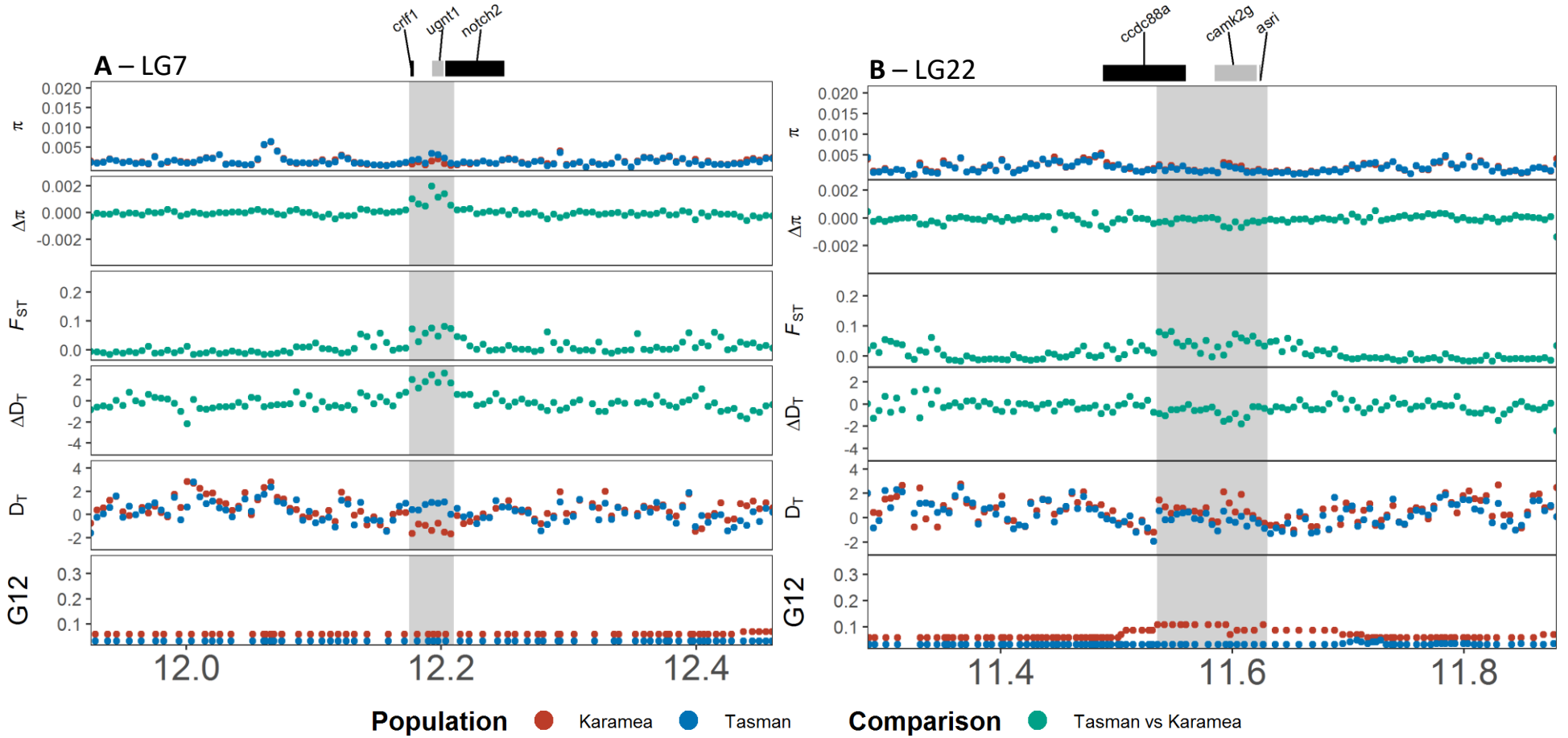

Figure 4.4: Two genomics regions containing candidate genes related to cold stress (A&B). Grey areas indicate the genomic regions containing outlier loci. X-axis shows the genomic location on the respective linkage group (LG) in Mbp. Genes located into the outlier are depicted at the top of the plots, full gene names can be found in Table S4.3.

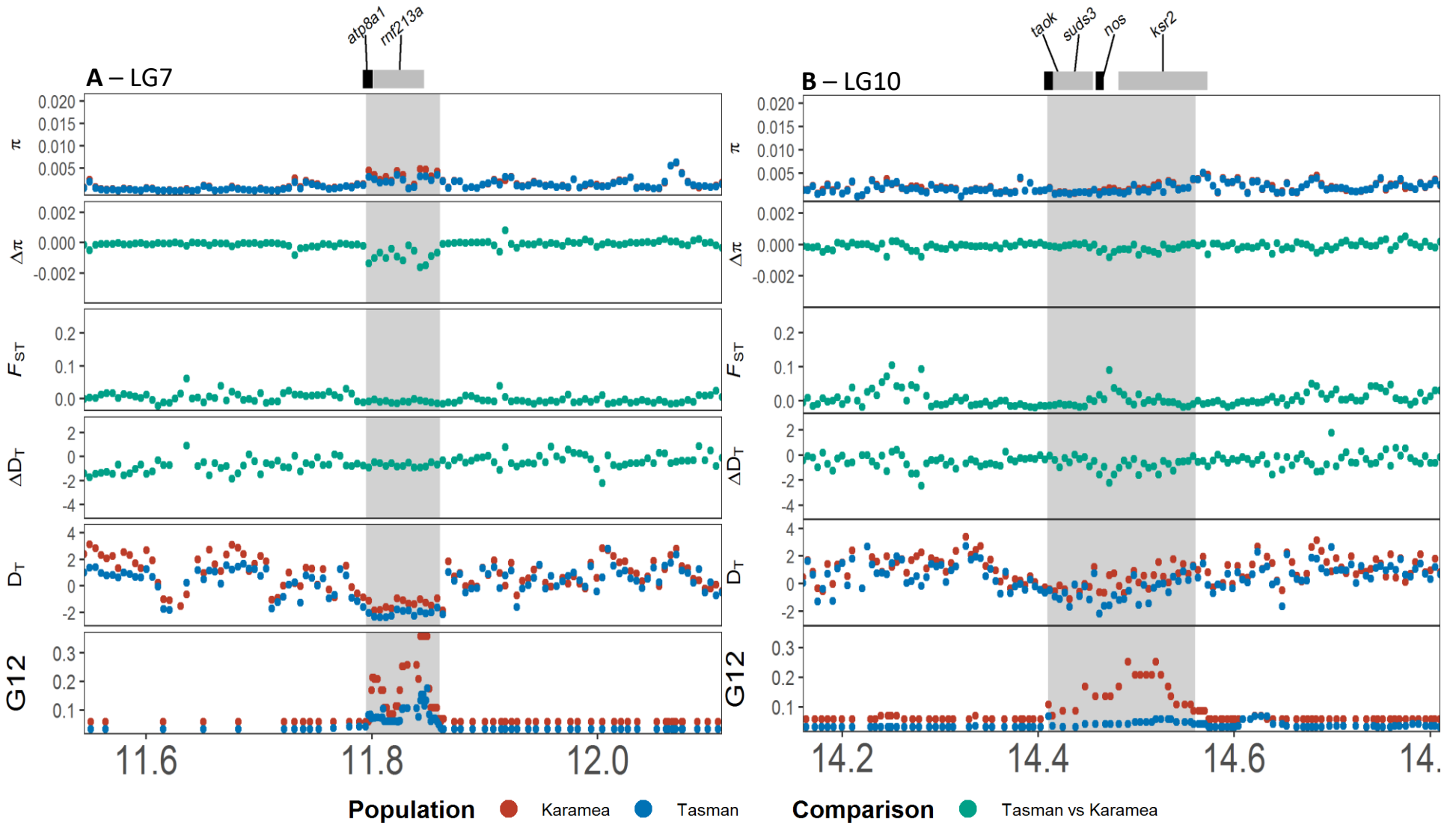

Figure 4.5: Two genomic regions showing signs of a selective sweep occurring in the Karamea Bight population (A&B). Grey areas indicate the genomic region containing outlier loci. X-axis shows the genomic location on the respective linkage group (LG) in Mbp. Genes located into the outlier are depicted at the top of the plots, full gene names can be found in Table S4.4.

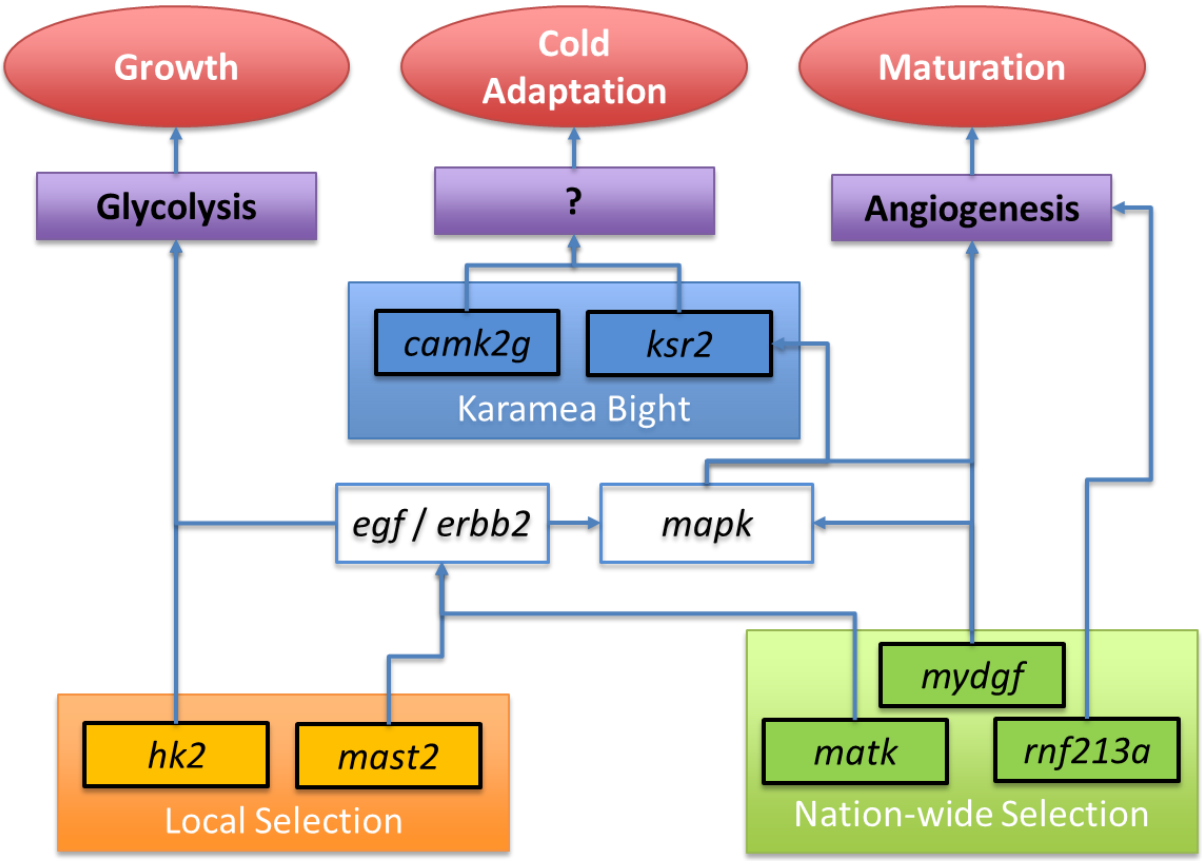

Figure 4.6: schematic representation of identified genes experiencing selection and their biological functions. Areas in blue, orange, and green indicate genes showing signs of divergent selection or selective sweeps. White boxes show signalling pathways related to identified genes and biological processes of interest.

4.4. Discussion

Marine teleost fish populations are often large and distributed across heterogeneous environments, providing suitable conditions for selection. This allows to evaluate the impacts of different processes on the genome, such as the role of natural environmental selection and the impact of anthropogenic pressures induced by fishing. Genome-wide sequencing of populations sampled across their distribution enabled the identification of genomic regions that contained adaptive variation. An analysis of 350 snapper genomes showed that 101 genomic regions deviated from neutral expectations. Subsequent analyses of these genomic areas identified genes and provided an opportunity to infer possible traits that may be experiencing selection. Some of the identified genes have previously been associated with growth, maturation rate and adaptation to cold in other species (Figure 4.6).

4.4.1. Adaptive variation at the local level

Localised patterns of adaptive genetic variation between samples found in the North and South genetic clusters suggest that these populations experience different

environmental conditions (Figure 4.2). While it is difficult to infer what environmental factors might be important in driving this pattern, it is possible that it is affecting growth rates. Two genes located in regions that showed a significant level of selection are linked to cell growth, i.e. *mast2*, *hk2*. *Mast2* is linked to the expression of the epidermal growth factor (Ahn et al., 1991), a gene that also showed moderate levels of adaptive variation (data not shown). In humans, *hk2* is a rate-limiting enzyme at the first step of glycolysis to catalyse cellular glucose into glucose-6-phosphate (Li et al., 2020), and has also been observed to play an important role in cell growth (Wellen et al., 2010). The detecting selection using estimates of relative population differentiation (F_{ST}) does not provide information on which population(s) experienced selection.

Analysis of LG10 suggested that the southern cluster has experienced balancing selection (Figure 4.2C). Here, a positive peak in Tajima's D statistic is indicating a lack of rare alleles, and high nucleotide diversity are indicative of long-term persistence of multiple alleles in the population (Charlesworth et al., 1997; Charlesworth, 2006; Hohenlohe et al., 2010). Although the same signals were undetected in the other two regions that showed strong signs of local selection, it is worth considering what could facilitate balancing selection in the southern cluster. There are several types of balancing selection, these include heterozygote advantage (over-dominance), frequency-dependant selection, and variation in fitness across spatial or temporal scales (Charlesworth, 2006; Llaurens et al., 2017). It is possible that variation in sea temperature across spatial scales is maintaining multiple alleles in the southern population. This common for panmictic marine fishes with high dispersal across large heterogeneous areas (Bernatchez, 2016). The southern cluster covers a large latitudinal range with a significant amount of spatial variation in sea surface temperatures (Figure S4.5). In addition, sea currents along the East Coast of the South Island pass through the Cook Strait and feed colder water into the Tasman Bay and Kapiti Coast areas.

Although selection on genes affecting growth is a predicted outcome of fishing-induced evolution (Enberg et al., 2012; Heino et al., 2015), it is unlikely that fishing is causing balancing selection in the southern cluster. This is because both the northern and the southern cluster are exposed to intense-fishing effort and a similar pattern should be detectable in the northern cluster. Differences in age structure and growth

rates between southern and northern clusters do suggest the regions experienced different exploitation histories (Blackwell & Gilbert, 2006, 2008; Walsh & Davies, 2004; Walsh et al., 2011). This could explain differences in selection between the northern and southern cluster. However, intense-fishing effort is a relatively recent selection pressure (last 60 years). Selection over such time scales means that intense-fishing effort may only be detectable if it was a rapid selective sweep in the population.

4.4.2. Nation-wide signs of selection

The detection of selective sweeps (i.e. reduction in genetic diversity in a genomic region under selection) in both genetic clusters suggests recent or ongoing selection in all of the sites sampled in this study (Figure 4.3). Local reduction of Tajima's D shows an excess of rare alleles within the regions of interest, consistent with selective sweeps (Vitti et al., 2013). This evidence of selection and its interpretation are consistent with the findings of a previous study that looked at adaptive radiation of three spine sticklebacks (*Gasterosteus aculeatus*) (Marques et al., 2018). Most studies that report the detection of selection in fish focus on genetic differences between populations (i.e. F_{ST}) (Clucas et al., 2019; Moore et al., 2014; Osterberg et al., 2018; Reid et al., 2016), similar to the detection of local selection in this study. This does not allow for the detection of natural selection occurring in populations that experience the same environmental conditions. This is observed in this chapter as differentiation between the two populations was limited in the genomic regions showing evidence of a selective sweep (Figure 4.3). The detection of selective sweeps through approaches that involved the detection of local reductions in genetic diversity (e.g. H12/G12, iHS, XP-EHH) would also allow for the detection of selection over shorter time-scales as differences in alleles frequencies between populations take time to establish (Sabeti et al., 2006).

The association of the *matk* signalling pathway with *erbb2* implies a wide range of possible downstream effects (Zrihan-Licht et al., 1997). For instance, overexpression of *matk* was shown to facilitate cell growth inhibition (Dokmanovic et al., 2015). Moreover, mutations in *erbb2* have been shown to activate the *mapk* pathway (Anglesio et al., 2008; Demidov et al., 2007), which has been associated with cell proliferation and a wide range of biological systems such as cardiovascular, nervous system, and immunology (Sun et al., 2015; Zhang & Liu, 2002). *Mapk* has also been linked to immune

responses in fish (Wei et al., 2020). The genomic region showing signs of a selective sweep that includes *matk* involves many genes. It is possible that a different gene is experiencing a selective sweep and *matk* is hitchhiking with it.

Both regions contained genes involved with angiogenesis (i.e. *rnf213a* and *mydgf*) (Korf-Klingebiel et al., 2015; Liu et al., 2011). Disruption of angiogenesis is associated with diseases such as cancer in humans (Shahi & Pineda, 2008), as well as playing a vital role during early development (Breier, 2000). In threespine sticklebacks (*Gasterosteus aculeatus*), changes in angiogenesis were shown to influence maturation rates and skew sex ratios (Furin et al., 2015). Changes in maturation have been observed in many teleost species, and are a known response to intense size-selection fishing effort, which can be a form of directional selection (Dieckmann & Heino, 2007; Enberg et al., 2012; Heino et al., 2015; Trippel, 1995). Intense fishing effort that removes the largest individuals from a population would facilitate an adaptive advantage for variants that maximised their reproduction output earlier in life (Enberg et al., 2012). A causal connection between the genes experiencing a directional selection and maturation is difficult to establish. However, it is noteworthy that the two regions that show evidence for recent or ongoing selection are potentially linked to adaptive variation reported for many other fished species (see Heino et al. (2015)). The detection of selective sweeps near genes associated with angiogenesis does not provide any specific evidence for a response to fishing, but it would be an interesting connection to explore in more detail.

The study could not exclude the presence of any other environmental conditions that could be causing the observed evolutionary changes. Various environmental conditions including habitat disturbance, trophic impacts, invasive species, and climate change have been identified as possible environmental conditions for snapper in New Zealand (Parsons et al., 2014). Contrasting genes under selection between snapper in New Zealand and Australia will provide useful information to further investigate what environmental conditions could be causing the observed selection.

4.4.3. Adaption to cold stress in Karamea Bight

Divergent selection between Tasman Bay and Karamea Bight is a possibility at the southern end of the snapper distribution. Karamea Bight could represent an

adaptive front that has a higher frequency of adaptive variants associated with reproduction and feeding in colder waters. A study looked at the effect of different temperature treatments on the differential expression of genes in yellow drum (*Nibea albiflora*) (Xu et al., 2018), and identified genes and gene-pathways that are likely involved in the cold-stress response. The number of differentially expressed genes ranged between 46 and 208 depending on the treatment comparison and tissue type. Two genes that also showed a pattern of differential expression between temperature treatments were also found to show signs of divergent selection and a selective sweep: *notch2* and *camk2* (Xu et al., 2018). Specifically, it was the gamma subunit from *camk2* (*camk2g*) that was identified as a gene potentially involved in the advantageous trait. While the relation between *notch2* and cold stress or other relevant biological function is not well established, *camk2* appears to have a strong association with cold stress tolerance (Cao et al., 2004).

The region that includes the coding region for *camk2g* also showed a moderate signal of a selective sweep, suggesting recent or ongoing selection. In addition, the selective sweep observed in the genomic region coding for *ksr2* adds to the hypothesis of cold stress adaptation in the Karamea Bight population (Figure 4.5B). *Ksr2* has been shown to regulate energy balance via feeding behaviour and adaptive thermogenesis (Guo et al., 2017). Finally, *ksr2* is likely involved in the downstream signalling of the *mapk* pathway (Liu et al., 2009).

It is unclear why the selective sweep in the region coding for *rnf123a* is stronger in the Karamea Bight population. Signals of selective sweeps are possibly inflated in the Karamea Bight population due to the smaller sample size (19 vs 37). However, the effect of a smaller sample size is likely observed in the difference in baseline values of G12 (Figure 4.5A), showing only a slight inflation. Similar, the divergent selection detected in the region coding for *mast2* does not show a direct connection to the cold stress hypothesis. Other environmental differences between the two locations likely cause the observed differences. For this study, specific interest went towards the cold stress hypothesis.

4.4.4 Future research and conclusion

Future efforts for genomic analyses should focus on obtaining a phased SNP data set for the detection of selective sweeps. The detection of large haplotype blocks (i.e. selective sweeps) provides the best resolution for detecting selection over such short timescales (Sabeti et al., 2006). Here, G12 provided the best approach for the detection of increased size of haplotype blocks using unphased data. G12 is an adaptation to the H12 statistic which uses phased data to detect selective sweeps (Garud et al., 2015; Harris et al., 2018). The power of this analysis would significantly increase if phased data sets were available. Statistics such as his and XPWEHH were also estimated with the current data using the R-package rehh (Gautier et al., 2017). Following Gautier's recommendations for using unphased data resulted in very stringent filtering of the data. This resulted in a lot of missing data, especially in regions of interest that were identified using other statistical tests. In addition, structural variants are likely to provide additional insights regarding selection. Structural variants such as inversion and copy number variants have been shown to play a significant role in adaptation (Wellenreuther et al., 2019b). Finally, the addition of genotype-environment association (GEA) could provide information on which environmental gradient could be facilitating evolutionary change (Grummer et al., 2019).

This study focussed on the regions that showed the strongest signs of selection. It has been well established that phenotypic changes are often the result of multiple genes of small effect (i.e. quantitative traits). Gene ontology (GO) analyses can provide better insights into the biological functions related to the regions that show signs of selection. Overlap in GO between genes under selection will help build the argument for specific traits being under selection. The addition of relevant phenotypic data such as age, size at age/maturation or sex will allow for quantitative trait loci (QTL) to be identified, which have already been related to changes in growth in captive snapper (Ashton et al., 2019b). Similar to GO analyses, QTL mapping helps to relate genomic regions or genes of small effect to a specific trait. Finally, RNA sequencing may provide insights into the differential expression of the genes identified to be under selection, and determine whether genes under selection are being up or down regulated in response to different environmental conditions (Xu et al., 2019).

Genome scans in snapper showed significant signs of selection across the genome, both within groups of localised samples and when considering the nation-wide. While the underlying environmental conditions are unclear, genes located in the regions under selection provide tentative support for adaptive responses to cold stress and possibly intense fishing pressure, and balancing selection in a population that covers a temperature gradient. The suggestions for future research will help provide more power to detect recent or ongoing selection and relate genes under selection to traits. Knowing what traits are under selection in a given species or population provides powerful insights into what environmental conditions are most relevant to the reproductive success of a species. Climate change and other anthropogenic perturbations are predicted to force species to adapt at unprecedented rates, and understanding the adaptive responses to these selective pressures may help provide better conservation management plans to protect new and emerging adaptive fronts.

4.5. Supplementary information

4.5.1. Supplementary Tables

Table S4.1: Genes located in genomic regions showing strong local selection

Figure panel	ID	Size	Full name	Gene code
Figure 4.2A	LG1	0.28	***Gypsy retrotransposon integrase-like protein 1***	***gin1***
			Zinc finger protein ZIC 4	*zic4*
			Transcription factor COE3	*ebf3*
Figure 4.2B	LG7	0.54	*Paired mesoderm homeobox protein 1*	*prrx1*
			Myomegalin	***pde4dip***
			Microtubule-associated serine/threonine-protein kinase 2	***mast2***
			Phosphatidylinositol 3-kinase regulatory subunit gamma	*pik3r3*
Figure 4.2C	LG10	0.35	*Low molecular weight neuronal intermediate filament*	*xefiltin*
			Methylcytosine dioxygenase TET2	*tet2*
			Phosphatidate cytidylyltransferase 2	*cds2*
			Neuropeptide FF receptor 1	*npffr1*
			Rho GTPase-activating protein 25	*arhgap25*
			Hexokinase-2	***hk2***
			Erlin-2	***erlin2***
			Zinc finger protein 703	*znf703*

Notes: Table refers to the selection scan plots in Figure 4.2. Size indicates the size of the genomic region under selection in Mbp.

Table S4.2: Genes located in genomic regions showing strong national selection

Figure panel	ID	Size	Gene name	Gene code
Figure 4.3A	LG7	0.07	Phospholipid-transporting ATPase IK	ATP8A1
			E3 ubiquitin-protein ligase rnf213-alpha	**rnf213a**
Figure 4.3B	LG7	0.4	Vesicle transport protein USE1	use1
			Unconventional myosin-IXb	myo9b
			Nexilin	nexilin
			TGF-beta-activated kinase 1 and MAP3K7-binding protein 3	tab1
			Sphingosine 1-phosphate receptor 1	s1pr1
			Oxygen-dependent coproporphyrinogen-III oxidase, mitochondrial	cpox
			Prolyl 3-hydroxylase 2	p3h2
			Nicalin-1	ncl1
			Ceramide synthase 1	crs1
			Claudin-1	cldn1
			Coatomer subunit epsilon	cope
			Tumor necrosis factor alpha-induced protein 8-like protein 1	tnfaip8
			Platelet glycoprotein V	gp5
			Leucine-rich repeat-containing protein 15	lrrc15
			Myeloid-derived growth factor	**mydgf**
			Copia protein	copia
			Protein-lysine 6-oxidase	lox
			Nicotinamide riboside kinase 2	nmrk2
			Caytaxin	atcay
			Zinc finger RNA-binding protein	zfr
			PEX5-related protein	pex5l
			Megakaryocyte-associated tyrosine-protein kinase	**matk**
			Retinal homeobox protein Rx1	rx1

Notes: Table refers to the selection scan plots in Figure 4.3. Size indicates the size of the genomic region under selection in Mbp.

Table S4.3: Candidate genes related to cold stress showing signs of divergent selection between Tasman Bay and Karamea Bight

Figure panel	LG	Size	Gene name	Gene code
Figure 4.4A	LG7	0.035	Cytokine Receptor Like Factor 1	crlf1
			UDP-N-acetylglucosamine transporter	ignt1
			Neurogenic locus notch homolog protein 2	**notch2**
Figure 4.4B	LG22	0.095	Girdin	ccdc88a
			Calcium/calmodulin-dependent protein kinase type II subunit gamma	**camk2g**
			Arylsulfatase I	Asri

Notes: Table refers to the selection scan plots in Figure 4.4. Size indicates the size of the genomic region under selection in Mbp.

Table S4.4: Genes located in genomic regions showing divergent signals in selective sweeps between Tasman Bay and Karamea Bight

Figure panel	LG	Size	Gene name	Gene code
Figure 4.5A	LG7	0.065	Phospholipid-transporting ATPase IA	atp8a1
			E3 ubiquitin-protein ligase RNF123	**rnf123a**
Figure 4.5B	LG10	0.15	Serine/threonine-protein kinase TAO3	taok3
			Nitric oxide synthase	nos
			Sin3 histone deacetylase corepressor complex component SDS3	suds3
			Kinase Suppressor Of Ras 2	**ksr2**

Notes: Table refers to the selection scan plots in Figure 4.5. Size indicates the size of the genomic region under selection in Mbp.

4.5.2. Supplementary Figures

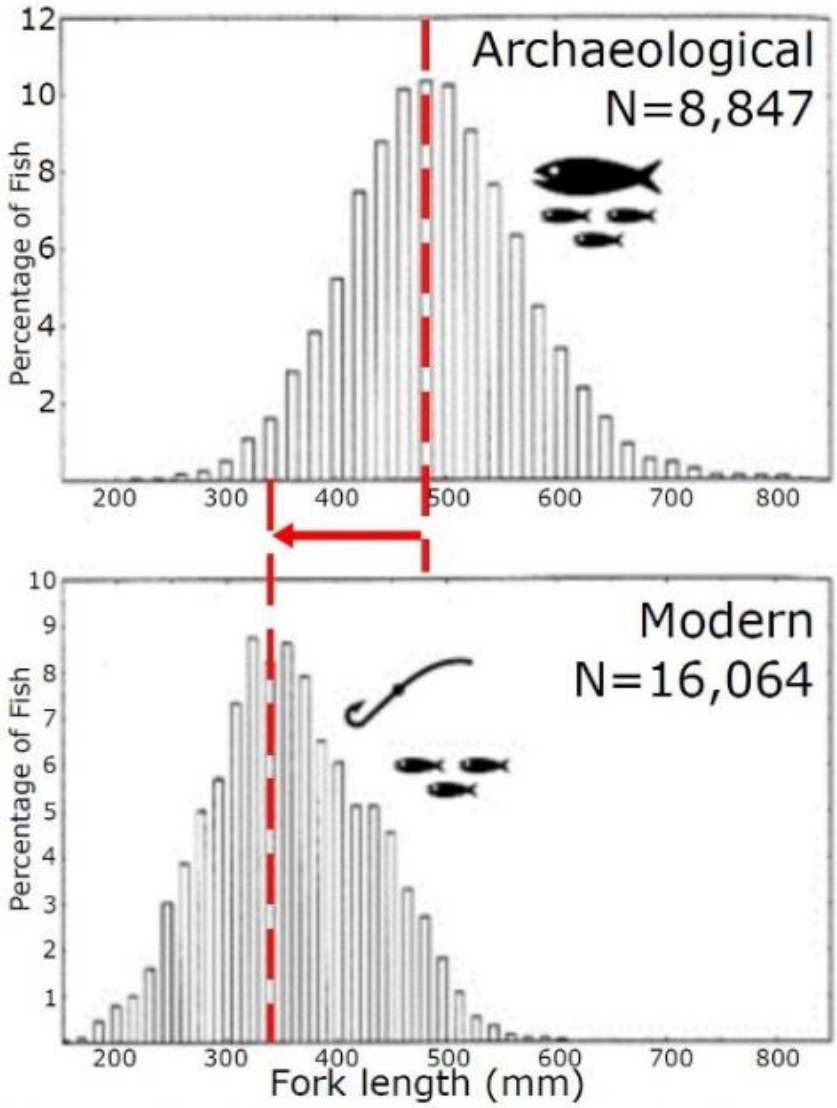

Figure S4.1: reduction on fork length of snapper over time. Source: Leach et al., 2006.

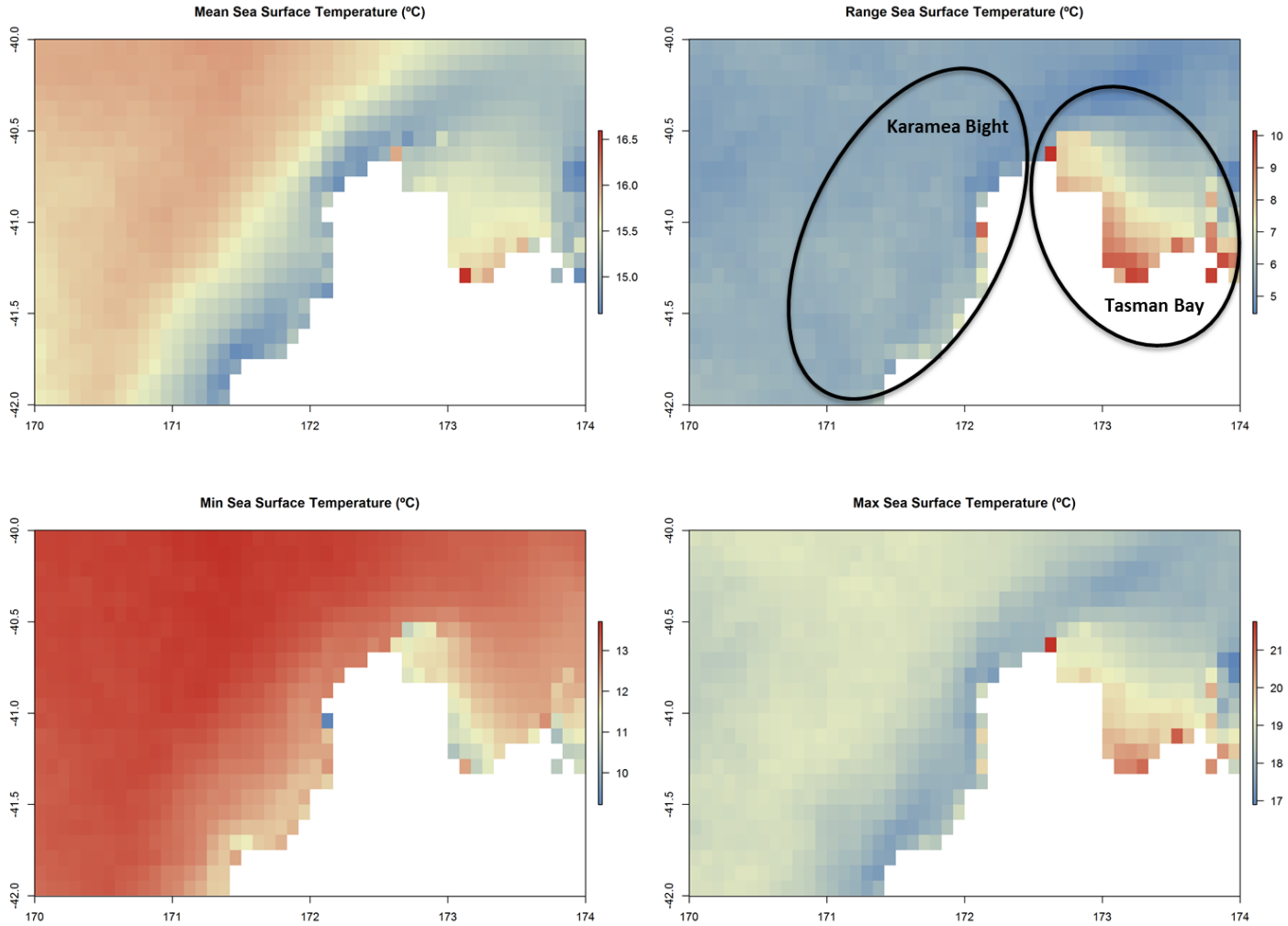

Figure S4.2: Maps showing the variation in temperature between Tasman Bay and Karamea Bight (indicated in top right panel). Temperature data has been obtained from bio-oracle.org.

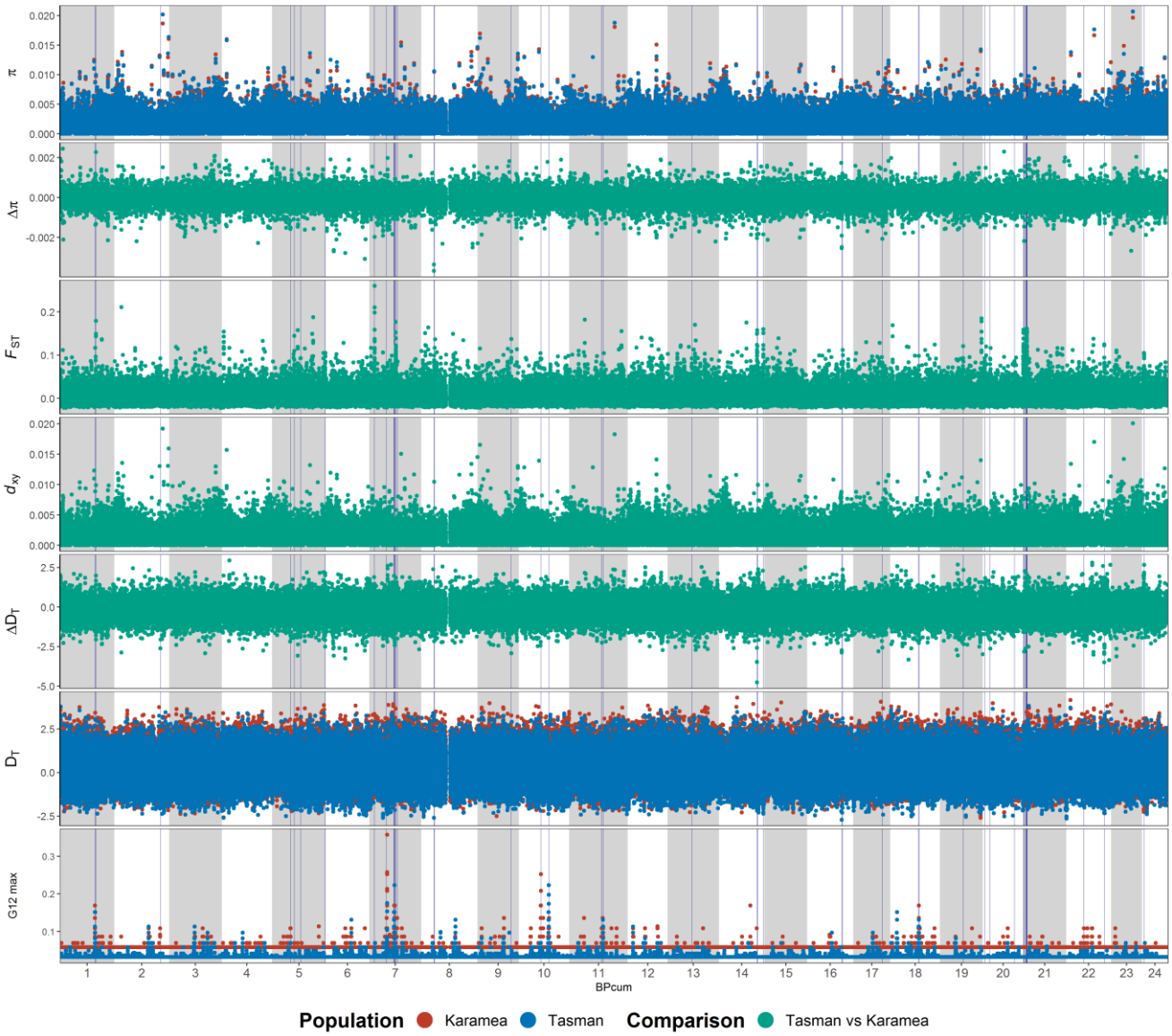

Figure S4.3: Selection involving Tasman Bay and Karamea Bight. Grey and white rectangles show different linkage groups. Vertical blue lines shows outlier regions possibly experiencing selection. Multiple regions were too small to be plotted properly.

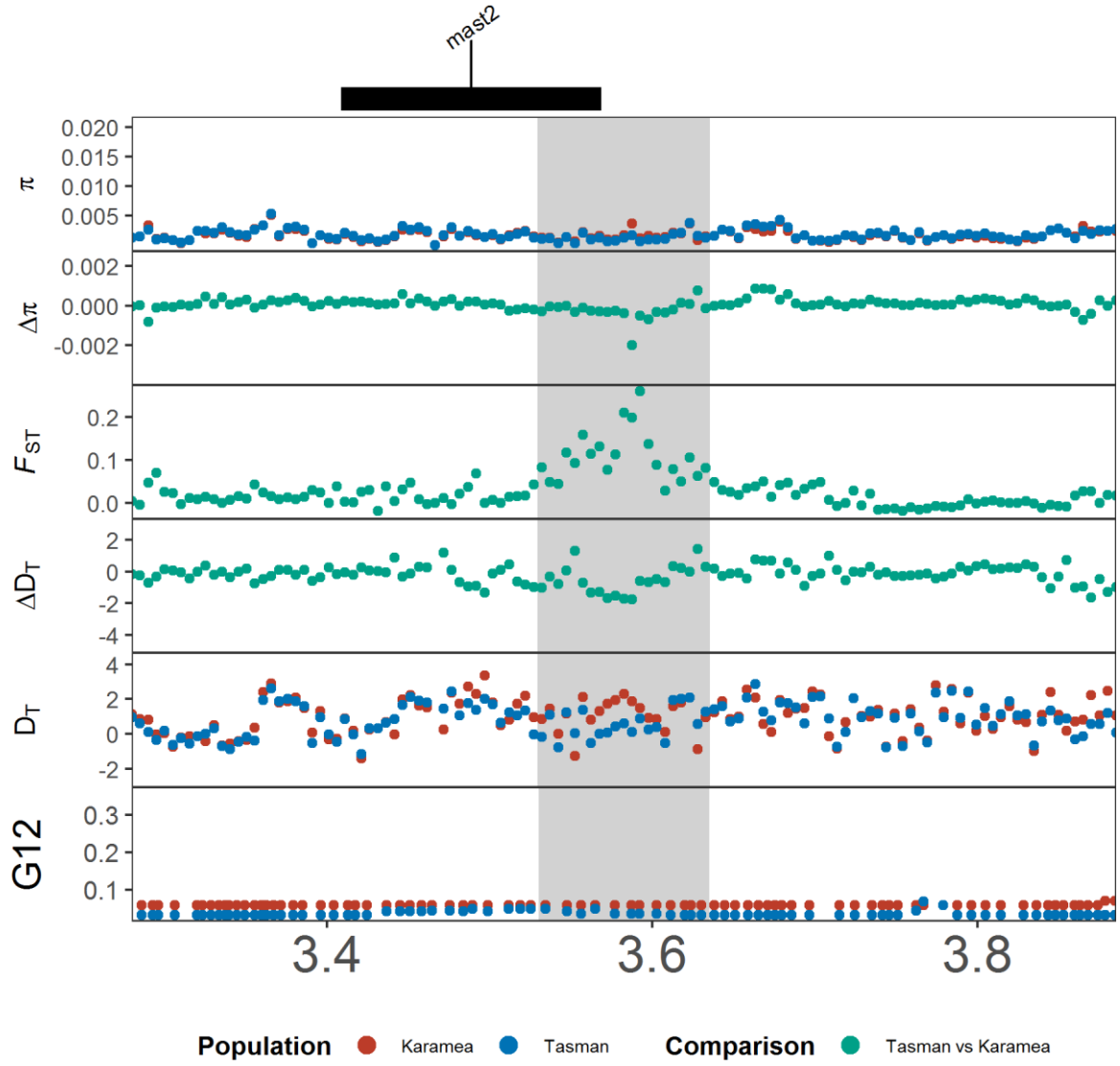

Figure S4.4: Divergent selection between Tasman Bay and Karamea Bight occurring on linkage group 7. This gene was also shown to experience local selection in the southern cluster.

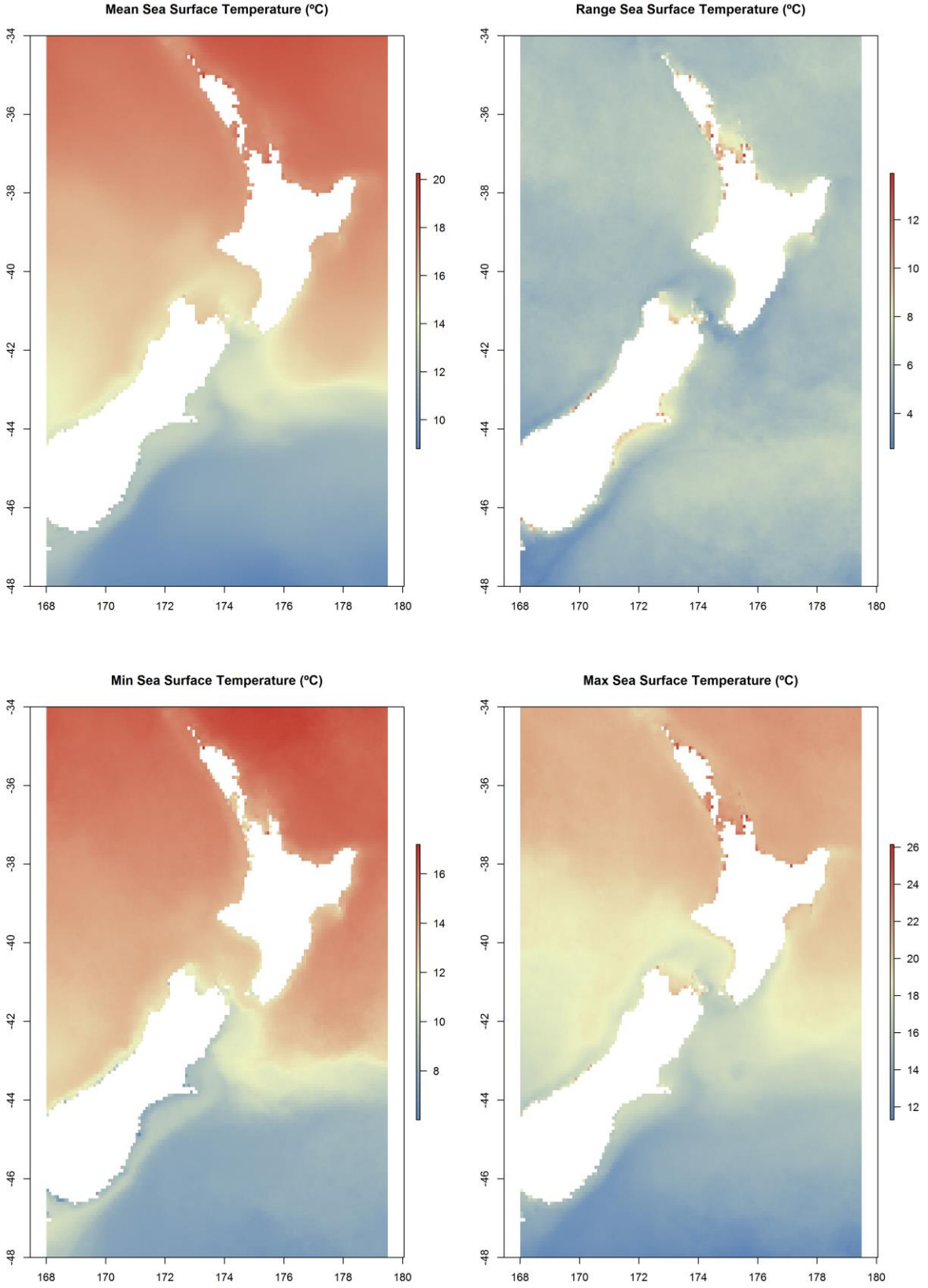

Figure S4.5: Variation in sea surface temperature across New Zealand to show the variation in temperature down the West Coast of New Zealand.

Chapter 5. General Discussion

This thesis generated the first whole-genome population data set for snapper using modern contemporary and ancient samples. The analyses of genome-wide markers provided an in-depth picture of the spatial distribution of variation, evolutionary history, and evidence of adaptive variation. In this chapter, the main findings of the research are summarised for each chapter, their relevance discussed and subsequently placed in a broader context.

5.1. Main findings

5.1.1. High levels of genetic connectivity between populations and presence of two genetic clusters

In Chapter 2, the use of 167,543 neutrally evolving loci showed the distribution of genetic diversity and connectivity between sampled sites. The AMOVA showed that 94.59% of neutral variation is observed within individuals, suggesting significant levels of gene flow between populations. Pairwise comparisons of genetic differentiation (F_{ST}) showed low levels of genetic differentiation, with significant evidence for isolation by distance (R^2 = 0.75, p-value = 0.002). This is typical for fish species with large population sizes and high chance of dispersal. Two genetic disjunctions were detected around Cape Reinga and Mahia Peninsula using F_{ST}, PCA, and DAPC creating two distinct genetic clusters (i.e. northern and southern cluster). Multiple individuals were sampled in areas that did not reflect their population of origin, showing evidence of movement between genetic clusters. Gisborne is an area where both genetic clusters appear to be mixing, but there was no evidence of reproduction in this area. Gene flow between the two stocks is likely occurring on the West Coast and Hawke's Bay, as indicated by high genetic similarities. In previous studies, this genetic similarity had been related to high levels of connectivity between the two locations via the D'Urville current. However, the results of this study suggest that the genetic similarity is caused by gene flow with the same genetic cluster (i.e. northern cluster). Levels of genetic diversity were not significantly different between populations, suggesting that fishing pressure has not lead to local reductions in genetic variation. This result is in contrast to previous studies that reported loss of genetic variation using microsatellites in Tasman Bay (Hauser et

al., 2002). Similar levels of genetic variation are likely maintained by the high level of connectivity between populations.

5.1.2. Detection of adaptive genetic variation relevant for the identification of demographically independent populations

Chapter 2 included the identification of putative adaptive SNPs. These SNPs were indicative that certain regions of the genome that resist homogenisation from migration exchange between different populations. Local environments tend to favour the reproductive success of the adapted resident individuals. Detection of putative adaptive SNPs using all samples from the populations identified 134 outliers with high confidence (false discovery rate (FDR) < 1%). The PCA produced using these SNPs showed a similar pattern compared to neutral SNPs on the first principle component (PC1), suggesting that there are environmental differences between the two genetic clusters. The second principle component (PC2) separated individuals from Kapiti Coast from the rest of the southern cluster. It is possible that individuals from Kapiti Coast are genetically related to a population that is thought to be present in the Marlborough Sound (Fisheries New Zealand, 2018).

More fine-scale population structure was detected when adaptive SNPs were identified for each genetic cluster. Here, reduced sample sizes, and smaller genetic differences required acceptance of higher false discovery rates (FDR < 10%). In the northern cluster, 30 adaptive SNPs showed that Gisborne was the most genetically differentiated. This was expected because the population was observed to consist of a mixture of northern and southern individuals. East Cape formed a genetically distinct cluster, while Northland, Hauraki Gulf, and Bay of Plenty overlapped. The identification of 40 putative adaptive SNPs that were able to segregate the three populations required acceptance of high false discovery rates (FDR < 50%). This implies that putative adaptive SNPs that are able to separate Northland, Hauraki Gulf, and Bay of Plenty, should be interpreted with extreme care.

In the southern cluster, the identification of 53 putative adaptive SNPs separated individuals from Kapiti Coast, Karamea Bight and Hawke's Bay into individual clusters. While Kapiti Coast and Hawke's Bay may represent demographically independent stocks, Karamea Bight is hypothesized to represent an adaptive front that is adapting to

colder waters (discussed in Chapter 4). Reproductively, Karamea Bight is likely dependent on Tasman Bay because this is the closest known spawning ground. Finally, 120 putative adaptive SNPs that were able to separate Tasman Bay and West coast individuals were identified with high confidence (FDR < 1%). This shows that the two populations may constitute demographically independent stocks. This was expected based on the known locations of breeding grounds, and monitoring data (Fisheries New Zealand, 2018; Parsons et al., 2014). The ability to differentiate between demographically independent populations is extremely valuable for fisheries management. The patterns detected here show promising results for future implementation into fisheries management of snapper stocks. The addition of more samples and use of structural variation is likely to provide additional statistical power and insights.

5.1.3. The influence of glacial cycles on the demographic history of snapper

In Chapter 3, the use of contemporary and ancient samples showed that snapper experienced strong phylogeographic structuring. Mitochondrial genomes indicated the presence of two lineages that diverged approximately 650,000 (490,000 – 840,000) years ago. The separation of populations was likely caused by reductions in sea level during glacial cycles. This hypothesis is most consistent with observations in other species that have similar phylogenetic lineages (Hewitt, 2000). It is unclear where the populations became separated, but it is highly likely to have occurred before snapper arrived in New Zealand (i.e. 650,000 years ago). It is possible that populations were already separated in the Indo-Pacific Ocean, a region known for facilitating phylogeographic structuring during glacial periods (Hewitt, 2000). This idea is supported by the observation that the split between the two phylogeographic clusters was not significantly different from the time estimate when *C. auratus* and *P. major* split 727,000 years ago (591,000-865,000).

Estimates of changes in population size show strong support for an exponential population size increase after the LGM. Changes in population abundance based on the Bayesian Skyline plot suggested a strong population increase approximately 10,000 years ago. The strong increase in new branches in the phylogenetic tree suggests population sizes increase approximately 20,000 (7,000-35,000) years ago. A post-glacial expansion is the most likely explanation for the observed increase in population abundance. This

takes into account inherent uncertainties with estimating the timing of events, and knowledge about key drivers of evolutionary change during this period. During this period, sea levels rose which reconnected fragmented populations, and increased sea temperatures allowed for southward expansion. These results show that current climate change is likely to continue the expansion of suitable snapper habitat, as is supported by ecological niche modelling (Brooks, 2019).

5.1.4. The association between genes under selection and traits related to growth, maturation, and cold-stress

In Chapter 4, selection scans identified genomic regions under selection and investigated which genes and traits are experiencing evolutionary change. The first selection scan contrasted the two genetic clusters that were identified in Chapter 2 (i.e. North and South cluster). The scan identified 101 genomic regions that showed evidence of selection; these contained 253 genes. Three genomic regions showed strong evidence of divergent (local) selection, which suggested the clusters are experiencing different selective pressures. One of the regions was associated with a retrotransposon (*gin1*) that could not be related to a specific biological function. The other two regions contained two genes near the peak of divergence that are associated with glycolysis which influences (cell-) growth (i.e. *mast2* and *hk2*). The regions containing *hk2* showed a lack of rare alleles ($T_D > 2$) in the southern cluster, consistent with balancing selection maintaining multiple alleles in the population. It is possible that variation in growth rate is maintained throughout the genetic cluster because of a latitudinal gradient in sea temperature. Temperature is known to influence growth in fish (Munday et al., 2017), including in snapper (Wellenreuther et al., 2019a). This temperature gradient is not present in the northern cluster, which would explain why multiple alleles are not maintained in the northern cluster.

Two genomic regions showed strong evidence of a selective sweep ($G12 > 0.2$ & $T_D < -2$) in both genetic clusters (nation-wide), indicative of recent or ongoing selection. Three genes of interest were identified within these regions (i.e. *mydgf*, *matk*, and *rnf213a*). *Mydgf* and *rnf213a* were associated with angiogenesis, which involves the development of new blood vessels. *Matk* is involved in a signalling pathway that has been shown to affect a wide range of biological functions, including glycolysis and angiogenesis.

Evidence of evolutionary changes in angiogenesis could be of particular interest because it was shown to affect maturation rates in threespine sticklebacks (*Gasterosteus aculeatus*) (Furin et al., 2015). Maturation is one of the life-history traits associated with fishing-induced evolution (Enberg et al., 2012). While the connection between the genes experiencing a selective sweep and maturation is speculative, it is noteworthy that the two regions that show evidence for recent or ongoing selection are potentially linked to adaptive responses observed in many fished species (Heino et al., 2015).

The second selection scan involved the identification of divergent selection between Tasman Bay and Karamea Bight. The hypothesis was that individuals from Karamea Bight represent an adaptive front that is adapting to colder waters. Selection was detected in 123 genomic regions containing 296 genes, of which 197 potentially experience divergent selection. Two genes that showed significant divergent selection were also observed to be differentially expressed in a study that looked at adaptation to cold stress in yellow drum (*Nibea albiflora*) (Xu et al., 2018). *Notch2* is involved in pathways signalling but could not directly be associated with biological processes related to cold stress. However, *camk2g* is a subunit from *camk2* which has a well-established function in relation to cold stress (Cao et al., 2004). Finally, Karamea Bight showed evidence of a selective sweep associated with *ksr2*. This gene has been shown to regulate energy balance via feeding behaviour and adaptive thermogenesis (Guo et al., 2017), which is related to cold stress.

5.2. Relevance

5.2.1. Implications for snapper fisheries in New Zealand

Snapper is one of New Zealand's largest inshore fisheries species. To effectively manage the species it is important that management areas accurately reflect demographically independent units. Currently, snapper are managed using five management areas with the majority of fishing occurring in two of these (SNA1 and SNA8). Stock assessments have already identified the presence of multiple demographically independent populations within these management areas (Parsons et al., 2014). The identification of genetic clusters that are consistent with the previously identified populations is promising. It shows that genomic methods provide a level of resolution that is relevant for applied fisheries management. The use of adaptive variation also identified the presence of genetically distinct populations along the Kapiti Coast. This had not been reported in previous studies. While speculative, it is possible that this population is related to a population present in the Marlborough Sounds. Findings like this help to refine stock boundaries for fisheries management.

Improvements in marker identification are still needed, particularly to distinguish between the populations present in SNA1. The three recognized populations are genetically very similar, and show high levels of connectivity. Increased sample sizes would provide more statistical power to detect outliers. Also, the use of structural variation is more likely to identify relevant adaptive variation that could detect more fine-scale population structure. Structural variation has been shown to play important roles in the adaptation to local environments (Hoban et al., 2016), and in snapper structural variation outweighs the number of SNPs three-fold (Catanach et al., 2019).

While adaptive variation plays an important role in stock identification, the use of thousands of neutrally evolving loci provide useful tools for monitoring genetic diversity, population abundances, and connectivity between populations (Luikart et al., 2018). Chapter 2 showed how snapper consists of two genetic clusters and that movement of individuals between populations is homogenising non-adaptive (neutral) variation. This implies that populations such as Tasman Bay may not be as genetically isolated as previously thought (Bernal-Ramírez et al., 2003; Hauser et al., 2002). It also

showed that populations that were thought to have high levels of connectivity (i.e. West Coast and Hawke's Bay) (Bernal-Ramírez et al., 2003; Smith et al., 1978), appear to be genetically similar, presumably because they experience gene flow with the northern cluster.

Genome-wide estimates of heterozygosity showed that there are no significant differences in genetic variation between populations (p = 0.069). This is in contrast to a previous study that reported loss of genetic variation in the Tasman Bay population due to fishing (Hauser et al., 2002). This raised serious concerns regarding the effects of fishing on maintaining genetically viable stocks. The study used microsatellite data to detect temporal changes in heterozygosity. This thesis did not include a temporal comparison within the same population, which made it impossible to corroborate the startling findings with an investigation of genome-wide levels of heterozygosity over time. It is possible that gene flow between Tasman Bay and other populations has since increased diversity, but there is currently no way of testing this hypothesis. Regardless, it is promising to see that Tasman Bay has genome-wide levels of heterozygosity that are similar to other populations. It implies that levels of genetic diversity can be maintained in populations that experienced large population reductions (Pinsky et al., 2021). While levels of genetic diversity in snapper appear stable, it is less clear whether intense size-selective fishing of snapper has caused evolutionary change.

5.2.2. Is there evidence for fishing-induced evolution in snapper?

Intense size-selective fishing has been suggested to cause significant reductions in growth rate and select for faster maturation (Heino et al., 2015). One important question is whether genetic changes are associated with the observed shifts in key life-history traits. Experiments have shown that size-selective removal of fish is able to modulate fish size and this could be correlated with changes in allele frequency (Therkildsen et al., 2019; van Wijk et al., 2013). However, possible confounding factors make it difficult to establish causal links between environmental factors and genes that show a pattern consistent with selection. Snapper is a prime candidate to test for fishing induced evolution because this species has experienced intense size-selective fishing pressure since the 1950s. Moreover, archaeological work also suggests the snapper has

reduced in average size (Leach, 2006).

In Chapter 4, two genomic regions in linkage group 7 showed significant evidence for increased homozygosity on a nation-wide level (G12 > 0.2), consistent with a selective sweep suggesting recent or ongoing selection. Selective sweeps are the best candidate for detecting fishing-induced evolution because local reductions in variation are one of the first detectable signals that selection is occurring (Sabeti et al., 2006; Vitti et al., 2013). Intense size-selective fishing in snapper is a relatively recent selective pressure on evolutionary timescales (100s and not 1000s of years).

There were three genes of significant interest located within the two regions (i.e. *mydgf*, *matk*, and *rnf213a*). These genes were either directly or via signalling pathways related to angiogenesis, a biological process that has been shown to affect maturation in threespine sticklebacks (*Gasterosteus aculeatus*) (Furin et al., 2015). While tentative, it is noteworthy that the two regions that show evidence for recent or ongoing selection are potentially linked to life-history traits associated with fishing-induced evolution, and definitely worth exploring in more detail. There could be a number of confounding environmental conditions that facilitate a similar evolutionary change.

5.2.3. Impacts of climate on snapper

Chapter 3 suggests that changes in global climate appear to have had a major impact on the evolutionary history of snapper. First, mitochondrial lineages that date back to approximately 650,000 years ago were possibly established through population fragmentation caused by reductions in sea levels during glacial maxima. While there is no direct evidence that glacial cycles caused this phylogeographic structuring, it consistent with what we know about phylogeographic processes in marine species during these time periods (Bowen et al., 2016). Second, analyses indicated that the LGM caused an exponential populations increase during the last 20,000 years. During this period, sea-level rise and increased ocean temperatures would have increased suitable habitat for snapper around New Zealand, allowing populations to expand. A causal connection between the ending of the LGM and the population increase in snapper around New Zealand is lacking. However, I argue that this is the most likely scenario, given what we know about the ecological preferences of snapper and changes in ocean habitat around the LGM. These two events suggest that snapper tend to experience

strong phylogeographic structuring and population reductions during glacial cycles.

Chapter 4 showed evidence that the local climate facilitates selection in contemporary snapper populations. A temperature gradient in the southern cluster could explain the detection of balancing selection near a gene associated with growth (*hk2*). Temperature is known to influence growth rates in fish (Munday et al., 2017), including snapper (Wellenreuther et al., 2019a). Multiple alleles could be maintained within the southern cluster because different alleles are being selected across the temperature gradient. This would also explain why balancing selection is not observed in the North cluster where this temperature gradient is not present. Another example of local climate potentially driving selection is around Karamea Bight, where water temperatures are considered to be on the boundary of the suitable habitat for snapper. Evidence of divergent selection between Karamea Bight and Tasman Bay near genes associated with cold stress (*camk2g* and *ksr2*) suggest that Karamea Bight individuals could represent an adaptive front.

These results suggest that snapper is primarily affected during periods of global cooling, and their distribution in part dictated by minimum ocean temperatures. With ocean temperatures expected to keep rising, it seems unlikely that snapper will experience direct deleterious effects from increased water temperatures in the near future. Snapper populations may even become more productive and continue to expand southwards as ocean temperatures continue to increase (Parsons et al., 2014). Increased temperatures could also affect patterns of reproduction. Snapper spawning is initiated when surface temperatures reach 14.8-16 C° around October (Scott & Pankhurst, 1992). Higher average ocean temperatures could shift the timing to earlier. This idea is supported by previous work based on ecological niche modelling (Brooks, 2019). The effects of climate change on snapper will unfortunately not be limited to increased ocean temperatures. Changes in ocean currents, ecological interactions, ocean acidification, and increase in extreme weather events could all pose serious risks for snapper in the future (Manning et al., 2015; Parsons et al., 2014). It is hence difficult to predict how snapper will be impacted by climate change in the future.

5.2.4. Application of aDNA in fisheries research

Advances in molecular methods have made it possible for ancient genomics to be applied to non-model species, including teleosts (Ferrari et al., 2021). The application of ancient genomics holds great potential to address questions about the evolutionary ecology of fish species that would otherwise be impossible (Oosting et al., 2019). One of the main limitations for advancing fish aDNA research is sample preservation. Fish bones are brittle, porous and very light, which is generally associated with poor DNA preservation. Sample preservation was also what limited the application of aDNA in this thesis.

One of the initial aims of this thesis was to test for fishing-induced evolution using ancient genomic variation to detect changes in allele frequencies over time. Our approach relied on the identification of samples with high levels of endogenous DNA (~50%) and then use those for whole-genome shotgun sequencing to obtain nuclear genomes. This approach was based on the observation that high levels of endogenous DNA could be extracted in other fish species (Boessenkool et al., 2017; Star et al., 2017). However, only one aDNA sample met these requirements, while the majority of other aDNA samples had levels of endogenous DNA below 10%.

Given what we know now about the average level of sample preservation of ancient snapper bones, it is important to assess what research methods and questions are feasible for future research. The time and resources required to identify enough high-quality ancient DNA sample with endogenous DNA levels above 50% using our initial approach turned out to be time-consuming and costly. Large numbers of ancient samples need to be extracted, quality checked, and skim-sequenced to obtain enough nuclear ancient genomes to get reliable allele frequency estimates (minimum of 15-20). Such an approach is unlikely to be accepted by most archaeological curators as it generally involves destructive sampling of a finite resource. If the sequencing of ancient nuclear markers is perused in the future the use of RNA-baits could provide a more reliable solution (Carpenter et al., 2013). Using this method, samples with >1% endogenous DNA can be used to obtain nuclear markers to test for evidence of selection on the genome over time, e.g. fishing-induced evolution. RNA-baits could be designed to target the exome using the transcriptome so only coding regions are extracted for

sequencing. This method is also time and resource-intensive which should be taken into consideration.

The sequencing of complete mitochondrial genomes had much higher success, with nearly one-third of extracted samples providing a complete mitochondrial genome. Sequencing of more mitochondrial genomes via shotgun sequencing is therefore a viable approach for snapper. Mitochondrial genomes provided fascinating insights regarding the evolutionary history of snapper. Adding more ancient mitochondrial genomes could provide better estimates for the timing of demographic events. This could provide information on what key environmental conditions could have initiated the exponential population size increase after the LGM.

5.2.5. Integration of genomics into fisheries management

The integration of genomics into fisheries research has been slow (Bernatchez et al., 2017). However, the recent advances in molecular methods and the reduction in price now allow genomics to provide impactful contributions to fisheries management.

The identification of demographically independent stocks is one of the focal points for using genomics in fisheries management. Genome-wide sequencing allows the genotyping of thousands of genetic makers and the identification of adaptive variation. Adaptive variation reflects parts of the genome that resist the homogenising effect of gene flow due to differences in environmental and ecological conditions between populations. Such genomic approaches have been used to detect fine-scale population structure between populations that conventional genetic methods were not able to detect (Barth et al., 2017; Benestan et al., 2015; Berg et al., 2016; Euclide et al., 2021; Petrou et al., 2021). This includes the results from Chapter 2. The identification of relevant markers using genomics can be used to design SNP panels that allow for low-cost genotyping of large numbers of individuals and population assignment based on these.

The increase in statistical power using genomics also enables more accurate assessments of population dynamics to account for such movement between populations. Large numbers of neutral markers can be used to estimate the level of gene flow, providing insights into the number of migrants that contribute to different gene

pools. Both neutral and adaptive markers can be used for genetic tagging and genetic assignment which provides information on the physical number of individuals that migrate regardless of reproductive success. Such methods could potentially replace more traditional tagging methods that tend to be more invasive. An extension to this method is close-kin-mark-recapture which does not require the same individuals to be tagged twice, but uses closely related individuals to estimated stock sizes (Bravington et al., 2016). There are definite drawbacks to using this method on fish because extremely large sample sizes are needed. This makes it difficult to design marker panels and to validate methods before implementing them on an industrial scale.

Fisheries management strategies often do not incorporate population genetic information for monitoring population trends, nor as a metric to quantify the long-term stability and resilience of a population. Current management goals primarily focus on maximising harvest targets, while ensuring that stock abundances remain stable (BMSY). The efficiency of these management strategy evaluation (MSE) based approaches heavily rely on active monitoring of the biological components of the population. Approaches such as monitoring changes in allele frequencies ($F_{temporal}$) could be very powerful in detecting reductions early on (Allendorf et al., 2010; Schwartz et al., 2007). The power of this method is highly dependent on the number of alleles, which is the main strength of genomics.

Apart from using genetic variation to monitor population trends, future discussions on sustainable fishing should include conversations about what levels of genetic diversity are needed to maintain genetically healthy and resilient stocks. One important question is what type of genetic diversity best represents the health of a population. For example, Chapter 2 raised the question of whether microsatellites would be an appropriate marker to monitor reductions in genetic diversity relevant to populations (Hauser et al., 2002). It is arguable that a combination of neutral and adaptive markers sampled across the genome will provide a much more useful metric to monitor genetic diversity. Such measurements are currently not included in MSE strategies, however, they could be vital to populations that are facing numerous environmental conditions.

5.2.6. How genomics is changing how we study biodiversity via adaptive genetic variation

The use of genomics has increased drastically over the last 10 years (Mardis, 2017). This is in large part due to the advances in sequencing technologies, reduced costs of sequencing, increase in genomic resources (e.g. number of annotated reference genomes), and development of software packages suited for whole-genome analyses. This has had a major impact on the study of population demography and eco-evolutionary processes in non-model species.

The genotyping of thousands of loci across the genome has enabled the identification of adaptive variation, providing valuable insights into the evolution of species. Adaptive variation allows for the identification of populations that are biologically relevant, highlighting the adaptation of populations to their local environment. Chapter 2 showed how this can be used to detect fine-scale population structure. The use of adaptive variation to study populations has been applied in a wide range of species including teleost fish (Vaux et al., 2021), crustaceans (Al-Breiki et al., 2018), birds (Lin et al., 2015), and mammals (Zhang et al., 2018). This is one of the exciting aspects of genomics because it provides insights into the evolution of species that conventional genetics was unable to provide (Luikart et al., 2018).

The true potential of genomics is the ability to identify the location of thousands of genetic markers in the genome, compare genomic regions, identify genes under selection, and study their interactions. The number of species for which this is possible has increased exponentially due to the fast-growing number of available (annotated) reference genomes (Fan et al., 2020; Koepfli et al., 2015). Combined with phenotypic and environmental data this information can be used to study which genes are associated with polygenic traits (Wellenreuther & Hansson, 2016), and which environmental conditions could be causing evolutionary change (Coscia et al., 2020). Chapter 4 showed how the application of such genomic resources leads to fascinating new insights into how natural selection affects the genome. Other examples include the identification of genes underlying the adaptive radiation of butterflies (Moest et al., 2020), the identification of adaption in plants populations possibly related to host-pathogen interaction (Bourgeois et al., 2018), and genes underlying the diversifying selection in

selected horse breeds (Gurgul et al., 2019). These examples highlight how adaptive variation can be used to study a wide-range species and of applications.

The increase in the number of species with genomic data sets enables comparative genomics to observe evolutionary patterns on a larger scale (Ellegren, 2008; Zhang et al., 2014). Chapter 4 was able to correlate genes under putative selection and adaptation to cold stress because previous studies reported similar results. Similar observations were reported between studies that looked at growth (Ahi, 2016; Jacobs et al., 2019; Jacobs et al., 2018), and salinity stress (Boutet et al., 2006; Momigliano et al., 2017; Zhang et al., 2016). Parallel adaptive evolution in response to temperature was observed in Atlantic cod (*Gadus morhua*) on opposite sides of the Atlantic (Bradbury et al., 2010). This is a specific example that shows how comparative studies between populations or species allows for the identification of predictable evolutionary responses. The ability to predict with confidence is the ultimate sign that science has developed into a meaningful body of theory. The detection of these overlapping adaptive responses will increase as adaptive genomic data is being generated for species. I propose that it is time for meta-studies to be done to identify trends in adaptive variation. This will enable large-scale evolutionary patterns to be observed, providing a better understanding of the evolutionary processes that influence biodiversity.

5.3. Conclusion

This thesis presents the first population genomic study of Australasian snapper in New Zealand, a species with a diverse genetic landscape and a rich evolutionary history. The detection of fine-scale population structure through adaptive differences between populations highlights the promising application of genomics in fisheries management. The study of mitochondrial lineages showed the effect of glacial cycles, providing insights into how New Zealand's marine fauna has been affected by major changes in global climate. Finally, the identification of genes and associated biological traits under selection has provided fundamental new insights regarding the environmental conditions that drive adaptive change and act on phenotypes.

Much can be learned about biodiversity through studying the genetic similarities between individuals. The research included in this thesis only scratches the surface of

the knowledge to be gained about snapper using genomics. Future efforts will continue the exploration of the diversity present in this species across time and space in an effort to expand our understanding of how past eco-evolutionary processes shaped the distribution of genetic variation in the present, and how it can be used to improve management of snapper in the future.

References

Ahi, E. P. (2016). Signalling pathways in trophic skeletal development and morphogenesis: insights from studies on teleost fish. *Developmental biology, 420*(1), 11-31.

Ahn, N. G., Seger, R., Bratlien, R. L., Diltz, C. D., Tonks, N. K., & Krebs, E. G. (1991). Multiple components in an epidermal growth factor-stimulated protein kinase cascade: In vitro activation of a myelin basic protein/microtubule-associated protein 2 kinase*. *Journal of Biological Chemistry, 266*(7), 4220-4227. doi:https://doi.org/10.1016/S0021-9258(20)64310-1

Aktas, C. (2015). Haplotypes: Haplotype inference and statistical analysis of genetic variation. *R package version, 1*.

Al-Breiki, R. D., Kjeldsen, S. R., Afzal, H., Al Hinai, M. S., Zenger, K. R., Jerry, D. R., . . . Delghandi, M. (2018). Genome-wide SNP analyses reveal high gene flow and signatures of local adaptation among the scalloped spiny lobster (Panulirus homarus) along the Omani coastline. *BMC genomics, 19*(1), 1-13.

Aljanabi, S. M., & Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Res, 25*(22), 4692-4693. doi:10.1093/nar/25.22.4692

Allendorf, F. W., England, P. R., Luikart, G., Ritchie, P. A., & Ryman, N. (2008). Genetic effects of harvest on wild animal populations. *Trends Ecol Evol, 23*(6), 327-337. doi:10.1016/j.tree.2008.02.008

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nat Rev Genet, 11*(10), 697-709. doi:10.1038/nrg2844

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. In: Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Anglesio, M. S., Arnold, J. M., George, J., Tinker, A. V., Tothill, R., Waddell, N., . . . Traficante, N. (2008). Mutation of ERBB2 provides a novel alternative mechanism for the ubiquitous activation of RAS-MAPK in ovarian serous low malignant potential tumors. *Molecular cancer research, 6*(11), 1678-1690.

Annala, J. H. (1996). New Zealand's ITQ system: have the first eight years been a success or a failure? *Reviews in Fish Biology and Fisheries, 6*(1), 43-62.

Araujo, H. A., Candy, J. R., Beacham, T. D., White, B., & Wallace, C. (2014). Advantages and challenges of genetic stock identification in fish stocks with low genetic resolution. *Transactions of the American Fisheries Society, 143*(2), 479-488.

Ashton, D. (2013). *Population Genetics of New Zealand Pagrus auratus and Genetic Variation of an Aquaculture Broodstock.* (Master), Victoria University of Wellington, Wellington.

Ashton, D. T., Hilario, E., Jaksons, P., Ritchie, P. A., & Wellenreuther, M. (2019a). Genetic diversity and heritability of economically important traits in captive Australasian snapper (Chrysophrys auratus). *Aquaculture, 505*, 190-198.

Ashton, D. T., Ritchie, P. A., & Wellenreuther, M. (2019b). High-density linkage map and QTLs for growth in snapper (Chrysophrys auratus). *G3: Genes, Genomes, Genetics*, g3. 200905.202018.

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., . . . Saunders, N. C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics, 18*(1), 489-522.

Barluenga, M., Stölting, K. N., Salzburger, W., Muschick, M., & Meyer, A. (2006). Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature, 439*(7077), 719-723.

Barot, S., Heino, M., O'Brien, L., & Dieckmann, U. (2004). Long-term trend in the maturation reaction norm of two cod stocks. *Ecological Applications, 14*(4), 1257-1271.

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution, 23*(1), 38-44. doi:https://doi.org/10.1016/j.tree.2007.09.008

Barrows, T. T., Juggins, S., De Deckker, P., Calvo, E., & Pelejero, C. (2007). Long-term sea surface temperature and climate change in the Australian–New Zealand region. *Paleoceanography, 22*(2).

Barth, J. M., Berg, P. R., Jonsson, P. R., Bonanomi, S., Corell, H., Hemmer-Hansen, J., . . . Knutsen, H. (2017). Genome architecture enables local adaptation of Atlantic cod despite high connectivity. *Molecular Ecology, 26*(17), 4452-4466. doi:10.1111/mec.14207

Beddington, J. R., Agnew, D. J., & Clark, C. W. (2007). Current problems in the management of marine fisheries. *science, 316*(5832), 1713-1716.

Begg, G. A., Friedland, K. D., & Pearce, J. B. (1999). Stock identification and its role in stock assessment and fisheries management: an overview. *Fisheries research, 43*(1), 1-8. doi:https://doi.org/10.1016/S0165-7836(99)00062-4

Bendich, A. J. (1987). Why do chloroplasts and mitochondria contain so many copies of their genome? *BioEssays, 6*(6), 279-282.

Benestan, L. (2019). Population genomics applied to fishery management and conservation. In *Population Genomics: Marine Organisms* (pp. 399-421): Springer.

Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., & Bernatchez, L. (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the a Merican Lobster (H Omarus americanus). *Molecular Ecology, 24*(13), 3299-3315.

Berg, P. R., Star, B., Pampoulie, C., Sodeland, M., Barth, J. M. I., Knutsen, H., . . . Jentoft, S. (2016). Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Scientific Reports, 6*(1), 23246. doi:10.1038/srep23246

Bernal-Ramírez, J., Adcock, G., Hauser, L., Carvalho, G., & Smith, P. (2003). Temporal stability of genetic population structure in the New Zealand snapper, Pagrus auratus, and relationship to coastal currents. *Marine Biology, 142*(3), 567-574.

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

Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M., Beacham, T. D., . . . Naish, K. A. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends in Ecology & Evolution, 32*(9), 665-680.

Berry, O., Tocher, M. D., & Sarre, S. D. (2004). Can assignment tests measure dispersal? *Molecular Ecology, 13*(3), 551-561.

Biswas, S., & Akey, J. M. (2006). Genomic insights into positive selection. *Trends in Genetics, 22*(8), 437-446.

Blackwell, R., & Gilbert, D. (2006). Age composition of commercial snapper landings in SNA 2. *New Zealand Fisheries Assessment Report*, 46.

Blackwell, R., & Gilbert, D. (2008). Age composition of commercial snapper landings in Tasman Bay/Golden Bay (SNA 7), 2006–07. *New Zealand Fisheries Assessment Report, 67*, 22.

Blunier, T., & Brook, E. J. (2001). Timing of Millennial-Scale Climate Change in Antarctica and Greenland During the Last Glacial Period. *science, 291*(5501), 109-112. doi:10.1126/science.291.5501.109

Boessenkool, S., Hanghøj, K., Nistelberger, H. M., Der Sarkissian, C., Gondek, A. T., Orlando, L., . . . Star, B. (2017). Combining bleach and mild predigestion improves ancient DNA recovery from bones. *Molecular ecology resources, 17*(4), 742-751.

Bollongino, R., Tresset, A., & Vigne, J.-D. (2008). Environment and excavation: Pre-lab impacts on ancient DNA analyses. *Comptes Rendus Palevol, 7*(2), 91-98. doi:https://doi.org/10.1016/j.crpv.2008.02.002

Bosch, S., Tyberghein, L., & De Clerck, O. (2017). sdmpredictors: an R package for species distribution modelling predictor datasets. *Marine Species Distributions: From data to predictive models, 49*.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., . . . Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol, 10*(4), e1003537.

Bourgeois, Y., Stritt, C., Walser, J. C., Gordon, S. P., Vogel, J. P., & Roulin, A. C. (2018). Genome-wide scans of selection highlight the impact of biotic and abiotic constraints in natural populations of the model grass Brachypodium distachyon. *The Plant Journal, 96*(2), 438-451.

Boutet, I., Ky, C. L., & Bonhomme, F. (2006). A transcriptomic approach of salinity response in the euryhaline teleost, Dicentrarchus labrax. *Gene, 379*, 40-50.

Bowen, B. W., Gaither, M. R., DiBattista, J. D., Iacchei, M., Andrews, K. R., Grant, W. S., . . . Briggs, J. C. (2016). Comparative phylogeography of the ocean planet. *Proc Natl Acad Sci U S A, 113*(29), 7962-7969. doi:10.1073/pnas.1602404113

Bowen, B. W., Muss, A., Rocha, L. A., & Grant, W. S. (2006). Shallow mtDNA Coalescence in Atlantic Pygmy Angelfishes (Genus Centropyge) Indicates a Recent Invasion from the Indian Ocean. *Journal of Heredity, 97*(1), 1-12. doi:10.1093/jhered/esj006

Bradbury, I. R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I. G., . . . Bentzen, P. (2010). Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society B: Biological Sciences, 277*(1701), 3725-3734. doi:doi:10.1098/rspb.2010.0985

Brandies, P., Peel, E., Hogg, C. J., & Belov, K. (2019). The value of reference genomes in the conservation of threatened species. *Genes, 10*(11), 846.

Brandström, M., & Ellegren, H. (2008). Genome-wide analysis of microsatellite polymorphism in chicken circumventing the ascertainment bias. *Genome research, 18*(6), 881-887.

Bravington, M. V., Skaug, H. J., & Anderson, E. C. (2016). Close-kin mark-recapture. *Statistical Science, 31*(2), 259-274.

Breier, G. (2000). Angiogenesis in embryonic development—a review. *Placenta, 21*, S11-S15.

Briggs, J. C. (1987). Antitropical distribution and evolution in the Indo-West Pacific Ocean. *Systematic Zoology, 36*(3), 237-247.

Broad Institute. (2020). Picard toolkit. Retrieved from GitHub Repository. http://broadinstitute.github.io/picard/

Brookes, A. J. (1999). The essence of SNPs. *Gene, 234*(2), 177-186. doi:https://doi.org/10.1016/S0378-1119(99)00219-X

Brooks, A. (2019). *Using Ecological Niche Modelling to Predict Climate Change Responses of Ten Key Fishery Species in Aotearoa New Zealand.* (Master), Victoria University of Wellington,

Bull, C. Y., Kiss, A. E., van Sebille, E., Jourdain, N. C., & England, M. H. (2018). The role of the New Zealand plateau in the Tasman Sea circulation and separation of the East Australian Current. *Journal of Geophysical Research: Oceans, 123*(2), 1457-1470.

Bushnell, B. (2017). BBMap. Retrieved from sourceforge.net/projects/bbmap/

Caddy, J. F., & Cochrane, K. L. (2001). A review of fisheries management past and present and some future perspectives for the third millennium. *Ocean & Coastal Management, 44*(9), 653-682. doi:https://doi.org/10.1016/S0964-5691(01)00074-6

Cano, J. M., Shikano, T., Kuparinen, A., & Merilä, J. (2008). Genetic differentiation, effective population size and gene flow in marine fishes: implications for stock management. *Journal of Integrated Field Science, 5*, 1-10.

Cao, W., Daniel, K. W., Robidoux, J., Puigserver, P., Medvedev, A. V., Bai, X., . . . Collins, S. (2004). p38 Mitogen-Activated Protein Kinase Is the Central Regulator of Cyclic AMP-Dependent Transcription of the Brown Fat Uncoupling Protein 1 Gene. *Molecular and Cellular Biology, 24*(7), 3057-3067. doi:10.1128/mcb.24.7.3057-3067.2004

Carpenter, M. L., Buenrostro, J. D., Valdiosera, C., Schroeder, H., Allentoft, M. E., Sikora, M., . . . Bustamante, C. D. (2013). Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *Am J Hum Genet, 93*(5), 852-864. doi:10.1016/j.ajhg.2013.10.002

Catanach, A., Crowhurst, R., Deng, C., David, C., Bernatchez, L., & Wellenreuther, M. (2019). The genomic pool of standing structural variation outnumbers single nucleotide polymorphism by threefold in the marine teleost Chrysophrys auratus. *Mol Ecol, 28*(6), 1210-1223. doi:10.1111/mec.15051

Charif, D., & Lobry, J. R. (2007). SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In *Structural approaches to sequence evolution* (pp. 207-232): Springer.

Charlesworth, B., Charlesworth, D., & Barton, N. H. (2003). The effects of genetic and geographic structure on neutral variation. *Annual review of ecology, evolution, and systematics, 34*(1), 99-125.

Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Research, 70*(2), 155-174.

Charlesworth, D. (2006). Balancing Selection and Its Effects on Sequences in Nearby Genome Regions. *Plos Genetics, 2*(4), e64. doi:10.1371/journal.pgen.0020064

Cheng, S. H., Gold, M., Rodriguez, N., & Barber, P. H. (2021). Genome-wide SNPs reveal complex fine scale population structure in the California market squid fishery (Doryteuthis opalescens). *Conservation genetics, 22*(1), 97-110. doi:10.1007/s10592-020-01321-2

Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., . . . McCabe, A. M. (2009). The last glacial maximum. *science, 325*(5941), 710-714.

Clucas, G. V., Lou, R. N., Overgaard Therkildsen, N., & Kovach, A. I. (2019). Novel signals of adaptive genetic variation in northwestern Atlantic cod revealed by whole genome sequencing. *Evolutionary Applications, 0*(ja). doi:10.1111/eva.12861

Conover, D. O., & Baumann, H. (2009). The role of experiments in understanding fishery-induced evolution. *Evolutionary Applications, 2*(3), 276-290.

Conover, D. O., & Munch, S. B. (2002). Sustaining fisheries yields over evolutionary time scales. *science, 297*(5578), 94-96.

Cooper, A., & Poinar, H. N. (2000). Ancient DNA: do it right or not at all. *science, 289*(5482), 1139-1139.

Coscia, I., Wilmes, S. B., Ironside, J. E., Goward-Brown, A., O'Dea, E., Malham, S. K., . . . Robins, P. E. (2020). Fine-scale seascape genomics of an exploited marine species, the common cockle Cerastoderma edule, using a multimodelling approach. *Evolutionary Applications, n/a*(n/a). doi:10.1111/eva.12932

Cossins, A. R., & Crawford, D. L. (2005). Fish as models for environmental genomics. *Nature Reviews Genetics, 6*(4), 324-333.

Costello, C., Ovando, D., Clavelle, T., Strauss, C. K., Hilborn, R., Melnychuk, M. C., . . . Cabral, R. B. (2016). Global fishery prospects under contrasting management regimes. *Proceedings of the National Academy of Sciences, 113*(18), 5125-5129.

Craig, M. T., Eble, J. A., Bowen, B. W., & Robertson, D. R. (2007). High genetic connectivity across the Indian and Pacific Oceans in the reef fish Myripristis berndti (Holocentridae). *Marine Ecology Progress Series, 334*, 245-254.

Dabney, J., Knapp, M., Glocke, I., Gansauge, M. T., Weihmann, A., Nickel, B., . . . Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc Natl Acad Sci U S A, 110*(39), 15758-15763. doi:10.1073/pnas.1314445110

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., . . . Sherry, S. T. (2011). The variant call format and VCFtools. *Bioinformatics, 27*(15), 2156-2158.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature methods, 9*(8), 772-772.

De-Santis, C., & Jerry, D. R. (2007). Candidate growth genes in finfish — Where should we be looking? *Aquaculture, 272*(1), 22-38. doi:https://doi.org/10.1016/j.aquaculture.2007.08.036

De la Cruz, O., & Raska, P. (2014). *Population structure at different minor allele frequency levels.* Paper presented at the BMC proceedings.

Demidov, O., Kek, C., Shreeram, S., Timofeev, O., Fornace, A., Appella, E., & Bulavin, D. (2007). The role of the MKK6/p38 MAPK pathway in Wip1-dependent regulation of ErbB2-driven mammary gland tumorigenesis. *Oncogene, 26*(17), 2502-2506.

Der Sarkissian, C., Allentoft, M. E., Ávila-Arcos, M. C., Barnett, R., Campos, P. F., Cappellini, E., . . . Ginolhac, A. (2015). Ancient genomics. *Phil. Trans. R. Soc. B, 370*(1660), 20130387.

Diaz Pauli, B., & Heino, M. (2014). What can selection experiments teach us about fisheries-induced evolution? *Biological journal of the Linnean Society, 111*(3), 485-503.

Dieckmann, U., & Heino, M. (2007). Probabilistic maturation reaction norms: their history, strengths, and limitations. *Marine Ecology Progress Series, 335*, 253-269. doi:DOI 10.3354/meps335253

Ding, S., Mishra, M., Wu, H., Liang, S., & Miyamoto, M. M. (2018). Characterization of hybridization within a secondary contact region of the inshore fish, Bostrychus sinensis, in the East China Sea. *Heredity, 120*(1), 51-62.

Dlabola, E., Wilson, G., Gorman, A., Riesselman, C., & Moy, C. (2015). A post-glacial relative sea-level curve from Fiordland, New Zealand. *Global and Planetary Change, 131*, 104-114.

Dohner, M., Phillips, N. E., & Ritchie, P. A. (2018). Fine-scale genetic structure across a New Zealand disjunction for the direct-developing intertidal whelk Cominella maculosa (Gastropoda: Buccinidae). *Biological journal of the Linnean Society, 123*(3), 593-602.

Dokmanovic, M., Wu, Y., Shen, Y., Chen, J., Hirsch, D. S., & Wu, W. J. (2015). Abstract B38: Trastuzumab-induced recruitment of Csk-homologous kinase (CHK) to ErbB2 receptor is associated with ErbB2-Y1248 phosphorylation and ErbB2 degradation to mediate cell growth inhibition. In: AACR.

Dolby, G. A. (2021). The hidden landscape: Evidence that sea-level change shaped the present population genomic patterns of marginal marine species. *Molecular Ecology, 30*(6), 1357-1360. doi:https://doi.org/10.1111/mec.15826

Donelson, J., Munday, P., McCormick, M., Pankhurst, N., & Pankhurst, P. (2010). Effects of elevated water temperature and food availability on the reproductive performance of a coral reef fish. *Marine Ecology Progress Series, 401*, 233-243.

Dorant, Y., Cayuela, H., Wellband, K., Laporte, M., Rougemont, Q., Mérot, C., . . . Bernatchez, L. (2020). Copy number variants outperform SNPs to reveal genotype–temperature association in a marine species. *Molecular Ecology, 29*(24), 4765-4782. doi:https://doi.org/10.1111/mec.15565

Doveri, S., Lee, D., Maheswaran, M., & Powell, W. (2008). Molecular markers-history, features and applications. *Principles and practices of plant genomics, Volume 1: genome mapping*, 23-67.

Duncan, K., Martin, A., Bowen, B., & De Couet, H. (2006). Global phylogeography of the scalloped hammerhead shark (Sphyrna lewini). *Molecular Ecology, 15*(8), 2239-2251.

Ellegren, H. (2008). Comparative genomics and the study of evolution by natural selection. *Molecular Ecology, 17*(21), 4586-4596.

Ellegren, H., & Galtier, N. (2016). Determinants of genetic diversity. *Nature Reviews Genetics, 17*(7), 422-433.

Elson, J. L., & Lightowlers, R. N. (2006). Mitochondrial DNA clonality in the dock: can surveillance swing the case? *Trends in Genetics, 22*(11), 603-607. doi:https://doi.org/10.1016/j.tig.2006.09.004

Enberg, K., Jørgensen, C., Dunlop, E. S., Varpe, Ø., Boukal, D. S., Baulier, L., . . . Heino, M. (2012). Fishing-induced evolution of growth: concepts, mechanisms and the empirical evidence. *Marine Ecology, 33*(1), 1-25.

Euclide, P. T., MacDougall, T., Robinson, J. M., Faust, M. D., Wilson, C. C., Chen, K. Y., . . . Ludsin, S. (2021). Mixed-stock analysis using Rapture genotyping to evaluate stock-specific exploitation of a walleye population despite weak genetic structure. *Evolutionary Applications, n/a*(n/a). doi:10.1111/eva.13209

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics, 32*(19), 3047-3048.

Fan, G., Song, Y., Yang, L., Huang, X., Zhang, S., Zhang, M., . . . He, S. (2020). Initial data release and announcement of the 10,000 Fish Genomes Project (Fish10K). *GigaScience, 9*(8). doi:10.1093/gigascience/giaa080

FAO. (2020). The State of World Fisheries and Aquaculture 2020.

Fauvelot, C., Bernardi, G., & Planes, S. (2003). Reductions in the mitochondrial DNA diversity of coral reef fish provide evidence of population bottlenecks resulting from Holocene sea-level change. *Evolution, 57*(7), 1571-1583.

Ferrari, G., Cuevas, A., Gondek-Wyrozemska, A. T., Ballantyne, R., Kersten, O., Pálsdóttir, A. H., . . . Star, B. (2021). The preservation of ancient DNA in archaeological fish bone. *Journal of Archaeological Science, 126*, 105317. doi:https://doi.org/10.1016/j.jas.2020.105317

Fisheries New Zealand. (2018). Fisheries Assessment Plenary, May 2018: stock assessments and stock status. Compiled by the Fisheries Science Group, Fisheries New Zealand, Wellington, New Zealand. 1 660p.

Fleming, A. M., Dohner, M. M., Phillips, N. E., & Ritchie, P. A. (2018). Genetic connectivity among populations of two congeneric direct-developing whelks varies across spatial scales. *New Zealand Journal of Marine and Freshwater Research, 52*(1), 100-117.

Foote, A. D., & Morin, P. A. (2015). Sympatric speciation in killer whales? *Heredity (Edinb), 114*(6), 537-538. doi:10.1038/hdy.2014.120

Foote, A. D., & Morin, P. A. (2016). Genome-wide SNP data suggest complex ancestry of sympatric North Pacific killer whale ecotypes. *Heredity (Edinb), 117*(5), 316-325. doi:10.1038/hdy.2016.54

Ford, J. K., Ellis, G. M., & Balcomb, K. C. (1996). *Killer whales: the natural history and genealogy of Orcinus orca in British Columbia and Washington*: UBC press.

Ford, J. K., Ellis, G. M., Barrett-Lennard, L. G., Morton, A. B., Palm, R. S., & Balcomb III, K. C. (1998). Dietary specialization in two sympatric populations of killer whales (Orcinus orca) in coastal British Columbia and adjacent waters. *Canadian Journal of Zoology, 76*(8), 1456-1471.

Fraser, C. I., Nikula, R., Spencer, H. G., & Waters, J. M. (2009). Kelp genes reveal effects of subantarctic sea ice during the Last Glacial Maximum. *Proceedings of the National Academy of Sciences, 106*(9), 3249-3253. doi:10.1073/pnas.0810635106

Furin, C. G., von Hippel, F. A., Postlethwait, J. H., Buck, C. L., Cresko, W. A., & O'Hara, T. M. (2015). Developmental timing of sodium perchlorate exposure alters angiogenesis, thyroid follicle proliferation and sexual maturation in stickleback. *General and Comparative Endocrinology, 219*, 24-35. doi:https://doi.org/10.1016/j.ygcen.2015.04.002

Gaither, M. R., Bowen, B. W., Bordenave, T.-R., Rocha, L. A., Newman, S. J., Gomez, J. A., . . . Craig, M. T. (2011). Phylogeography of the reef fish Cephalopholis argus (Epinephelidae) indicates Pleistocene isolation across the Indo-Pacific Barrier with contemporary overlap in the Coral Triangle. *BMC evolutionary biology, 11*(1), 1-16.

Gartside, D., & Kirkegaard, I. (2010). A history of fishing. *Interactions: Food, Agriculture And Environment, 2*, 70-80.

Garud, N. R., Messer, P. W., Buzbas, E. O., & Petrov, D. A. (2015). Recent selective sweeps in North American Drosophila melanogaster show signatures of soft sweeps. *Plos Genetics, 11*(2), e1005004.

Gautier, M., Klassmann, A., & Vitalis, R. (2017). rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Molecular ecology resources, 17*(1), 78-90.

Gillooly, J. F., Allen, A. P., West, G. B., & Brown, J. H. (2005). The rate of DNA evolution: Effects of body size and temperature on the molecular clock. *Proceedings of the National Academy of Sciences of the United States of America, 102*(1), 140-145. doi:10.1073/pnas.0407735101

Graham, C. F., Glenn, T. C., McArthur, A. G., Boreham, D. R., Kieran, T., Lance, S., . . . Somers, C. M. (2015). Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Mol Ecol Resour, 15*(6), 1304-1315. doi:10.1111/1755-0998.12404

Grant, W. S., Liu, M., Gao, T., & Yanagimoto, T. (2012). Limits of Bayesian skyline plot analysis of mtDNA sequences to infer historical demographies in Pacific herring (and other species). *Mol Phylogenet Evol, 65*(1), 203-212. doi:10.1016/j.ympev.2012.06.006

Grant, W. S., & Waples, R. S. (2000). Spatial and Temporal Scales of Genetic Variability in Marine and Anadromous Species: Implications for Fisheries Oceanography. In P. J. Harrison & T. R. Passons (Eds.), *Fisheries Oceanography An Integrative Approach to Fisheries Ecology and Management*. 2000.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., . . . Fritz, M. H.-Y. (2010). A draft sequence of the Neandertal genome. *science, 328*(5979), 710-722.

Grummer, J. A., Beheregaray, L. B., Bernatchez, L., Hand, B. K., Luikart, G., Narum, S. R., & Taylor, E. B. (2019). Aquatic Landscape Genomics and Environmental Effects on Genetic Variation. *Trends Ecol Evol, 34*(7), 641-654. doi:10.1016/j.tree.2019.02.013

Gu, S., Yi, M.-R., He, X.-B., Lin, P.-S., Liu, W.-H., Luo, Z.-S., . . . Yan, Y.-R. (2021). Genetic diversity and population structure of cutlassfish (Lepturacanthus savala) along the coast of mainland China, as inferred by mitochondrial and microsatellite DNA markers. *Regional Studies in Marine Science, 43*, 101702. doi:https://doi.org/10.1016/j.rsma.2021.101702

Guo, L., Costanzo-Garvey, D. L., Smith, D. R., Neilsen, B. K., MacDonald, R. G., & Lewis, R. E. (2017). Kinase Suppressor of Ras 2 (KSR2) expression in the brain regulates energy balance and glucose homeostasis. *Molecular Metabolism, 6*(2), 194-205. doi:https://doi.org/10.1016/j.molmet.2016.12.004

Gurgul, A., Jasielczuk, I., Semik-Gurgul, E., Pawlina-Tyszko, K., Stefaniuk-Szmukier, M., Szmatoła, T., . . . Bugno-Poniewierska, M. (2019). A genome-wide scan for diversifying selection signatures in selected horse breeds. *PloS one, 14*(1), e0210751.

Halpern, B. S., Frazier, M., Potapenko, J., Casey, K. S., Koenig, K., Longo, C., . . . Walbridge, S. (2015). Spatial and temporal changes in cumulative human impacts on the world's ocean. *Nature Communications, 6*(1), 7615. doi:10.1038/ncomms8615

Han, L., & Abney, M. (2013). Using identity by descent estimation with dense genotype data to detect positive selection. *European Journal of Human Genetics, 21*(2), 205-211.

Hansen, K. D., Brenner, S. E., & Dudoit, S. (2010). Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic acids research, 38*(12), e131-e131. doi:10.1093/nar/gkq224

Harris, A. M., Garud, N. R., & DeGiorgio, M. (2018). Detection and classification of hard and soft sweeps from unphased genotypes by multilocus genotype identity. *Genetics, 210*(4), 1429-1452.

Harrison, R. G. (1989). Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology & Evolution, 4*(1), 6-11.

Hauser, L., Adcock, G. J., Smith, P. J., Ramirez, J. H., & Carvalho, G. R. (2002). Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (Pagrus auratus). *Proc Natl Acad Sci U S A, 99*(18), 11742-11747. doi:10.1073/pnas.172242899

Heino, M., Diaz Pauli, B., & Dieckmann, U. (2015). Fisheries-Induced Evolution. *Annual Review of Ecology, Evolution, and Systematics, Vol 46, 46*, 461-+. doi:10.1146/annurev-ecolsys-120213-054339

Heino, M., Dieckmann, U., & Godo, O. R. (2002). Measuring probabilistic reaction norms for age and size at maturation. *Evolution, 56*(4), 669-678. doi:DOI 10.1111/j.0014-3820.2002.tb01378.x

Hess, J., Matala, A., & Narum, S. (2011). Comparison of SNPs and microsatellites for fine-scale application of genetic stock identification of Chinook salmon in the Columbia River Basin. *Molecular ecology resources, 11*, 137-149.

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature, 405*(6789), 907-913.

Hey, J., Won, Y. J., Sivasundar, A., Nielsen, R., & Markert, J. A. (2004). Using nuclear haplotypes with microsatellites to study gene flow between recently separated Cichlid species. *Molecular Ecology, 13*(4), 909-919.

Hilborn, R. (2012). The evolution of quantitative marine fisheries management 1985–2010. *Natural Resource Modeling, 25*(1), 122-144.

Hilborn, R., Amoroso, R. O., Anderson, C. M., Baum, J. K., Branch, T. A., Costello, C., . . . Ye, Y. (2020). Effective fisheries management instrumental in improving fish stock status. *Proceedings of the National Academy of Sciences, 117*(4), 2218-2224. doi:10.1073/pnas.1909726116

Hilborn, R., & Ovando, D. (2014). Reflections on the success of traditional fisheries management. *ICES Journal of Marine Science, 71*(5), 1040-1046.

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., . . . Whitlock, M. C. (2016). Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist, 188*(4), 379-397.

Hoffmann, A. A., & Sgro, C. M. (2011). Climate change and evolutionary adaptation. *Nature, 470*(7335), 479-485.

Hofreiter, M., Serre, D., Poinar, H. N., Kuch, M., & Pääbo, S. (2001). ancient DNA. *Nature Reviews Genetics, 2*(5), 353-359.

Hofreiter, M., & Shapiro, B. (2012). *Ancient DNA: Methods and Protocols*: Humana Press Incorporated.

Hogner, S., Laskemoen, T., Lifjeld, J. T., Porkert, J., Kleven, O., Albayrak, T., . . . Johnsen, A. (2012). Deep sympatric mitochondrial divergence without reproductive isolation in the common redstart P hoenicurus phoenicurus. *Ecology and Evolution, 2*(12), 2974-2988.

Hohenlohe, P. A., Funk, W. C., & Rajora, O. P. (2020). Population genomics for wildlife conservation and management. *Mol Ecol, n/a*(n/a). doi:10.1111/mec.15720

Hohenlohe, P. A., Phillips, P. C., & Cresko, W. A. (2010). Using population genomics to detect selection in natural populations: key concepts and methodological considerations. *International journal of plant sciences, 171*(9), 1059-1071.

Hutchinson, W. F., van Oosterhout, C., Rogers, S. I., & Carvalho, G. R. (2003). Temporal analysis of archived samples indicates marked genetic changes in declining North Sea cod (Gadus morhua). *Proc Biol Sci, 270*(1529), 2125-2132. doi:10.1098/rspb.2003.2493

Jacobs, A., Carruthers, M., Eckmann, R., Yohannes, E., Adams, C. E., Behrmann-Godel, J., & Elmer, K. R. (2019). Rapid niche expansion by selection on functional genomic variation after ecosystem recovery. *Nature ecology & evolution, 3*(1), 77-86.

Jacobs, A., Carruthers, M., Yurchenko, A., Gordeeva, N. V., Alekseyev, S. S., Hooker, O., . . . Adams, C. E. (2018). Convergence in form and function overcomes non-parallel evolutionary histories in Arctic charr. *bioRxiv*, 265272.

Jenkins, T. L., Castilho, R., & Stevens, J. R. (2018). Meta-analysis of northeast Atlantic marine taxa shows contrasting phylogeographic patterns following post-LGM expansions. *PeerJ, 6*, e5684.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics, 24*(11), 1403-1405.

Jombart, T., Archer, F., Schliep, K., Kamvar, Z., Harris, R., Paradis, E., . . . Lapp, H. (2017). apex: phylogenetics with multiple genes. *Molecular ecology resources, 17*(1), 19-26.

Jones, A. T., Lavery, S. D., Le Port, A., Wang, Y. G., Blower, D., & Ovenden, J. (2019). Sweepstakes reproductive success is absent in a New Zealand snapper (Chrysophrus auratus) population protected from fishing despite "tiny" Ne /N ratios elsewhere. *Mol Ecol, 0*(ja). doi:10.1111/mec.15130

Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P. L., & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics, 29*(13), 1682-1684. doi:10.1093/bioinformatics/btt193

Jørgensen, H. B., Hansen, M. M., Bekkevold, D., Ruzzante, D. E., & Loeschcke, V. (2005). Marine landscapes and population genetic structure of herring (Clupea harengus L.) in the Baltic Sea. *Molecular Ecology, 14*(10), 3219-3234.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ, 2*, e281.

Keighley, X., Bro-Jorgensen, M. H., Ahlgren, H., Szpak, P., Ciucani, M. M., Sanchez Barreiro, F., . . . Olsen, M. T. (2021). Predicting sample success for large-scale ancient DNA studies on marine mammals. *Mol Ecol Resour, n/a*(n/a). doi:10.1111/1755-0998.13331

Kim, B.-M., Amores, A., Kang, S., Ahn, D.-H., Kim, J.-H., Kim, I.-C., . . . Park, H. (2019). Antarctic blackfin icefish genome reveals adaptations to extreme environments. *Nature ecology & evolution, 3*(3), 469-478. doi:10.1038/s41559-019-0812-7

Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature, 217*(5129), 624-626.

Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., . . . Andersen, Ø. (2016). Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology, 25*(10), 2130-2143. doi:https://doi.org/10.1111/mec.13592

Knapp, M., Clarke, A. C., Horsburgh, K. A., & Matisoo-Smith, E. A. (2012). Setting the stage–building and working in an ancient DNA laboratory. *Annals of Anatomy-Anatomischer Anzeiger, 194*(1), 3-6.

Knutsen, H., Olsen, E. M., Jorde, P. E., Espeland, S. H., Andre, C., & Stenseth, N. C. (2011). Are low but statistically significant levels of genetic differentiation in marine fishes 'biologically meaningful'? A case study of coastal Atlantic cod. *Molecular Ecology, 20*(4), 768-783.

Koenigstein, S., Mark, F. C., Gößling-Reisemann, S., Reuter, H., & Poertner, H. O. (2016). Modelling climate change impacts on marine fish populations: process-based integration of ocean warming, acidification and other environmental drivers. *Fish and Fisheries, 17*(4), 972-1004.

Koepfli, K.-P., Paten, B., Scientists, G. K. C. o., & O'Brien, S. J. (2015). The Genome 10K Project: a way forward. *Annu. Rev. Anim. Biosci., 3*(1), 57-111.

Kolde, R., & Kolde, M. R. (2015). Package 'pheatmap'. *R package, 1*(7), 790.

Korf-Klingebiel, M., Reboll, M. R., Klede, S., Brod, T., Pich, A., Polten, F., . . . Wollert, K. C. (2015). Myeloid-derived growth factor (C19orf10) mediates cardiac repair following myocardial infarction. *Nature Medicine, 21*(2), 140-149. doi:10.1038/nm.3778

Koufopanou, V., Reid, D. G., Ridgway, S. A., & Thomas, R. H. (1999). A Molecular Phylogeny of the Patellid Limpets (Gastropoda: Patellidae) and Its Implications for the Origins of Their Antitropical Distribution. *Molecular phylogenetics and evolution, 11*(1), 138-156. doi:https://doi.org/10.1006/mpev.1998.0557

Lande, R. (1976). Natural selection and random genetic drift in phenotypic evolution. *Evolution*, 314-334.

Lande, R., & Barrowclough, G. (1987). Effective population size, genetic variation, and their use in population. *Viable populations for conservation*, 87-123.

Larson, W. A., Seeb, L. W., Everett, M. V., Waples, R. K., Templin, W. D., & Seeb, J. E. (2014). Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook salmon (Oncorhynchus tshawytscha). *Evolutionary Applications, 7*(3), 355-369.

Law, C. S., Rickard, G. J., Mikaloff-Fletcher, S. E., Pinkerton, M. H., Behrens, E., Chiswell, S. M., & Currie, K. (2018). Climate change projections for the surface ocean around New Zealand. *New Zealand Journal of Marine and Freshwater Research, 52*(3), 309-335.

Layton, K. K. S., Dempson, B., Snelgrove, P. V. R., Duffy, S. J., Messmer, A. M., Paterson, I. G., . . . Bradbury, I. R. (2020). Resolving fine-scale population structure and fishery exploitation using sequenced microsatellites in a northern fish. *Evolutionary Applications, 13*(5), 1055-1068. doi:https://doi.org/10.1111/eva.12922

Le Port, A., Montgomery, J., Smith, A., Croucher, A., McLeod, I., & Lavery, S. (2017). Temperate marine protected area provides recruitment subsidies to local fisheries. *Proceedings of the Royal Society B: Biological Sciences, 284*(1865), 20171300.

Leach, F. (1997). *A guide to the identification of fish remains from New Zealand archaeological sites*: Archaeozoology Laboratory, Museum of New Zealand Te Papa Tongarewa.

Leach, F. (2006). Fishing in Pre-European New Zealand. *Archaeofauna*(15).

Leach, F., & Davidson, J. (2000). Pre-European catches of snapper (Pagrus auratus) in northern New Zealand. *Journal of Archaeological Science, 27*(6), 509-522.

Lewin, H. A., Robinson, G. E., Kress, W. J., Baker, W. J., Coddington, J., Crandall, K. A., . . . Gilbert, M. T. P. (2018). Earth BioGenome Project: Sequencing life for the future of life. *Proceedings of the National Academy of Sciences, 115*(17), 4325-4333.

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

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature, 475*(7357), 493-496.

Li, W. (1997). Molecular evolution. *Sunderland, MA: Sinauer Associates*.

Li, W. C., Huang, C. H., Hsieh, Y. T., Chen, T. Y., Cheng, L. H., Chen, C. Y., . . . Chang, K. W. (2020). Regulatory Role of Hexokinase 2 in Modulating Head and Neck Tumorigenesis. *Front Oncol, 10*(176), 176. doi:10.3389/fonc.2020.00176

Lin, R., Du, X., Peng, S., Yang, L., Ma, Y., Gong, Y., & Li, S. (2015). Discovering all transcriptome single-nucleotide polymorphisms and scanning for selection signatures in ducks (Anas platyrhynchos). *Evolutionary Bioinformatics, 11*, EBO. S21545.

Liu, L., Channavajhala, P. L., Rao, V. R., Moutsatsos, I., Wu, L., Zhang, Y., . . . Qiu, Y. (2009). Proteomic characterization of the dynamic KSR-2 interactome, a signaling scaffold complex in MAPK pathway. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 1794*(10), 1485-1495.

Liu, S., & Hansen, M. M. (2017). PSMC (pairwise sequentially Markovian coalescent) analysis of RAD (restriction site associated DNA) sequencing data. *Molecular ecology resources, 17*(4), 631-641. doi:https://doi.org/10.1111/1755-0998.12606

Liu, W., Morito, D., Takashima, S., Mineharu, Y., Kobayashi, H., Hitomi, T., . . . Koizumi, A. (2011). Identification of RNF213 as a Susceptibility Gene for Moyamoya Disease and Its Possible Role in Vascular Development. *PloS one, 6*(7), e22542. doi:10.1371/journal.pone.0022542

Llaurens, V., Whibley, A., & Joron, M. (2017). Genetic architecture and balancing selection: the life and death of differentiated variants. *Molecular Ecology, 26*(9), 2430-2448. doi:https://doi.org/10.1111/mec.14051

Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. *Molecular Ecology, 23*(9), 2178-2192. doi:https://doi.org/10.1111/mec.12725

Lowe, W. H., & Allendorf, F. W. (2010). What can genetics tell us about population connectivity? *Molecular Ecology, 19*(15), 3038-3051.

Lowe, W. H., Kovach, R. P., & Allendorf, F. W. (2017). Population Genetics and Demography Unite Ecology and Evolution. *Trends in Ecology & Evolution*.

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics, 4*(12), 981-994.

Luikart, G., Kardos, M., Hand, B. K., Rajora, O. P., Aitken, S. N., & Hohenlohe, P. A. (2018). Population genomics: advancing understanding of nature. In *Population genomics* (pp. 3-79): Springer.

Mackay, T. F., Stone, E. A., & Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics, 10*(8), 565-577.

Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R. (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature ecology & evolution, 2*(12), 1940-1955. doi:10.1038/s41559-018-0717-x

Manel, S., Gaggiotti, O. E., & Waples, R. S. (2005). Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology & Evolution, 20*(3), 136-142.

Manning, M., Lawrence, J., King, D. N., & Chapman, R. (2015). Dealing with changing risks: a New Zealand perspective on climate change adaptation. *Regional Environmental Change, 15*(4), 581-594.

Mardis, E. R. (2017). DNA sequencing technologies: 2006–2016. *Nature protocols, 12*, 213. doi:10.1038/nprot.2016.182

Marques, D. A., Jones, F. C., Di Palma, F., Kingsley, D. M., & Reimchen, T. E. (2018). Experimental evidence for rapid genomic adaptation to a new niche in an adaptive radiation. *Nature ecology & evolution, 2*(7), 1128-1138. doi:10.1038/s41559-018-0581-8

Marshall, D. C., Hill, K. B., Fontaine, K. M., Buckley, T. R., & Simon, C. (2009). Glacial refugia in a maritime temperate climate: cicada (Kikihia subalpina) mtDNA phylogeography in New Zealand. *Molecular Ecology, 18*(9), 1995-2009.

Marske, K. A., Leschen, R. A., Barker, G. M., & Buckley, T. R. (2009). Phylogeography and ecological niche modelling implicate coastal refugia and trans-alpine dispersal of a New Zealand fungus beetle. *Molecular Ecology, 18*(24), 5126-5142.

Marty, L., Rochet, M.-J., & Ernande, B. (2014). Temporal trends in age and size at maturation of four North Sea gadid species: cod, haddock, whiting and Norway pout. *Marine Ecology Progress Series, 497*, 179-197.

Mather, N., Traves, S. M., & Ho, S. Y. W. (2020). A practical introduction to sequentially Markovian coalescent methods for estimating demographic history from genomic data. *Ecol Evol, 10*(1), 579-589. doi:10.1002/ece3.5888

Matschiner, M. (2016). Fitchi: haplotype genealogy graphs based on the Fitch algorithm. *Bioinformatics, 32*(8), 1250-1252. doi:10.1093/bioinformatics/btv717

McCulloch, G. A., Wallis, G. P., & Waters, J. M. (2010). Onset of glaciation drove simultaneous vicariant isolation of alpine insects in New Zealand. *Evolution: International Journal of Organic Evolution, 64*(7), 2033-2043.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., . . . Daly, M. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research, 20*(9), 1297-1303.

McKeown, N. J., Arkhipkin, A. I., & Shaw, P. W. (2017). Regional genetic population structure and fine scale genetic cohesion in the Southern blue whiting Micromesistius australis. *Fisheries research, 185*, 176-184. doi:https://doi.org/10.1016/j.fishres.2016.09.006

McKinney, G. J., Seeb, J. E., & Seeb, L. W. (2017). Managing mixed-stock fisheries: genotyping multi-SNP haplotypes increases power for genetic stock identification. *Canadian Journal of Fisheries and Aquatic Sciences, 74*(4), 429-434.

Meyer, M., Kircher, M., Gansauge, M. T., Li, H., Racimo, F., Mallick, S., . . . Paabo, S. (2012). A high-coverage genome sequence from an archaic Denisovan individual. *science, 338*(6104), 222-226. doi:10.1126/science.1224344

Ministry for Primary Industries. (2021). Quota Management System. Retrieved from https://www.mpi.govt.nz/fishing-aquaculture/fisheries-management/quota-management-system/

Ministry of Fisheries. (2011). *Operational guidelines for New Zealand's harvest strategy standard*.

Miya, M., Kawaguchi, A., & Nishida, M. (2001). Mitogenomic Exploration of Higher Teleostean Phylogenies: A Case Study for Moderate-Scale Evolutionary Genomics with 38 Newly Determined Complete Mitochondrial DNA Sequences. *Molecular Biology and Evolution, 18*(11), 1993-2009. doi:10.1093/oxfordjournals.molbev.a003741

Moest, M., Van Belleghem, S. M., James, J. E., Salazar, C., Martin, S. H., Barker, S. L., . . . Nadeau, N. J. (2020). Selective sweeps on novel and introgressed variation shape mimicry loci in a butterfly adaptive radiation. *PLoS biology, 18*(2), e3000597.

Momigliano, P., Jokinen, H., Fraimout, A., Florin, A.-B., Norkko, A., & Merilä, J. (2017). Extraordinarily rapid speciation in a marine fish. *Proceedings of the National Academy of Sciences, 114*(23), 6074-6079. doi:10.1073/pnas.1615109114

Moore, J. S., Bourret, V., Dionne, M., Bradbury, I., O'Reilly, P., Kent, M., . . . Bernatchez, L. (2014). Conservation genomics of anadromous Atlantic salmon across its North American range: outlier loci identify the same patterns of population structure as neutral loci. *Mol Ecol, 23*(23), 5680-5697. doi:10.1111/mec.12972

Morin, P. A., Luikart, G., Wayne, R. K., & the, S. N. P. w. g. (2004). SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution, 19*(4), 208-216. doi:https://doi.org/10.1016/j.tree.2004.01.009

Morin, P. A., Martien, K. K., & Taylor, B. L. (2009). Assessing statistical power of SNPs for population structure and conservation studies. *Molecular ecology resources, 9*(1), 66-73.

Moritz, C. (1994a). Applications of mitochondrial DNA analysis in conservation: a critical review. *Molecular Ecology, 3*(4), 401-411.

Moritz, C. (1994b). Defining 'evolutionarily significant units' for conservation. *Trends in Ecology & Evolution, 9*(10), 373-375.

Mueller, R. L. (2006). Evolutionary rates, divergence dates, and the performance of mitochondrial genes in Bayesian phylogenetic analysis. *Systematic Biology, 55*(2), 289-300.

Munday, P. L., Donelson, J. M., & Domingos, J. A. (2017). Potential for adaptation to climate change in a coral reef fish. *Global Change Biology, 23*(1), 307-317.

Myers, R. A., Hutchings, J. A., & Barrowman, N. J. (1997). Why do Fish Stocks Collapse? The Example of Cod in Atlantic Canada. *Ecological Applications, 7*(1), 91-106. doi:10.2307/2269409

Nam, B.-H., Yoo, D., Kim, Y.-O., Park, J. Y., Shin, Y., Shin, G.-h., . . . Kwak, W. (2019). Whole genome sequencing reveals the impact of recent artificial selection on red sea bream reared in fish farms. *Scientific Reports, 9*(1), 6487. doi:10.1038/s41598-019-42988-z

Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution, 32*(1), 268-274.

Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics, 23*(1), 263-286.

Oleksiak, M. (2010). Genomic approaches with natural fish populations. *Journal of Fish Biology, 76*(5), 1067-1093.

Olsen, E. M., Heino, M., Lilly, G. R., Morgan, M. J., Brattey, J., Ernande, B., & Dieckmann, U. (2004). Maturation trends indicative of rapid evolution preceded the collapse of northern cod. *Nature, 428*(6986), 932-935.

Oosting, T., Hilario, E., Wellenreuther, M., & Ritchie, P. A. (2020). DNA degradation in fish: Practical solutions and guidelines to improve DNA preservation for genomic research. *Ecol Evol, 10*(16), 8643-8651. doi:10.1002/ece3.6558

Oosting, T., Star, B., Barrett, J. H., Wellenreuther, M., Ritchie, P. A., & Rawlence, N. J. (2019). Unlocking the potential of ancient fish DNA in the genomic era. *Evol Appl, 12*(8), 1513-1522. doi:10.1111/eva.12811

Osterberg, J. S., Cammen, K. M., Schultz, T. F., Clark, B. W., & Di Giulio, R. T. (2018). Genome-wide scan reveals signatures of selection related to pollution adaptation in non-model estuarine Atlantic killifish (Fundulus heteroclitus). *Aquatic Toxicology, 200*, 73-82. doi:https://doi.org/10.1016/j.aquatox.2018.04.017

Ovenden, J. R., Berry, O., Welch, D. J., Buckworth, R. C., & Dichmont, C. M. (2015). Ocean's eleven: a critical evaluation of the role of population, evolutionary and molecular genetics in the management of wild fisheries. *Fish and Fisheries, 16*(1), 125-159. doi:10.1111/faf.12052

Pahnke, K., Zahn, R., Elderfield, H., & Schulz, M. (2003). 340,000-Year Centennial-Scale Marine Record of Southern Hemisphere Climatic Oscillation. *science, 301*(5635), 948-952. doi:10.1126/science.1084451

Palsbøll, P. J., Berube, M., & Allendorf, F. W. (2007). Identification of management units using population genetic data. *Trends in Ecology & Evolution, 22*(1), 11-16.

Papa, Y., Halliwell, A. G., Morrison, M. A., Wellenreuther, M., & Ritchie, P. A. (2021). Phylogeographic structure and historical demography of tarakihi (Nemadactylus macropterus) and king tarakihi (Nemadactylus n.sp.) in New Zealand. *New Zealand Journal of Marine and Freshwater Research*, 1-25. doi:10.1080/00288330.2021.1912119

Papa, Y., Oosting, T., Valenza-Troubat, N., Wellenreuther, M., & Ritchie, P. A. (2020). Genetic stock structure of New Zealand fish and the use of genomics in fisheries management: an overview and outlook. *New Zealand Journal of Zoology*, 1-31. doi:10.1080/03014223.2020.1788612

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics, 26*(3), 419-420.

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics, 35*(3), 526-528.

Parsons, D., Sim-Smith, C., Cryer, M., Francis, M., Hartill, B., Jones, E., . . . Morrison, M. (2014). Snapper (Chrysophrys auratus): a review of life history and key vulnerabilities in New Zealand. *New Zealand Journal of Marine and Freshwater Research, 48*(2), 256-283.

Parsons, D. M., Morrison, M. A., MacDiarmid, A. B., Stirling, B., Cleaver, P., Smith, I. W., & Butcher, M. (2009). Risks of shifting baselines highlighted by anecdotal accounts of New Zealand's snapper (Pagrus auratus) fishery. *New Zealand Journal of Marine and Freshwater Research, 43*(4), 965-983.

Pastene, L. A., Goto, M., Kanda, N., Zerbini, A. N., Kerem, D., Watanabe, K., . . . Larsen, F. (2007). Radiation and speciation of pelagic organisms during periods of global warming: the case of the common minke whale, Balaenoptera acutorostrata. *Molecular Ecology, 16*(7), 1481-1495.

Pauly, D., Christensen, V., Guénette, S., Pitcher, T. J., Sumaila, U. R., Walters, C. J., . . . Zeller, D. (2002). Towards sustainability in world fisheries. *Nature, 418*(6898), 689.

Peltier, W., & Fairbanks, R. G. (2006). Global glacial ice volume and Last Glacial Maximum duration from an extended Barbados sea level record. *Quaternary Science Reviews, 25*(23-24), 3322-3337.

Peñalba, J. V., & Wolf, J. B. (2020). From molecules to populations: appreciating and estimating recombination rate variation. *Nature Reviews Genetics, 21*(8), 476-492.

Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R., . . . Misra, R. (2007). Diet and the evolution of human amylase gene copy number variation. *Nature genetics, 39*(10), 1256-1260.

Petrou, E. L., Fuentes-Pardo, A. P., Rogers, L. A., Orobko, M., Tarpey, C., Jiménez-Hidalgo, I., . . . Hauser, L. (2021). Functional genetic diversity in an exploited marine species and its relevance to fisheries management. *Proceedings of the Royal Society B: Biological Sciences, 288*(1945), 20202398. doi:doi:10.1098/rspb.2020.2398

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution, 31*(7), 1929-1936.

Pine, W. E., Pollock, K. H., Hightower, J. E., Kwak, T. J., & Rice, J. A. (2003). A review of tagging methods for estimating fish population size and components of mortality. *Fisheries, 28*(10), 10-23.

Pinsky, M. L., Eikeset, A. M., Helmerson, C., Bradbury, I. R., Bentzen, P., Morris, C., . . . Star, B. (2021). Genomic stability through time despite decades of exploitation in cod on both sides of the

Atlantic. *Proceedings of the National Academy of Sciences, 118*(15), e2025453118. doi:10.1073/pnas.2025453118

Poplin, R., Ruano-Rubio, V., DePristo, M. A., Fennell, T. J., Carneiro, M. O., Van der Auwera, G. A., . . . Banks, E. (2018). Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, 201178. doi:10.1101/201178

Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., . . . De Filippo, C. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature, 505*(7481), 43-49.

Prunier, J., Giguère, I., Ryan, N., Guy, R., Soolanayakanahally, R., Isabel, N., . . . Porth, I. (2019). Gene copy number variations involved in balsam poplar (Populus balsamifera L.) adaptive variations. *Molecular Ecology, 28*(6), 1476-1490.

Punt, A. E., Butterworth, D. S., de Moor, C. L., De Oliveira, J. A., & Haddon, M. (2016). Management strategy evaluation: best practices. *Fish and Fisheries, 17*(2), 303-334.

Qiu, F., Li, H., Lin, H., Ding, S., & Miyamoto, M. M. (2016). Phylogeography of the inshore fish, Bostrychus sinensis, along the Pacific coastline of China. *Molecular phylogenetics and evolution, 96*, 112-117. doi:https://doi.org/10.1016/j.ympev.2015.11.020

R Core Team. (2013). R: A language and environment for statistical computing.

Railsback, L. B., Gibbard, P. L., Head, M. J., Voarintsoa, N. R. G., & Toucanne, S. (2015). An optimized scheme of lettered marine isotope substages for the last 1.0 million years, and the climatostratigraphic nature of isotope stages and substages. *Quaternary Science Reviews, 111*, 94-106.

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics, 197*(2), 573-589.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology, 67*(5), 901.

Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., . . . Ward, R. (2001). Linkage disequilibrium in the human genome. *Nature, 411*(6834), 199-204.

Reid, N. M., Proestou, D. A., Clark, B. W., Warren, W. C., Colbourne, J. K., Shaw, J. R., . . . Whitehead, A. (2016). The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *science, 354*(6317), 1305-1308. doi:10.1126/science.aah4993

Rizzi, E., Lari, M., Gigli, E., De Bellis, G., & Caramelli, D. (2012). Ancient DNA studies: new perspectives on old samples. *Genetics Selection Evolution, 44*(1), 1-19.

Roure, B., Baurain, D., & Philippe, H. (2013). Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Mol Biol Evol, 30*(1), 197-214. doi:10.1093/molbev/mss208

Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., . . . Lander, E. (2006). Positive natural selection in the human lineage. *science, 312*(5780), 1614-1620.

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., . . . The, S. N. P. C. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature, 449*(7164), 913-918. doi:10.1038/nature06250

Sahrhage, D., & Lundbeck, J. (2012). *A history of fishing*: Springer Science & Business Media.

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., & Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *science, 230*(4732), 1350-1354.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, 74*(12), 5463-5467.

Sato, Y., Miya, M., Fukunaga, T., Sado, T., & Iwasaki, W. (2018). MitoFish and MiFish Pipeline: A Mitochondrial Genome Database of Fish with an Analysis Pipeline for Environmental DNA Metabarcoding. *Molecular Biology and Evolution, 35*(6), 1553-1555. doi:10.1093/molbev/msy074

Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature genetics, 46*(8), 919-925.

Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology & Evolution, 16*(7), 372-380. doi:https://doi.org/10.1016/S0169-5347(01)02198-X

Schroeder, G. K., Lad, C., Wyman, P., Williams, N. H., & Wolfenden, R. (2006). The time required for water attack at the phosphorus atom of simple phosphodiesters and of DNA. *Proc Natl Acad Sci U S A, 103*(11), 4052-4055. doi:10.1073/pnas.0510879103

Schubert, M., Ermini, L., Der Sarkissian, C., Jónsson, H., Ginolhac, A., Schaefer, R., . . . McCue, M. (2014). Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nature protocols, 9*(5), 1056.

Schwartz, M. K., Luikart, G., & Waples, R. S. (2007). Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution, 22*(1), 25-33.

Scott, S., & Pankhurst, N. (1992). Interannual variation in the reproductive cycle of the New Zealand snapper Pagrus auratus (Bloch & Schneider)(Sparidae). *Journal of Fish Biology, 41*(5), 685-696.

Shafer, A. B., Cullingham, C. I., Cote, S. D., & Coltman, D. W. (2010). Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. *Molecular Ecology, 19*(21), 4589-4621.

Shahi, P. K., & Pineda, I. F. (2008). Tumoral angiogenesis: review of the literature. *Cancer investigation, 26*(1), 104-108.

Shin, G. H., Shin, Y., Jung, M., Hong, J. M., Lee, S., Subramaniyam, S., . . . Park, C. I. (2018). First Draft Genome for Red Sea Bream of Family Sparidae. *Front Genet, 9*(643), 643. doi:10.3389/fgene.2018.00643

Siddall, M., Rohling, E. J., Almogi-Labin, A., Hemleben, C., Meischner, D., Schmelzer, I., & Smeed, D. (2003). Sea-level fluctuations during the last glacial cycle. *Nature, 423*(6942), 853-858.

Sims, D. W., & Southward, A. J. (2006). Dwindling fish numbers already of concern in 1883. *Nature, 439*(7077), 660-660.

Skog, A., Zachos, F., Rueness, E., Feulner, P., Mysterud, A., Langvatn, R., . . . Hartl, G. (2009). Phylogeography of red deer (Cervus elaphus) in Europe. *Journal of Biogeography, 36*(1), 66-77.

Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *science, 236*(4803), 787-792.

Smith, I., & James-Lee, T. (2010). *Data for an Archaeozoological Analysis of Marine Resource Use in Two New Zealand Study Areas*. Retrieved from http://hdl.handle.net/10523/5945

Smith, P. J. (1979). Esterase gene frequencies and temperature relationships in the New Zealand snapper Chrysophrys auratus. *Marine Biology, 53*(4), 305-310.

Smith, P. J., Francis, R., & Paul, L. (1978). Genetic variation and population structure in the New Zealand snapper. *New Zealand Journal of Marine and Freshwater Research, 12*(4), 343-350.

Speller, C. F., Hauser, L., Lepofsky, D., Moore, J., Rodrigues, A. T., Moss, M. L., . . . Yang, D. Y. (2012). High potential for using DNA from ancient herring bones to inform modern fisheries management and conservation. *PloS one, 7*(11), e51122. doi:10.1371/journal.pone.0051122

Star, B., Boessenkool, S., Gondek, A. T., Nikulina, E. A., Hufthammer, A. K., Pampoulie, C., . . . Barrett, J. H. (2017). Ancient DNA reveals the Arctic origin of Viking Age cod from Haithabu, Germany. *Proc Natl Acad Sci U S A, 114*(34), 9152-9157. doi:10.1073/pnas.1710186114

Stein, L. (2001). Genome annotation: from sequence to biology. *Nature Reviews Genetics, 2*(7), 493-503.

Stiller, J., da Fonseca, R. R., Alfaro, M. E., Faircloth, B. C., Wilson, N. G., & Rouse, G. W. (2020). Using ultraconserved elements to track the influence of sea-level change on leafy seadragon populations. *Molecular Ecology*.

Sun, D., Ge, Y., & Cheng, Q. (2019). Genetic diversity of eight wild populations of Pampus argenteus along the coast of China inferred from fifteen polymorphic microsatellite markers. *Brazilian Journal of Oceanography, 67*.

Sun, Y., Liu, W.-Z., Liu, T., Feng, X., Yang, N., & Zhou, H.-F. (2015). Signaling pathway of MAPK/ERK in cell proliferation, differentiation, migration, senescence and apoptosis. *Journal of Receptors and Signal Transduction, 35*(6), 600-604. doi:10.3109/10799893.2015.1030412

Swain, D. P., Sinclair, A. F., & Mark Hanson, J. (2007). Evolutionary response to size-selective mortality in an exploited fish population. *Proceedings of the Royal Society B: Biological Sciences, 274*(1613), 1015-1022. doi:10.1098/rspb.2006.0275

Sylvester, E. V. A., Bentzen, P., Bradbury, I. R., Clément, M., Pearce, J., Horne, J., & Beiko, R. G. (2018). Applications of random forest feature selection for fine-scale genetic population assignment. *Evolutionary Applications, 11*(2), 153-165. doi:https://doi.org/10.1111/eva.12524

Tabata, K., & Taniguchi, N. (2000). Differences between Pagrus major and Pagrus auratus through mainly mtDNA control region analysis. *Fisheries Science, 66*(1), 9-18.

Terhorst, J., Kamm, J. A., & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature genetics, 49*(2), 303.

Teske, P. R., Von der Heyden, S., McQuaid, C. D., & Barker, N. P. (2011). A review of marine phylogeography in southern Africa. *South African Journal of Science, 107*(5-6), 43-53.

Therkildsen, N. O., Nielsen, E. E., Swain, D. P., & Pedersen, J. S. (2010). Large effective population size and temporal genetic stability in Atlantic cod (Gadus morhua) in the southern Gulf of St. Lawrence. *Canadian Journal of Fisheries and Aquatic Sciences, 67*(10), 1585-1595.

Therkildsen, N. O., Wilder, A. P., Conover, D. O., Munch, S. B., Baumann, H., & Palumbi, S. R. (2019). Contrasting genomic shifts underlie parallel phenotypic evolution in response to fishing. *science, 365*(6452), 487-490. doi:10.1126/science.aaw7271

Thioulouse, J., Dray, S., Dufour, A.-B., Siberchicot, A., Jombart, T., & Pavoine, S. (2018). *Multivariate analysis of ecological data with ade4*: Springer.

Thorstad, E. B., Rikardsen, A. H., Alp, A., & Økland, F. (2013). The use of electronic tags in fish research–an overview of fish telemetry methods. *Turkish Journal of Fisheries and Aquatic Sciences, 13*(5), 881-896.

Tørresen, O. K., Brieuc, M. S., Solbakken, M. H., Sørhus, E., Nederbragt, A. J., Jakobsen, K. S., . . . Jentoft, S. (2018). Genomic architecture of haddock (Melanogrammus aeglefinus) shows expansions of innate immune genes and short tandem repeats. *BMC genomics, 19*(1), 240.

Trippel, E. A. (1995). Age at maturity as a stress indicator in fisheries. *BioScience, 45*(11), 759-771.

Uusi-Heikkilä, S., Sävilammi, T., Leder, E., Arlinghaus, R., & Primmer, C. R. (2017). Rapid, broad-scale gene expression evolution in experimentally harvested fish populations. *Molecular Ecology, 26*(15), 3954-3967.

Uusi-Heikkilä, S., Whiteley, A. R., Kuparinen, A., Matsumura, S., Venturelli, P. A., Wolter, C., . . . Killen, S. S. (2015). The evolutionary legacy of size-selective harvesting extends from genes to populations. *Evolutionary Applications, 8*(6), 597-620.

Väli, Ü., Einarsson, A., Waits, L., & Ellegren, H. (2008). To what extent do microsatellite markers reflect genome-wide genetic diversity in natural populations? *Molecular Ecology, 17*(17), 3808-3817.

van der Valk, T., Pečnerová, P., Díez-del-Molino, D., Bergström, A., Oppenheimer, J., Hartmann, S., . . . Dalén, L. (2021). Million-year-old DNA sheds light on the genomic history of mammoths. *Nature, 591*(7849), 265-269. doi:10.1038/s41586-021-03224-9

van Etten, J. (2017). R package gdistance: distances and routes on geographical grids. *Journal of Statistical Software, 76*(1), 1-21.

van Wijk, S. J., Taylor, M. I., Creer, S., Dreyer, C., Rodrigues, F. M., Ramnarine, I. W., . . . Carvalho, G. R. (2013). Experimental harvesting of fish populations drives genetically based shifts in body size and maturation. *Frontiers in Ecology and the Environment, 11*(4), 181-187.

Vaux, F., Bohn, S., Hyde, J. R., & O'Malley, K. G. (2021). Adaptive markers distinguish North and South Pacific Albacore amid low population differentiation. *Evolutionary Applications, n/a*(n/a). doi:10.1111/eva.13202

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., . . . Holt, R. A. (2001). The sequence of the human genome. *science, 291*(5507), 1304-1351.

Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting Natural Selection in Genomic Data. *Annual Review of Genetics, 47*(1), 97-120. doi:10.1146/annurev-genet-111212-133526

Voight, B. F., Kudaravalli, S., Wen, X., & Pritchard, J. K. (2006). A map of recent positive selection in the human genome. *PLoS biology, 4*(3), e72.

Walker, M., Johnsen, S., Rasmussen, S. O., Popp, T., Steffensen, J. P., Gibbard, P., . . . Björck, S. (2009). Formal definition and dating of the GSSP (Global Stratotype Section and Point) for the base of the Holocene using the Greenland NGRIP ice core, and selected auxiliary records. *Journal of Quaternary Science: Published for the Quaternary Research Association, 24*(1), 3-17.

Wallis, G. P. (2019). Thirty years of conservation genetics in New Zealand: what have we learnt? *Journal of the Royal Society of New Zealand, 49*(3), 320-346.

Wallis, G. P., & Trewick, S. A. (2009). New Zealand phylogeography: evolution on a small continent. *Molecular Ecology, 18*(17), 3548-3580.

Walsh, C. (2003). Snapper catch-at-length and catch-at-age heterogeneity between strata in East Northland longline landings. *New Zealand Fisheries Assessment Report, 200311*, 11.

Walsh, C., & Davies, N. (2004). Length and age composition of commercial snapper landings in SNA 8, 2003–04. *New Zealand Fisheries Assessment Report, 56*, 18.

Walsh, C., McKenzie, J., Buckthought, D., Armiger, H., Ferguson, H., Smith, M., . . . Miller, A. (2011). Age composition of commercial snapper landings in SNA 1, 2009–10. *New Zealand Fisheries Assessment Report, 54*(53), 28.

Walsh, C., McKenzie, J. R., Bian, R., Armiger, H., Buckthought, D., Smith, M. D., . . . Miller, A. (2012). *Snapper Catch-at-length and Catch-at-age Heterogeneity Between Spatial Strata in SNA 2 Bottom Trawl Landings, 2007-08 and 2008-09*: Ministry for Primary Industries.

Walsh, M. R., Munch, S. B., Chiba, S., & Conover, D. O. (2006). Maladaptive changes in multiple traits caused by fishing: impediments to population recovery. *Ecology letters, 9*(2), 142-148.

Wang, K., Shen, Y., Yang, Y., Gan, X., Liu, G., Hu, K., . . . He, S. (2019). Morphology and genome of a snailfish from the Mariana Trench provide insights into deep-sea adaptation. *Nature ecology & evolution, 3*(5), 823-833. doi:10.1038/s41559-019-0864-8

Wang, X., Que, P., Heckel, G., Hu, J., Zhang, X., Chiang, C.-Y., . . . Liu, Y. (2019). Genetic, phenotypic and ecological differentiation suggests incipient speciation in two Charadrius plovers along the Chinese coast. *BMC evolutionary biology, 19*(1), 135. doi:10.1186/s12862-019-1449-5

Wang, Z.-D., Liao, J., Huang, C.-Q., Long, S.-S., Zhang, S., Guo, Y.-S., . . . Liu, C.-W. (2018). Significant genetic differentiation of Gobiopterus lacustris, a newly recorded transparent goby in China. *Mitochondrial DNA Part A, 29*(5), 785-791.

Waples, R. S. (1998). Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *Journal of Heredity, 89*(5), 438-450.

Waples, R. S. (2016). Tiny estimates of the N e/N ratio in marine fishes: Are they real? *Journal of Fish Biology, 89*(6), 2479-2504.

Waples, R. S., & Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology, 15*(6), 1419-1439.

Waples, R. S., Punt, A. E., & Cope, J. M. (2008). Integrating genetic data into management of marine resources: how can we do it better? *Fish and Fisheries, 9*(4), 423-449. doi:https://doi.org/10.1111/j.1467-2979.2008.00303.x

Waples, R. S., & Yokota, M. (2007). Temporal Estimates of Effective Population Size in Species With Overlapping Generations. *Genetics, 175*(1), 219-233. doi:10.1534/genetics.106.065300

Ward, R. D. (2000). Genetics in fisheries management. *Hydrobiologia, 420*(1), 191-201. doi:Doi 10.1023/A:1003928327503

Waters, J. M., Craw, D., Youngson, J. H., & Wallis, G. P. (2001). Genes meet geology: fish phylogeographic pattern reflects ancient, rather than modern, drainage connections. *Evolution, 55*(9), 1844-1851.

Webb, W. C., Marzluff, J. M., & Omland, K. E. (2011). Random interbreeding between cryptic lineages of the Common Raven: evidence for speciation in reverse. *Molecular Ecology, 20*(11), 2390-2402.

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

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 1358-1370.

Weir, J. T., Haddrath, O., Robertson, H. A., Colbourne, R. M., & Baker, A. J. (2016). Explosive ice age diversification of kiwi. *Proceedings of the National Academy of Sciences, 113*(38), E5580-E5587.

Weissman, D. B., & Hallatschek, O. (2017). Minimal-assumption inference from population-genomic data. *Elife, 6*, e24836.

Wellen, K. E., Lu, C., Mancuso, A., Lemons, J. M., Ryczko, M., Dennis, J. W., . . . Thompson, C. B. (2010). The hexosamine biosynthetic pathway couples growth factor-induced glutamine uptake to glucose metabolism. *Genes & development, 24*(24), 2784-2799.

Wellenreuther, M., & Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends in Ecology & Evolution, 33*(6), 427-440.

Wellenreuther, M., & Hansson, B. (2016). Detecting Polygenic Evolution: Problems, Pitfalls, and Promises. *Trends in Genetics, 32*(3), 155-164. doi:https://doi.org/10.1016/j.tig.2015.12.004

Wellenreuther, M., Le Luyer, J., Cook, D., Ritchie, P. A., & Bernatchez, L. (2019a). Domestication and Temperature Modulate Gene Expression Signatures and Growth in the Australasian Snapper Chrysophrys auratus. *G3 (Bethesda), 9*(1), 105-116. doi:10.1534/g3.118.200647

Wellenreuther, M., Merot, C., Berdan, E., & Bernatchez, L. (2019b). Going beyond SNPs: the role of structural genomic variants in adaptive evolution and species diversification. *Mol Ecol, 0*(ja). doi:10.1111/mec.15066

Wen, J., Sun, X., Chen, H., Liu, H., Lai, R., Li, J., . . . Sheng, W. (2016). Mutation of rnf213a by TALEN causes abnormal angiogenesis and circulation defects in zebrafish. *Brain Research, 1644*, 70-78. doi:https://doi.org/10.1016/j.brainres.2016.04.051

Whitlock, M. C., & Lotterhos, K. E. (2015). Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of F ST. *The American Naturalist, 186*(S1), S24-S36.

Williams, A. N., Ulm, S., Sapienza, T., Lewis, S., & Turney, C. S. (2018). Sea-level change and demography during the last glacial termination and early Holocene across the Australian continent. *Quaternary Science Reviews, 182*, 144-154.

Worm, B., Barbier, E. B., Beaumont, N., Duffy, J. E., Folke, C., Halpern, B. S., . . . Palumbi, S. R. (2006). Impacts of biodiversity loss on ocean ecosystem services. *science, 314*(5800), 787-790.

Xu, D., You, Q., Chi, C., Luo, S., Song, H., Lou, B., & Takeuchi, Y. (2018). Transcriptional response to low temperature in the yellow drum (Nibea albiflora) and identification of genes related to cold stress. *Comp Biochem Physiol Part D Genomics Proteomics, 28*, 80-89. doi:10.1016/j.cbd.2018.07.003

Xu, S., Zhao, L., Xiao, S., & Gao, T. (2019). Whole genome resequencing data for three rockfish species of Sebastes. *Scientific Data, 6*(1), 97. doi:10.1038/s41597-019-0100-z

Yang, C., Wu, F., Lu, X., Jiang, M., Liu, W., Yu, L., . . . Wen, H. (2017). Growth arrest specific gene 2 in tilapia (Oreochromis niloticus): molecular characterization and functional analysis under low-temperature stress. *BMC Molecular Biology, 18*(1), 18. doi:10.1186/s12867-017-0095-y

You, X., Bian, C., Zan, Q., Xu, X., Liu, X., Chen, J., . . . Shi, Q. (2014). Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes. *Nature Communications, 5*(1), 5594. doi:10.1038/ncomms6594

Zeldis, J. (1993). Applicability of egg surveys for spawning-stock biomass estimation of snapper, orange roughy, and hoki in New Zealand. *Bulletin of Marine Science, 53*(2), 864-890.

Zeldis, J., & Francis, R. (1998). A daily egg production method estimate of snapper biomass in Hauraki Gulf, New Zealand. *ICES Journal of Marine Science, 55*(3), 522-534.

Zhang, C., Ni, P., Ahmad, H. I., Gemingguli, M., Baizilaitibei, A., Gulibaheti, D., . . . Xiao, C. (2018). Detecting the population structure and scanning for signatures of selection in horses (Equus caballus) from whole-genome sequencing data. *Evolutionary Bioinformatics, 14*, 1176934318775106.

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., . . . Meredith, R. W. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *science, 346*(6215), 1311-1320.

Zhang, G., Li, L., Meng, J., Qi, H., Qu, T., Xu, F., & Zhang, L. (2016). Molecular basis for adaptation of oysters to stressful marine intertidal environments. *Annual review of animal biosciences, 4*, 357-381.

Zhang, W., & Liu, H. T. (2002). MAPK signal pathways in the regulation of cell proliferation in mammalian cells. *Cell Research, 12*(1), 9-18. doi:10.1038/sj.cr.7290105

Zhao, D., Li, Q., Kong, L., & Yu, H. (2017). Cryptic diversity of marine gastropod Monodonta labio (Trochidae): did the early Pleistocene glacial isolation and sea surface temperature gradient jointly drive diversification of sister species and/or subspecies in the Northwestern Pacific? *Marine Ecology, 38*(4), e12443.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics, 28*(24), 3326-3328.

Zrihan-Licht, S., Lim, J., Keydar, I., Sliwkowski, M. X., Groopman, J. E., & Avraham, H. (1997). Association of csk-homologous kinase (CHK)(formerly MATK) with HER-2/ErbB-2 in breast cancer cells. *Journal of Biological Chemistry, 272*(3), 1856-1863.

Appendix

Appendix 1

Rapid Salt-extraction protocol (adapted Aljanabi and Martinez (1997))

A relatively cheap and quick protocol (~7 hours) that gives good results if there is sufficient tissue. This protocol also has the benefit of not using nasty phenol! 15-100mg tissue, **5-50mg fin clips**

Ingredients:

- 10 mg / mL proteinase K
- 20 mg / mL RNase A
- 1M Tris pH 8.0 Stock Solution
- 0.5M EDTA pH 8.0 Stock solution
- 1M NaCl Stock Solution
- Extraction buffer (0.4 M NaCl, 10 mM Tris-HCl pH 8.0, and 2 mM EDTA pH 8.0: Make from stock solutions)
- TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0)
- 10% sodium dodecyl sulfate (warm to re-dissolve)
- 5M sodium chloride (NaCl saturated dH_2O: Autoclave)
- 100% isopropanol chilled
- 70% ethanol chilled
- Distilled water in squirt bottle

Equipment:

- Thermomixer
- Refrigerated Centrifuge
- Pipettes
- 1.5 mL microcentrifuge tubes
- Forceps
- Scissors
- Kimwipes
- Bunsen burner

Tissue collection for high molecular weight genomic DNA (for GBS, not required for PCR)

Tissue collection and storage has a huge impact on the quality of your DNA. For GBS libraries you need high quality DNA >20kbp. There is some evidence that heat treating the tissue before storing it in the freezer prevents the DNA from degrading. Worth trialling for your species if possible. Here's how to do the heat treatment on fresh tissue:

1. Chill 96% Ethanol in 1.5ml ethanol in 2ml screw top tubes (with an o-ring) on ice
2. Add fresh tissue and allow ethanol to soak into the sample.

3. Within 2 hours heat the tube to 80°C for 5 mins (you can improvise in the field by boiling water and adding the samples for 5 mins)
4. Chill on ice and store in -20°C freezer.

For any extractions it is best to put samples straight into ethanol and freeze them, try to avoid freezing the tissue first.

Sample preparation

1. Select the samples to be used from the freezer and hold on ice during sample prep.
2. Remove the fin-clip/tissue from ethanol storage and cut a portion onto a clean kimwipe in a petri dish.
3. **Weigh ~20-40 mg of sample into a 1.5 mL microcentrifuge tube** and return finclip/tissue to storage. Squash out the ethanol and weigh. For tissue, consider rinsing in water, and re-drying with a kimwipe.
4. Clean forceps and scissors before further sampling:
    a. Wash with distilled water
    b. Wipe with clean paper towel
    c. Wash with ethanol
    d. Sterilise by passing through a flame

Cell lysis

1. Add **480 µL cell lyses buffer or** (400 µL of DNA extraction buffer, 80 µL of 10% sodium dodecyl sulphate (SDS)). You may need to warm the SDS to about 40°C to dissolve it.
2. Heat tubes to 80°C for 5 min to inactivate enzymes such as DNase.
3. Cool samples on ice.
4. Add **10 µL of 10 mg/mL proteinase K** to each tube.
5. Incubate at 56°C for 1.5-3 hr. Invert the sample every 15 minutes during digestion to mix or set orbital mixer to 300rpm throughout digestion. If necessary, add more proteinase K. *Overnight option*: incubate at 37°C overnight

Protein precipitation and removal

6. Spin max speed 5 min.
7. Transfer supernatant to new tube using wide bore tips. Keep pipette tip close to the surface and occasionally mix surface layer with pipette tip without dislodging pellet in the bottom of the tube.
8. **Add 320 µL of 5M sodium chloride** to each tube and mix by inverting the tubes 60 times.
9. Spin max speed 5 min.
10. Transfer supernatant to new tube using wide bore tips.

RNA digestion

11. **Add 5µL of 100 ug/uL RNase A** to each tube, and incubate for 5 min at RT.
12. Spin max speed 5 min.

13. Transfer supernatant to new tube using wide bore tips.

Note: If RNA is coming through in samples increase the digestion time (Sample digestion step 4) or amount of RNase A used in step 5.

DNA precipitation

1. **Add 525µL of chilled 100% isopropanol** to the supernatant, inverted several times, and hold at -20°C for ~1 hr- overnight if more convenient.
2. Centrifuge tubes for 20 min at 13K xg and 4°C.
3. Carefully remove supernatant without disturbing pellet.
4. **Add 1mL of chilled 70% ethanol** to tubes and invert several times.
5. Centrifuge for 10 min at 13K xg and 4°C.
6. Remove the supernatant and air dry the pellet for approximately 15-30 min at 25°C. Do not over-dry your samples.

DNA rehydration

1. **Add 30-100 µL of TE buffer** to the dried pellet and leave to rehydrate for >30 min (or overnight)
2. Gently agitate (flick) tube to re-suspend DNA.
3. Store DNA in fridge until checking quality (no longer than 48 hrs) or store in freezer for long term storage. Remember to label your samples well and date them.

DNA QC

1. Quantify DNA on the Nanodrop. Dilute and re-measure if your DNA is too concentrated. Record ng/ul, 260/280 (1.8 pure DNA; <1.7 indicates protein contamination) ratio and 260/230 ratio (<1.5 indicates salt contamination).
2. Analyse 500ng DNA on a 1% gel. Use a high molecular weight ladder such as lambda Hind III.
3. For GBS quantification using the Qubit is required.

Appendix 2

Test for Heterozygosity over sequencing depth

To verify whether the expected sequencing depth would be sufficient to capture all genomic diversity, levels of observed Heterozygosity (H_O) over sequencing depth were assessed using the samples provided by Plant and Food Research. These individuals had previously been sequenced at a depth of ~30 X. All 12 individuals were subsampled at chosen fractions of the total number of reads (i.e. 0.05, 0.1, 0.2, 0.3, 0.5, and 0.75). Subsampling was done using reformat tool from BBtools (Bushnell, 2017). Reads were aligned to the reference genome using the bam_pipeline, paleomix v1.2.12.3 (detailed description of alignment in section 2.2.4.) (Schubert et al., 2014). The folded Site Frequency Spectra (fSFS) was estimated for each individual in angsd v0.931. The SFS's was folded because the ancestral states of the SNPs were unknown.

Genotype likelihoods were estimated using ANGSD v0.931. In addition, gcc v8.2.0 and htslib v1.9 were used for dependencies for running ANGSD.

The 12 high-coverage samples obtained from PFR were used to assess the change in heterozygosity over sequencing depth. Read files were subsampled (0.05, 0.1, 0.2, 0.3, 0.4, 0.5, and 0.75 of the total number of reads), and aligned to the reference genome (as described in section 2.2.1.4.). Site allele frequency likelihoods were estimated in ANGSD (-dosaf 1) under a folded site frequency spectrum (-fold 1), because the ancestral state of each allele is unknown. The site frequency spectrum (SFS) was produced for each sample and sequencing depth independently using realSFS. Heterozygosity was inferred from the SFS and plotted against the sequencing depth of the sample.

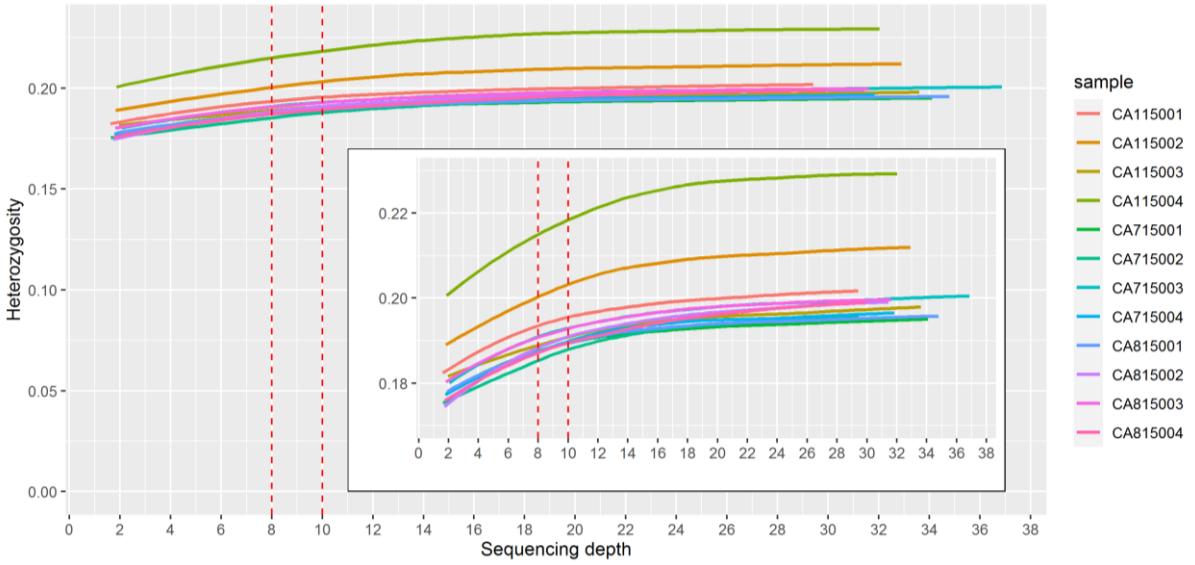

Figure A.1 Heterozygosity plotted over sequencing depth of 12 high-coverage individuals. Vertical lines indicate the minimum coverage (x=8), and average observed coverage (x=10).

Figure A.1 Heterozygosity plotted over sequencing depth of 12 high-coverage individuals. Vertical lines indicate the minimum coverage (x=8), and average observed coverage (x=10).

Appendix 3

Description of SNP filtering chapter 2

SNP filtering parameters were assessed by plotting the number of sites retained over a range of values. A minimum allele frequency (MAF) of 0.01 was determined as an appropriate threshold because this means seven alleles have to be falsely genotyped for a locus to be used in the data set (with a sample size of 350 diploid individuals). This setting meant only 33% of sites were removed compared to 55% when using the more conventional MAF threshold of 0.05 (Figure A.2A). Another advantage is that more rare alleles are retained within the data set, improving the chances of detecting fine-scale genetic differences between populations (De la Cruz & Raska, 2014). This is particularly useful in species with very large effective population sizes and high rates of dispersal (i.e. fish). The high-quality DNA recovered from the samples and high Phred scores of the reads were reflected in the large number of shared SNPs among individuals (Figure A.2B).

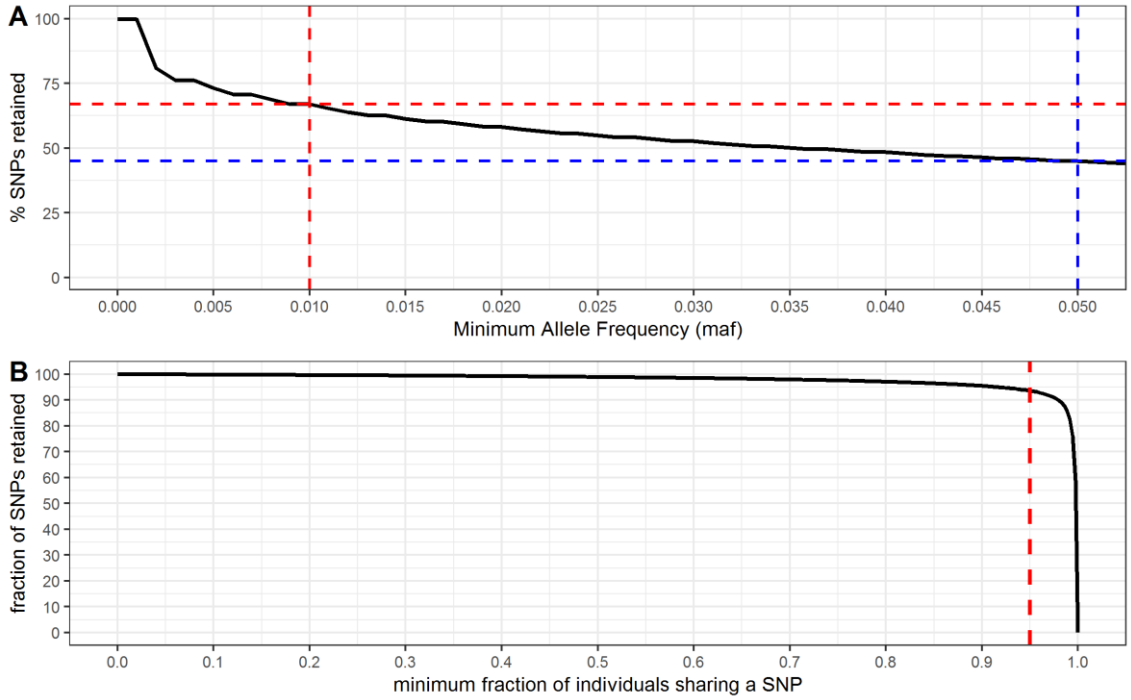

Figure A.2: A) The number of SNPs retained based on minimum allele frequency (maf) and minimum allele count (mac). The red lines highlights the intersect of maf=0.01, the value used in this thesis. The blue lines highlights the intersect of maf=0.05, the default parameters commonly used in population genetic research. B) The percentage of SNPs retained as a function of the minimum fraction of individuals sharing a particular SNP.

After initial quality filtering, 93.6% of sites were shared among 95% of the individuals within the data set. Also, GATK's joint calling algorithm uses allele information across samples to call genotypes, but it can falsely call heterozygous loci when only a single read is available for an individual in the data set. The minimum depth filter (minDP 2) for individual genotypes was

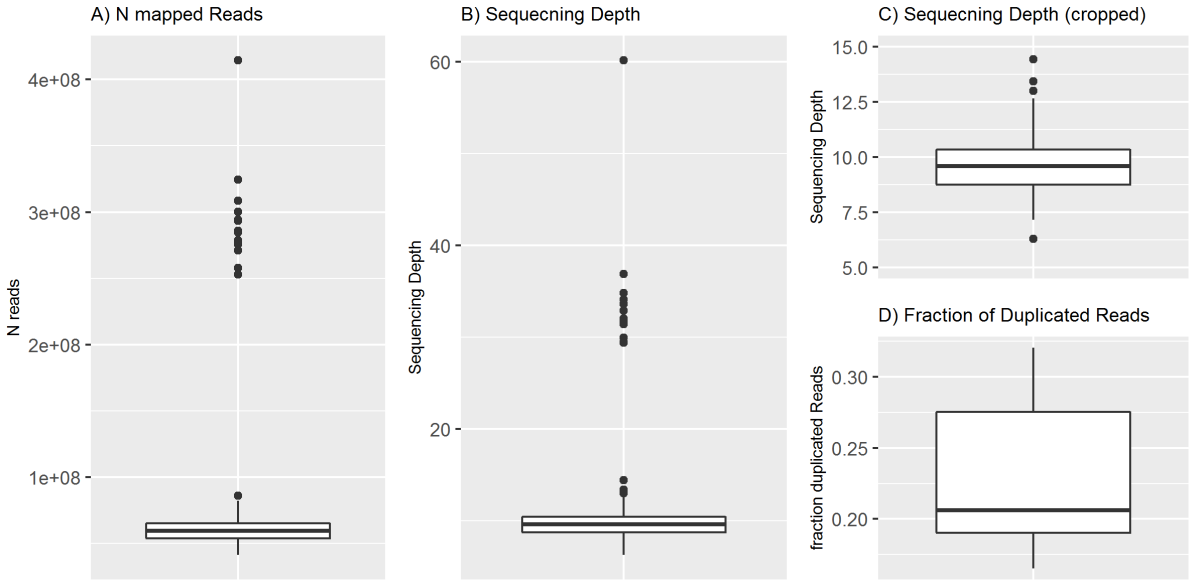

Figure A.3 : A) number of mapped reads per individual, B&C) sequencing doeth per individual, D) fraction of duplicated reads per individual

used as a correction. Plots of heterozygosity per individual show higher levels for the high-coverage individuals, which means heterozygosity was being underestimated in the majority of the individuals (Figure S2.5B). While increasing the coverage per individual would improve heterozygosity estimates, there was no evidence suggesting the difference in coverage introduced biases in the data set. For instance, PCA's do not group high-coverage samples as a consequence of differences in heterozygosity levels (Figure 2.2 & Figure 2.3). The relatively high proportion of sites removed due to allelic imbalance can most likely be attributed to the average sequencing depth of all samples. With an average depth of 10.2x and variable sequencing depths

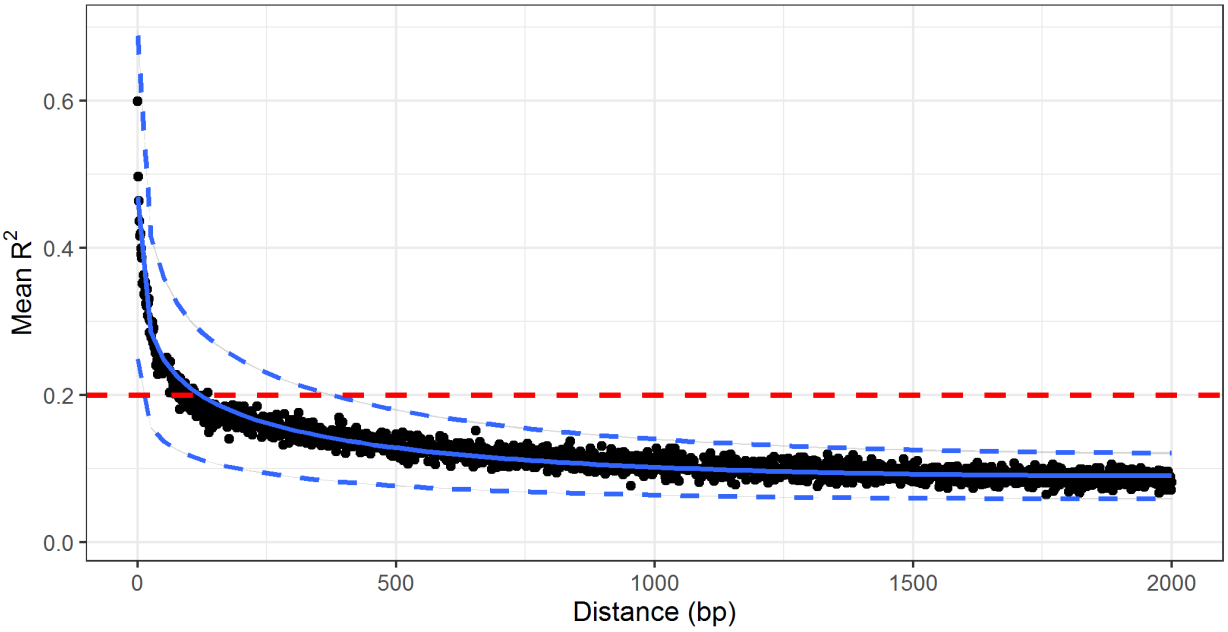

Figure A.4: Linkage disequilibrium decay over genomics distance (in base pairs). The horizontal line dashed line (red) shows commonly used threshold identifying independent degradation of sites.

across the genome, a proportion of all sites will lack the statistical power to reject the unbalanced sampling alleles. The effect of this filter step on the final number of sites was relatively low due to the subsequent, highly stringed, pruning step. To ensure the final data set consisted of independently segregating sites, the data set was pruned for a minimum distance of 1.500 bp between SNPs. This was based on the decay of linkage disequilibrium (LD) ($R^2 < 0.2$) (Figure S2.8). Fish generally have large effective population sizes, and thus high recombination rates, causing associations between sites to rapidly decay (Peñalba & Wolf, 2020). The SNPs in the final data set were at a higher density while segregating independently in the population.