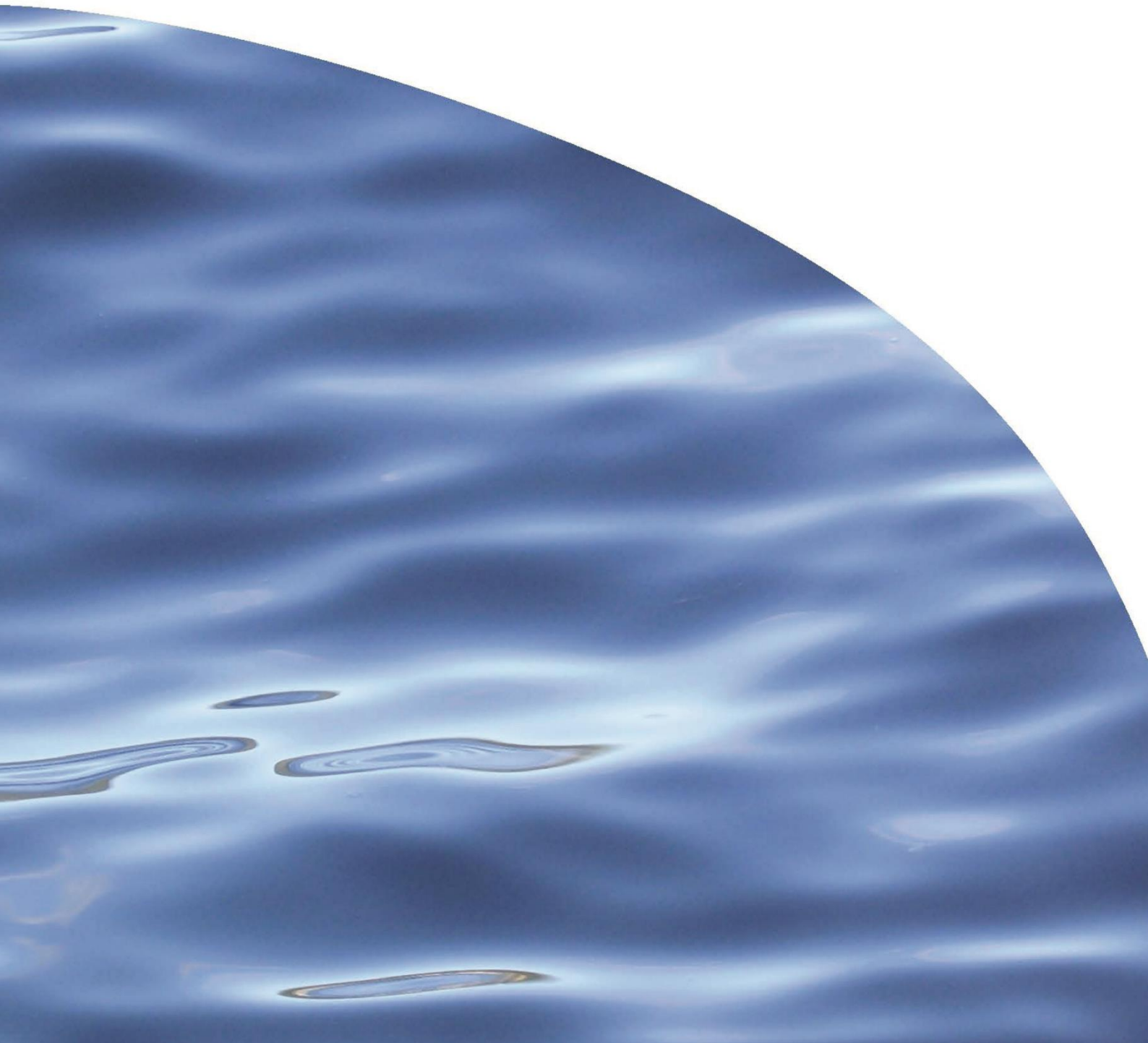REPORT NO. 3416

NEW ZEALAND WILD CHINOOK SALMON POPULATION GENOMICS

NEW ZEALAND WILD CHINOOK SALMON POPULATION GENOMICS

JANE SYMONDS [1], RASMUS GABRIELSSON [1], MONICA VALLENDER[2], KEN DODDS[2], SHANNON CLARKE[2], SEUMAS WALKER[1]

1. CAWTHRON INSTITUTE
2. AGRESEARCH, INVERMAY

Prepared for Fish & Game New Zealand

REVIEWED BY:
Nick King

APPROVED FOR RELEASE BY:
Serean Adams

EXECUTIVE SUMMARY

The overall goal of this project was to estimate the level of population divergence among New Zealand Chinook salmon (*Oncorhynchus tshawytscha*) populations. To achieve this, genotyping-by-sequencing (GBS) was used to genotype 383 wild Chinook salmon from 12 South Island populations. Due to smaller sample sizes in five of the populations sampled, the analysis focussed on seven populations with 35 or more individuals (Lake Paringa, Lake Hawea, Lake Wakatipu, Waitaki River, Rangitata River, Rakaia River, Wairau River). A total of 120,332 single nucleotide polymorphisms (SNPs) were detected in these seven populations which enabled population structure and genetic diversity within and among these populations to be estimated using different statistical methods.

The results from the different methods were in agreement and are summarised as follows:

1. The two landlocked populations (Lake Hawea and Lake Wakatipu) are clearly different from the other sea run populations and from each other. They also have very small effective population sizes and many of the genotyped individuals within each population were inbred and related. This is consistent with the isolated history of these stocks and the lack of migration from other rivers into these closed populations. The results suggest these populations were only founded by a small number of individuals or have since gone through one or more genetic bottlenecks that has reduced the number of breeding individuals in each population.

2. The other five main river (sea run) populations were more closely related to each other, but two populations, the West Coast Lake Paringa catchment and Wairau River populations, were separated using principal component analysis (PCA). Most of the individuals belonging to these populations formed separate clusters, consistent with their more distant geographic locations from the other main rivers. Similarly, the PCA results for the three other East Coast river populations (Waitaki River, Rangitata River, Rakaia River), showed most of the Waitaki individuals to be separated from the other two rivers, whereas individuals collected from the Rakaia and Rangitata rivers could not be differentiated. This was also consistent with other estimates of differentiation, and the higher mean relatedness between these two populations.

3. Some evidence of mixing was identified between some populations. For example, a few individuals from the Paringa system clustered more closely with the Rakaia and Rangitata River samples. This is consistent with the history of these stocks and Fish & Game's release of hatchery-reared yearling salmon smolt from Rakaia River origin into the Paringa catchment for fishery enhancement and management purposes.

4. The ancestral Waitaki River population has the largest effective population size (Ne) and is the most genetically diverse population. The effective population size estimated for the Waitaki River in this study was also similar to the estimate of Ne = 598 by Kinnison et al. (2002).

Overall, the results are consistent with the previous studies; some differentiation was detected among the main populations, albeit lower between the four main East Coast rivers. The population differences detected among the larger rivers could be the result of genetic drift over time and isolation by distance. Population differences may also occur as a result of interannual variation at different locations and this needs to be considered when interpreting the results from a single timepoint at a given location.

The estimates of genetic diversity for the seven main populations have uncovered some potential issues that should be considered when developing a strategy for managing New Zealand wild salmon populations. For example, the low Ne of the landlocked populations and the Lake Paringa and Wairau River populations. Relatedness was also higher in the landlocked and Lake Paringa populations, and inbreeding levels were higher in the landlocked populations. Conversely, the Rangitata and Waitaki Rivers represent more genetically diverse populations, with higher numbers of SNPs, higher heterozygosity and Ne estimates, and lower levels of relatedness. The Rangitata River also has the lowest level of inbreeding, whereas inbreeding was higher in the Waitaki population complex.

This preliminary study provides some recommendations for additional sampling to obtain a more complete picture of the genetic differentiation and diversity of New Zealand South Island Chinook salmon populations. This includes increasing sample sizes for some populations, sampling from additional rivers and sampling at different time from the same populations to assess interannual variation. It is suggested that that a stock management plan be developed once additional sampling and analysis has been completed. Where there is a risk that individuals within a population may be related and matings are managed by Fish & Game, GBS could be used to estimate relatedness between spawning individuals to avoid crossing relatives helping maintain genetic diversity.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

1. INTRODUCTION

1.1. Background

The overall goal of this project was to estimate the level of population divergence among New Zealand Chinook salmon (*Oncorhynchus tshawytscha*) populations. Chinook salmon is native to the Pacific Northwest coast of North America and was introduced to New Zealand over a hundred years ago (McDowall 1994a). The species quickly established wild sea-run populations in South Island rivers, and there are also several self-sustaining land-locked populations in lakes. Research in the 1990s (reviewed by Quinn et al., 2001 and Kinnison et al., 2002) indicated that, as Pacific salmon populations can diverge after colonization, genetic changes may have taken place since the introductions. However, these studies only included a small number of populations and used DNA markers such as microsatellites to examine population structure.

Recent advances in population genomics make it possible to detect previously unidentified population structure, and explore adaptive divergence, potentially revolutionizing the way genetic data are used to manage wild populations (e.g. Saint-Pe et al., 2019). This project therefore aimed to obtain samples from a range of locations across all major New Zealand wild Chinook salmon populations, facilitating the collection of genome-wide sequence information using genotyping-by-sequencing (GBS) and a comprehensive South Island wide population genomic analyses. GBS generates data on thousands of single nucleotide (SNP) markers. Although each SNP has less polymorphic information (biallelic) compared to microsatellite markers, which can have many alleles, this is overcome and superseded by utilizing thousands of SNP markers simultaneously. For example, AgResearch's work on sheep has found 11 microsatellite markers are equivalent to 50 SNPs. SNPs are abundant in the sheep genome and sheep genotyping is now possible using 600,000 SNPs, providing a much more powerful tool for genomic analysis.

The SNP based genotyping results provide valuable information about the current genetic stock structure of New Zealand's wild Chinook salmon populations, assisting sports fishery management, enhancement and conservation efforts. The findings will improve the understanding of how an introduced sportfish species establishes and forms locally adapted populations, which may help guide future management decisions.

1.2. Acclimatisation, biology and historical stock management

The successful establishment of self-sustaining populations of Chinook salmon in New Zealand is the most well-documented transplant of anadromous Pacific salmon outside their native range. New Zealand Chinook salmon have been the subject of numerous studies, with over 60 papers published in the scientific literature. Below, we briefly outline relevant aspects of the introduction, biology and history of hatchery supplementation. For

further details see more comprehensive reviews by Deans et al. (2004), Unwin and Gabrielsson (2018) and Quinn (2018).

All Chinook salmon in New Zealand derive from introductions to the Waitaki River system, primarily the Hakataramea River, between 1901 and 1907 (McDowall 1994a). Introduced Chinook were imported as embryos from the Sacramento River system, California, in particular a Battle Creek population that returned to freshwater in late summer/early fall. By 1907 anadromous fish, resulting from releases of the imported salmon, were observed spawning in the Hakataramea River. Within 5–10 years other spawning populations formed by natural straying and colonization in the other major river drainages along the East Coast of the South Island (McDowall 1994a), where self-sustaining populations still exist today. Salmon naturally dispersed to many rivers, primarily north of the Waitaki system. But notably, progeny from salmon returning to the Waitaki system were also intentionally introduced to the headwater lakes of the southwardly located Clutha River catchment starting in 1917, along with some West Coast catchments (McDowall 1994b). Hence, all Chinook salmon populations in New Zealand originally derive from a common ancestral source introduced to the Waitaki drainage, and their current distribution appear mostly to be formed by natural straying and colonization processes.

The most abundant wild New Zealand stocks are those of the large, glacier-fed, braided East Coast rivers draining the South Island main divide, primarily the Waimakariri, Rakaia, Rangitata, and Waitaki (Quinn & Unwin 1993), which historically supported annual spawning populations of up to 20,000 fish (West & Goode 1987). Smaller runs occur in the Clutha River, having been greatly reduced in number after completion of the Roxburgh Dam in 1956 (Jellyman 1987), and in non-glacial East Coast rivers such as the Waiau, Clarence, Hurunui, and Opihi. Sporadic runs are recorded well outside this range (McDowall 1990), particularly on the South Island West Coast where spawning populations of up to several hundred fish return to the tributaries of Lake Paringa and Lake Mapourika. However, since 2000, spawning runs have been unusually poor and returns to the four main East Coast sea run salmon populations (Waitaki, Rangitata, Rakaia and Waimakariri) have fallen to record low levels, particularly in recent years. This prompted fishery managers to introduce a series of new fishing regulations for the 2019/20 season aimed at significantly reducing harvest.

Chinook salmon display great variation in life history traits, morphology, behavior and other characteristics. This diversity partly reflects differences in rearing conditions, but also genetic adaptations to local environmental conditions, as a consequence of strong natal stream homing behavior (Quinn 2018). However, Chinook salmon can also display a high degree of phenotypic plasticity and can undergo rapid evolution (within 90 years) in response to changes in environmental conditions (Quinn et al., 2001; Hendry and Stearns 2004).

Thomas Quinn, Martin Unwin and Mike Kinnison took advantage of the New Zealand introduction and conducted a large-scale study to examine the environmental and genetic

2

basis for and contemporary evolution of life-history traits (reviewed by Quinn et al., 2001 and Quinn 2018). A key finding from this work was that it provided conclusive evidence of local adaptation (Unwin et al., 2003) such as a 50% higher survival and 'home court' advantage for locally adapted stock. New Zealand salmon were found to vary in size (weight/length) at a given age, age at maturity, fecundity and timing of migration and spawning; traits commonly used in North America to describe and separate distinct populations. Experiments also showed very strong genetic control over some traits, especially the timing of migration and spawning. A combination of environmental and genetic controls influenced other traits such as growth rate, age at maturity and egg production. The genetic control over timing of adult migration and reproduction is believed to be especially important. Firstly, local regimes of natural selection will favor fish that emerge from their redds (nests), enter the ocean and return to their spawning grounds at locally appropriate times of the year. Secondly, as their spawning dates begin to differ individual populations become increasingly isolated from each other. Thus, in combination with geographical isolation the evolution of fitness-related traits such as run time and spawning date help accelerate the divergence of populations in other traits. For a more comprehensive summary of the outcomes and implications, see Deans et al. (2004) and Quinn (2018), along with recent findings from efforts to characterize the ancestry and genetic diversity of Chinook salmon populations expanding through South America's Patagonia (Correa and Moran 2017).

From the late 1970s, considerable effort was made to develop a commercial ocean ranching[1] industry based on Chinook salmon (Deans et al. 2004). In total, over a 20-year period, almost 50 million hatchery salmon were released into South Island rivers. During the ocean-ranching boom, returns of hatchery fish had a significant impact on the recreational salmon fishery. However, hatchery fish tended to be smaller for their age and males typically matured one year earlier than wild born salmon. Scientists also warned that at a population level, hatchery fish may merely be replacing wild fish, as expected if salmon production is density dependent at some critical phase in the life cycle. A consistent pattern across most ocean ranching hatchery release programmes was for a period of early success, followed by a steady decline in salmon survival and return rates. Hence, along with an increasingly lively debate about the adverse impacts of hatcheries on the resilience of wild populations (reviewed by Holmes 2018), sustained poor return rates over 30+ years demonstrate that the viability of hatchery supplementation as a management tool for the New Zealand salmon fishery is limited (Deans et al., 2004; Unwin and Gabrielsson 2018; Holmes 2018).

Despite this, Fish & Game still release large numbers of hatchery produced Chinook salmon smolt annually in an attempt to enhance the East Coast sea-run fisheries, particularly in Canterbury region. A recent review of contemporary stocking practices within New Zealand (Holmes 2018) showed that in total, close to four million salmonids have been

---

[1] The practice of raising young salmon in a hatchery for some, or all, of their first year of life, and releasing them to the ocean in the hope that enough adults will survive and return to the point of release to generate a viable business.

released over the last five years. Chinook salmon make up the majority of these releases (> 75%), mainly in Canterbury. The report expressed concerns that historic and current hatchery practices may have negatively impacted wild sea-run salmon populations. This is consistent with independent recommendations provided by Willis (2018) who cautioned that hatchery operations should avoid mixing locally adapted wild stocks and end the practice of releasing surplus salmon smolt from commercial aquaculture operations into wild sea-run salmon populations. To better guide future management of wild salmon populations, Willis (2018) recommended Fish & Game develop a genetic baseline for population identification and examine if genetic differences can be linked to locally adapted river specific life history characteristics.

1.3. Population genetics microsatellite DNA results

Population differences among wild New Zealand Chinook salmon were last studied in the 1990s to 2000s by Martin Unwin and colleagues at NIWA in collaboration with Thomas Quinn and Michael Kinnison, based at the time at the School of Fisheries, University of Washington, USA. Together, they published multiple papers describing phenotypic and genotypic differences among the NZ populations they studied (e.g. Kinnison et al., 1998; 2002; 2008; 2011; Quinn et. al., 1993; 1996; 2000; 2001; Unwin et al., 2000; 2003).

In one of these papers, Kinnison et al. (2002), 11 microsatellite DNA markers were used to look at salmon population differentiation and effective population sizes after approximately 30 generations since introduction to New Zealand. Samples were analysed from 13 populations from three major catchments (Waimakariri River, Rakaia River and Waitaki River), including samples from the same populations but collected in different years. The study showed that detectable population structure had arisen among the New Zealand populations, consistent with the results from their other papers on the process of evolution and local adaptation by new salmon populations. At the time of the study, the effective population size, Ne, was estimated to be 643 and 598 for the Waimakariri and Waitaki salmon, respectively.

1.4. Population genomics - GBS

Over the last two decades the fast-growing development of high throughput DNA sequencing and new cost-effective methods for sequence-based genotyping have revolutionized our ability to understand genetic variation in populations. These methods now make it possible to assess DNA sequence variation with less bias and at a fraction of previous costs. One such method is genotyping-by-sequencing (GBS), which allows the analysis of 1000s of single nucleotide polymorphisms (SNPs) spread throughout the genome in many individuals (e.g. Elshire et al., 2011). Each SNP represents a difference in a single DNA building block, called a nucleotide, in the DNA sequences being compared.

SNPs represent the most abundant form of genetic variation in genomes and are therefore a very useful tool for detecting differences between individuals or populations.

In 2013, AgResearch obtained funding through MBIE for the "Genomics for Production and Security in a Biological Economy" programme (C10X1306) to develop GBS in a variety of species, including farmed NZ king salmon. This was successful and the salmon farming companies are now employing GBS for a range of applications such as pedigree assignment, estimation of genetic diversity, relatedness and inbreeding, and are aiming to employ genomic selection in the future (Symonds et al., 2019). These applications make use of statistical methods developed by Dodds et al. (2015), enabling a cost-effective DNA profiling tool that assesses approximately 2 % of the genome in the method utilized for salmon. This GBS approach coupled with high-throughput, high quality DNA extraction methods (Clarke et al., 2014) and reduced sequencing costs, provides information such as genetic diversity indices for a similar cost to previous genotyping methods such as microsatellite DNA based methods. The cost is particularly important for smaller companies, or organizations such as Fish & Game, where resources are limited but the analysis of hundreds of individuals is required to provide informative results.

The GBS pipeline developed for NZ salmon generates tens of thousands of SNPs per individual, usually in the range to 35,000 to 45,000 SNPs. Dodds et al. (2018a, b) have developed a suite of population genomics methods to estimate genetic diversity indices, relatedness and population structure based on these SNPs. Through collaboration between Fish& Game, Cawthron Institute, AgResearch and University of Otago, these methods were employed in this study to estimate population structure and genetic diversity for wild NZ salmon.

1.5. Scope of report

Cawthron Institute was commissioned by Fish & Game New Zealand in 2018 to produce this report as a first step towards gaining a better understanding of the current population structure of wild New Zealand Chinook salmon using new population genomics tools, namely genotyping-by-sequencing (GBS). A total of 420 samples were collected from around the South Island from 12 populations and submitted to GenomNZ at AgResearch for genotyping using GBS.

The aims of this report are firstly, to summarise the sampling and genotyping results and secondly, to present the results of the population genomics statistical analysis that was completed to assess genetic diversity and population structure among NZ Chinook salmon populations. These results are then discussed in comparison to previous population genetics studies. The report concludes with a discussion and recommended guidelines for using GBS to monitor and better understand wild populations and improve broodstock management for future hatchery supplementation.

2. METHODS

2.1. Sample collection

The aim was to sample adult salmon from as many of the main South Island river systems as possible on the East Coast, as well as the landlocked lakes (e.g. Lake Hawea, Lake Wakatipu) and some samples from the West Coast to represent the range of distribution of wild Chinook salmon in New Zealand. Fish & Game funding was available to genotype 400 individuals.

Sample collection was coordinated by Helen Trotter (Otago Fish & Game) with Rasmus Gabrielsson and Seumas Walker from Cawthron, with significant support from regional Fish & Game staff and anglers. The aim was to collect 50 fin clip samples from adult wild salmon per "population" and preserve them in 100 % ethanol. Tubes with ethanol and a data recording form were provided to ensure date, location and lifestage were recorded for each sample.

2.2. GBS, SNP analysis and population differentiation

The DNA extraction and GBS was carried out by GenomNZ and AgResearch according to the methods described in Clarke et al. (2014) and Dodds et al. (2015).

To identify SNPs, a reference-based approach was used, which utilised the reference genome of a Chinook salmon from a Canadian hatchery (Christensen et al., 2018). The identified SNPs were then filtered and underwent a rigorous quality control process, including removal of any SNPs that were potentially located in the duplicated part of the salmon genome.

Following data processing, the KGD program (Kinship using GBS with Depth adjustment, Dodds et al., 2015) was used to analyse the data and establish relatedness between the samples and determine individual inbreeding levels. The KGD analysis also included principal component analysis (PCA) which is a method used in population genetics to visualize genetic distance between populations.

Another method used to determine genetic differences between populations is known as the fixation index, Fst. Fst values between populations were estimated using KGD-popgen based on methods develop by Weir (1996). Fst is a measure of the SNP alleles in common across groups being compared and also considers differences in allele frequencies between populations. The more SNPs that are found only in some populations and are not present in others (i.e. are fixed in these populations), the higher the Fst estimates between these populations are likely to be. Values can range from 0 to 1. High Fst implies a considerable degree of differentiation among populations.

2.3. Genetic diversity estimates

Levels of genetic diversity within populations were estimated using the software packages KGD, KGD-Popgen and GUS-LD (Bilton et al., 2018) developed by AgResearch in the statistical analysis package R to analyse GBS SNP data (https://github.com/AgResearch/KGD; https://github.com/AgResearch/GUS-LD).

Four different estimates of diversity were determined separately within each population:

1. Number of variable SNPs detected within a population

2. Minor Allele Frequency (MAF)

Minor allele frequency (MAF) is the frequency at which the second most common SNP allele in a particular sequence occurs in a given population. MAF is widely used in population genetics studies because it provides information to differentiate between common and rare variants in populations and their distribution.

3. Heterozygosity

A salmon is heterozygous if its DNA contains two different alleles for a particular SNP. For analysis, mean heterozygosity is calculated across all SNPs and is a valuable parameter to estimate the degree of genetic variation within a population. A higher heterozygosity implies more genetic variability.

Expected heterozygosity (He) is the proportion of SNPs, with the observed allele frequencies, expected to be heterozygous in a randomly mating population (ranging from 0 to 1.0).

Observed heterozygosity (Ho) is the actual observed proportion of heterozygotes in the population being analysed, averaged over all SNPs.

Ho*/He* is the ratio of raw observed heterozygosity to expected heterozygosity taking into account read depth (the number of times that a given SNP in the genome has been sequenced during the genotyping of an individual).

If Ho is lower than expected, this could be due to forces such as inbreeding. If Ho is higher than expected, it could be due to the mixing of two previously isolated populations.

4. Effective population size (Ne)

The effective population size is the size of an idealized population (random mating between and within sexes including selfing) that would give the same rate of increase in inbreeding as the population under study. The smaller the Ne, the less likely a population would be able to adapt to change due to decreased genetic variation.

Ne was calculated using the mean linkage disequilibrium (r^2) between SNPs on 34 different chromosomes, as defined in the Chinook salmon reference genome. The SNPs selected were equidistant across the chromosome, and 100 pairs per chromosome were utilised. The r^2 values were calculated using the method of Bilton et al. (2018).

These diversity indices were calculated for the 7 individual populations where SNPs for 35 or more individuals were available for analysis after the quality control process. In addition, all individuals from all 12 populations sampled were combined to calculate these indices for New Zealand wild salmon (n = 383) and also for only the sea run (SR) populations (n= 294) or land locked (n= 89) populations.

3. RESULTS

3.1. Populations sampled and SNP genotyping results

A total of 420 samples sourced from the 12 populations shown in Figure 1 and Table 1 were genotyped. As all samples were sequenced regardless of DNA quality, filtering the resulting SNP data so that accurate interpretations could be made while retaining as many SNPs as possible was a priority. As part of the quality control process, 37 samples with low read depth were removed and 383 samples remained for analysis. Of the 37 samples removed, 1 was from Lake Wakatipu, 1 from Hurunui River, 1 from Lake Mapourika, 1 from Lower Clutha, 6 from Lake Paringa, 14 from Wairau River, 7 from the Rakaia and 3 from the Rangitata and Waitaki rivers.
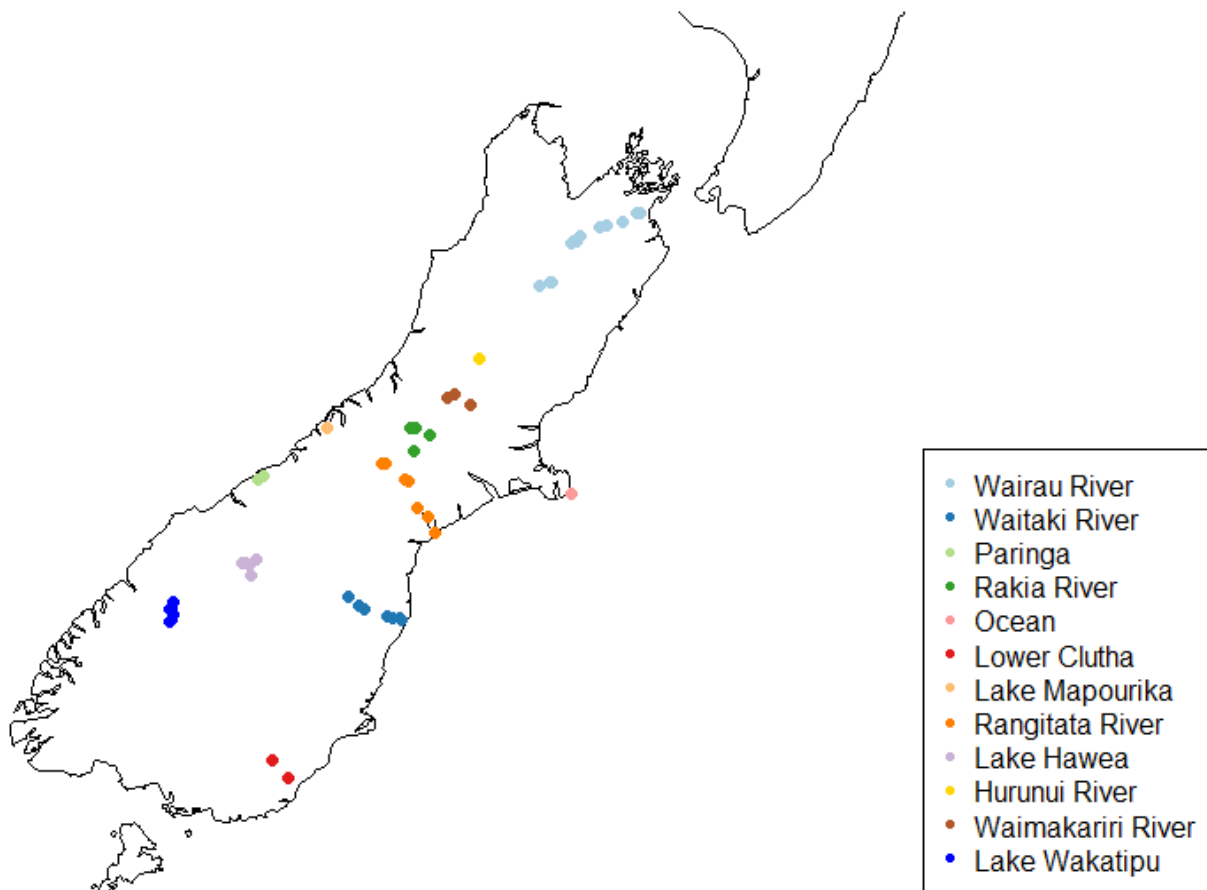

Figure 1.   Map of the South Island and location of the samples collected and their population origin.

.

Table 1.   Populations sampled, Fish & Game region, number of individuals sampled (with the number included in the SNP analysis following filtering in parentheses), timing of sampling and designation as either a landlocked or sea run population for the seven populations with sample sizes of 35 or more. The Central South Island (CSI) Fish & Game region, including the South Canterbury area. The same color coding used in Table 1 is used in Figures 1, 2, 3, 4 and 6 to identify the populations.

Colour code	Population	Region	Number of Individuals	Year collected	Sea Run or Land Locked
	Wairau River	Nelson-Marlborough	68 (54)	June 2015-April 2018	Sea Run
	Hurunui River	North Canterbury	5 (4)	May 2018	
	Waimakariri River	North Canterbury	9 (9)	June-July 2018	
	Rakaia River	CSI & North Canterbury	42 (35)	Feb-Aug 2018	Sea Run
	Rangitata River	CSI	65 (62)	2017-early 2018	Sea Run
	Lake Mapourika	West Coast	6 (5)	June 2018	
	Paringa	West Coast	57 (51)	Dec 2017-June 2018	Sea Run
	Waitaki River	CSI	56 (53)	2017-2018	Sea Run
	Lake Hawea	Otago	48 (48)	May, Oct, Nov 2017	Landlocked
	Lake Wakatipu	Otago	42 (41)	April, May, October 2017	Landlocked
	Lower Clutha	Otago	6 (5)	2016-2017	
	Ocean	CSI	16 (16)	Jan 2017	

After the initial SNP filtering process 252,662 SNPs remained. This number was further reduced through additional quality control processes, leaving a total of 120,811 SNPs for the analysis with a mean sample depth of 3.59. The number of informative (i.e. variable) SNPs ranged from 73,608 (Lake Wakatipu) to 97,229 (Waitaki River) in the seven populations where at least 35 individuals contributed to the SNP analysis following completion of the quality control process (Table 2).

3.2. Relatedness and population structure

The KGD software was used to calculate the relatedness between all possible combinations of the 383 individuals genotyped. The KGD method is an unbiased estimate, given the SNP allele frequencies, of relatedness obtained by using only SNPs with genotype calls in both of the individuals being analysed. This relatedness is depicted in a genomic relationship matrix (GRM) heat map (Figure 2). This displays the self-relatedness on the diagonal and relatedness between individuals on the off-diagonal with each group color-coded on the axis using the same population colour code used in Table 1. The heatmap helps understand the relationships between populations (color coded) and individuals.

Figure 2 clearly shows that the two populations at the bottom left of the GRM, Lake Hawea (lilac) and Lake Wakatipu (dark blue), are not closely related to the other populations, or each other, but many related individuals exist within these populations. Lake Paringa (light green), Wairau River (light blue), Waitaki River (medium blue) and the Rakaia River (dark green) also formed defined clusters representing the majority of the individuals in these populations.

Table 2 shows the average relatedness among the 7 main populations and confirms the interpretation of the data in Figure 2.
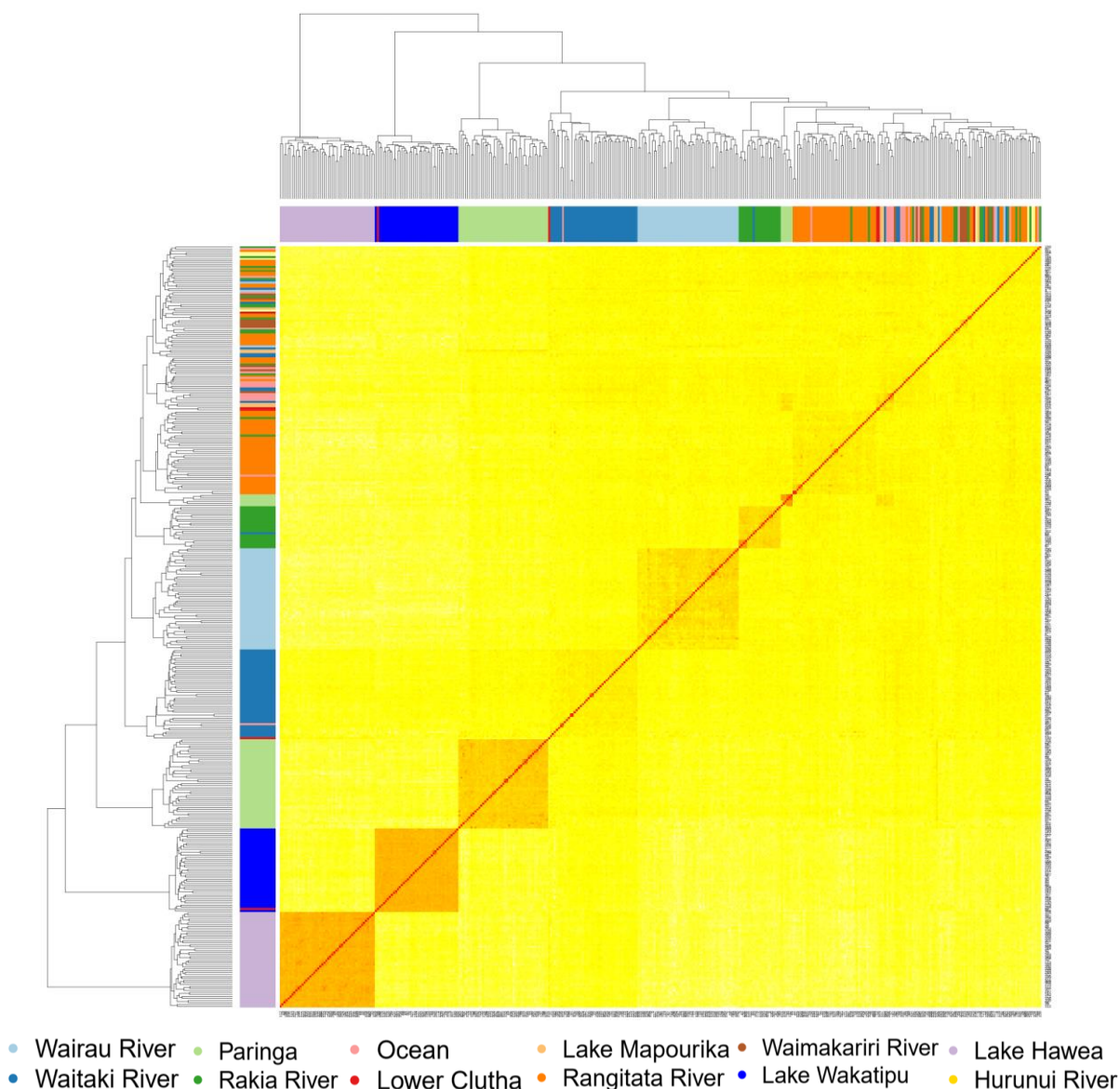

Figure 2.   Genomic relationship matrix heat map between all 383 individuals. Self-relatedness on the diagonal and relatedness between individuals on the off diagonal. Colours in the side bars refer to the population origin (see Table 1). The darker and redder the colour, the more related the individuals.

Table 2.   Average relatedness among the seven main populations (n=344). Diagonal values are the mean of individual relatedness within a population. Conditional formatting has been applied to the relatedness values, green = lower relatedness to red = higher relatedness.

Population	Lake Hawea	Lake Wakatipu	Paringa	Rakaia River	Rangitata River	Wairau River	Waitaki River	Total
Lake Hawea	0.223	-0.013	-0.020	-0.022	-0.023	-0.024	-0.001	0.016
Lake Wakatipu	-0.012	0.216	-0.014	-0.016	-0.014	-0.021	0.002	0.015
Paringa	-0.018	-0.013	0.111	-0.001	0.003	-0.004	0.013	0.014
Rakaia River	-0.021	-0.015	0.001	0.052	0.026	0.020	0.018	0.011
Rangitata River	-0.023	-0.014	0.004	0.026	0.039	0.020	0.015	0.011
Wairau River	-0.025	-0.019	-0.002	0.018	0.021	0.078	0.011	0.013
Waitaki River	-0.002	0.002	0.014	0.015	0.015	0.012	0.040	0.014
Total	0.016	0.014	0.015	0.010	0.011	0.013	0.014	0.013

Principal component analysis (PCA) was also used to visualize the variation between individuals and population groups (Figure 3). The axes of a PCA plot are arbitrary scales that show the most variation accounted for in a dataset (PCA 1) and the second most variation which can be accounted for in a dataset (PCA 2). This allows interpretation of groups or data points and how different, or how much they vary from one another, as the distance separating them in any given direction is directly correlated to how much difference/variation there is. This method is used in population genetics to visualize the relatedness and genetic distance between populations.

The PCA plot (Figure 3) is consistent with the GRM and shows the Lake Hawea and Lake Wakatipu populations clearly separated from the other populations and from each other. The other 10 populations cluster together, with the Waitaki and Wairau river samples forming their own clusters within this group.

A PCA plot without the Lake Hawea and Lake Wakatipu samples and only the 5 sea run populations where 35 of more individuals were successfully genotyped is shown in Figure 4. The Wairau River, Lake Paringa and Waitaki River samples form separate clusters, with a small number of individuals from these populations clustering close to or within the Rakaia and Rangitata River sample cluster. This is consistent with the results in the GRM heat map (Figure 2).
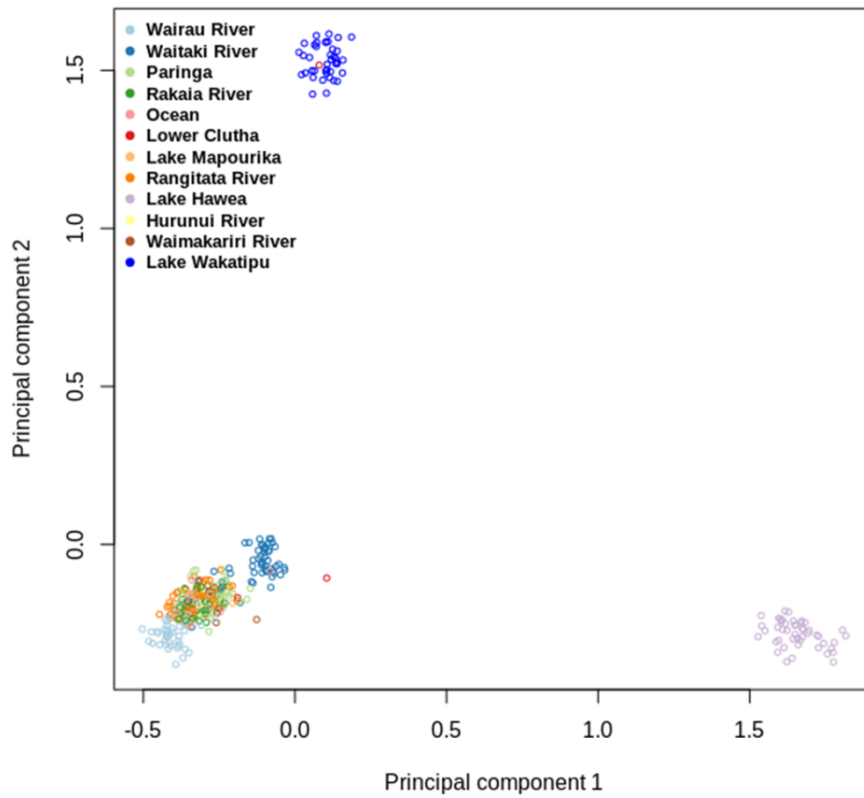
Figure 3.   Principal component analysis of all 12 populations and 383 individuals, as reported by KGD. PC1=21% of variation, PC2=15%.
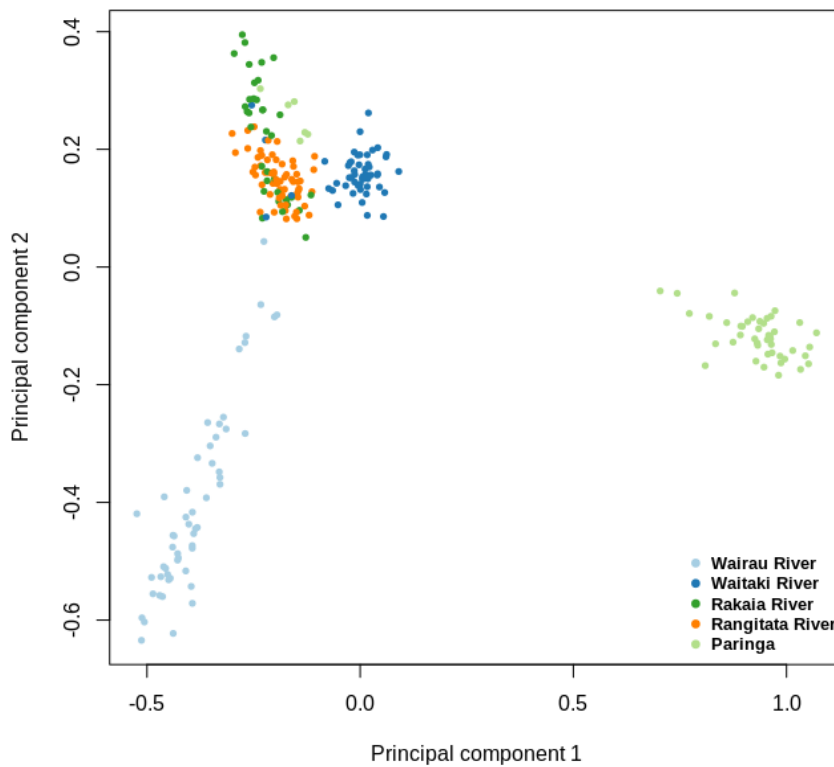

Figure 4.   Principal component analysis of only the five non-landlocked (sea run) populations with at least 35 or more individuals genotyped. PC1 = 14.5% of variation, PC2=4% of variation.

The Fst analysis also showed the highest differentiation between the landlocked lakes and the other populations (Table 3). Less differentiation was observed among salmon sampled from the four East Coast rivers (Rakaia, Rangitata, Wairau and Waitaki) surveyed in this study. The Fst values among population subgroups within the main river systems are also shown in Table 1 in Appendix 2. Most of the within-river Fst values were consistent with the overall analysis for each river; similar Fst values were obtained for all the subgroups within a river system. Exceptions are the lower Fst values between the Wairau Rainbow River and Upper Wairau River samples and higher Fst values between the Lower and Upper Wairau individuals.

The distances (kilometres) between the four East Coast rivers and the Fst values were plotted and a regression analysis was completed. This showed that the Fst values between these four populations tended to increase the further the distance between the populations (r^2 = 0.678), suggesting that gene flow between these populations increases with geographic distance.

Table 3.   Pairwise mean Fst values between the seven 7 main populations.

Population	Paringa	Lake Wakatipu	Rakaia River	Rangitata River	Wairau River	Waitaki River
Lake Hawea	0.030	0.036	0.027	0.025	0.028	0.022
Paringa		0.029	0.017	0.015	0.019	0.014
Lake Wakatipu			0.026	0.023	0.027	0.021
Rakaia River				0.009	0.013	0.010
Rangitata River					0.011	0.009
Wairau River						0.012

3.3. Estimates of genetic diversity

A total of 120,811 SNPs was detected in the 383 individuals genotyped (Table 4) and their frequency distribution is shown in Figure 5. In Figure 5, all the populations are combined and the MAF distribution shows a higher proportion of the lower MAF SNPs than for the higher MAF SNPs in these populations. This distribution is encouraging as a flatter distribution would be expected if genetic variation was low overall. MAF distribution graphs were also completed for the individual populations and examples are provided in Appendix 1.
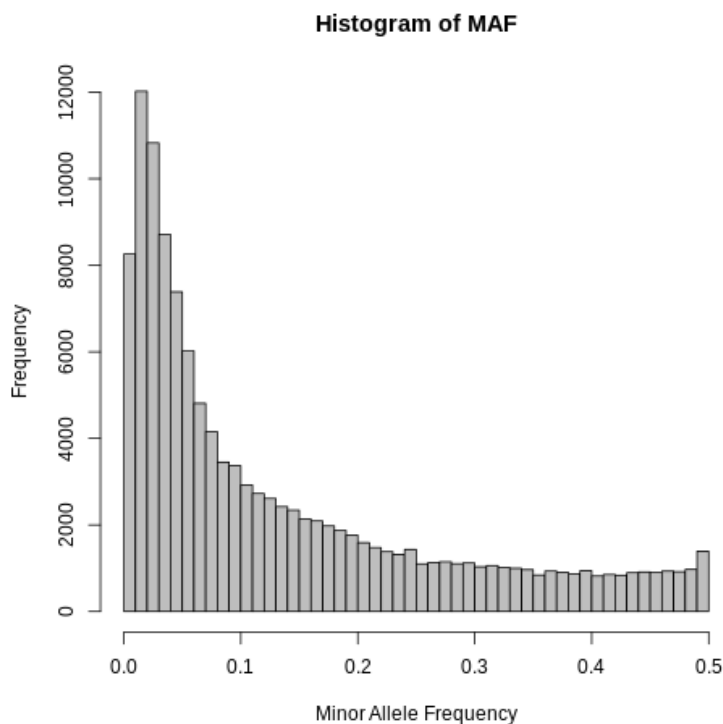
Figure 5.   Minor allele frequency distribution from all 383 individuals genotyped.

As a comparison of genetic diversity, levels of genetic variation for all 12 populations combined, for only the sea run populations (SR) and within each separate population, were determined by four different estimates of diversity: number of variable SNPs detected within the group, MAF, heterozygosity and effective population size, Ne. These are summarised in Table 4. The proportion of MAF >0.2 indicates the proportion of SNPs that were found at a higher frequency in the population and the higher this value, the less diverse the population.

For the seven main populations, all four diversity indices showed a consistent pattern, lower in the land locked lakes and higher in the Waitaki River followed by the Rangitata River. Observed heterozygosity was consistently higher than expected in all seven populations.

Table 4.   Population statistics related to diversity. The proportion of SNPs with MAF > 0.2 is relative to the polymorphic (MAF > 0) SNPs. Ne is the effective population size, while n is the number of individuals sampled and successfully genotyped per population. Ho is the observed heterozygosity. He is the expected heterozygosity for the underlying genotypes (not directly observed). Ho*/He* is the observed heterozygosity in relation to the expected heterozygosity when taking into account sequencing depth.

Group	n	#SNPs, MAF > 0	Proportion MAF > 0.2	Ho*/He*	Ho	He	Ne
All 12 populations							
All	383	120811	0.239	1.202	0.222	0.185	531
Sea Run	294	119335	0.243	1.202	0.222	0.185	641
Lake Hawea	48	75402	0.379	1.325	0.209	0.158	9
Lake Wakatipu	41	73608	0.384	1.202	0.189	0.157	10
7 main populations							
All	344	120332	0.240	1.191	0.222	0.186	∞
Sea Run	255	117946	0.246	1.208	0.224	0.185	455
Lake Hawea	48	75369	0.379	1.328	0.211	0.159	9
Lake Wakatipu	41	73615	0.384	1.298	0.204	0.157	10
Lake Paringa	51	83345	0.346	1.264	0.212	0.168	21
Rakaia River	35	82054	0.352	1.252	0.210	0.168	69
Rangitata River	62	94558	0.308	1.250	0.221	0.177	113
Wairau River	54	86058	0.341	1.260	0.216	0.171	28
Waitaki River	53	97229	0.308	1.253	0.226	0.180	501

3.4. Relatedness and inbreeding

As well as population-based statistics, the GBS data were analysed to estimate how inbred each individual was, and how related each individual was to all the other individuals genotyped, within the same population and also between populations. Figure 6 shows the number of individuals within each inbreeding range for the seven main populations (n=344). Figure 7 shows the same data but as a box plot for each population.
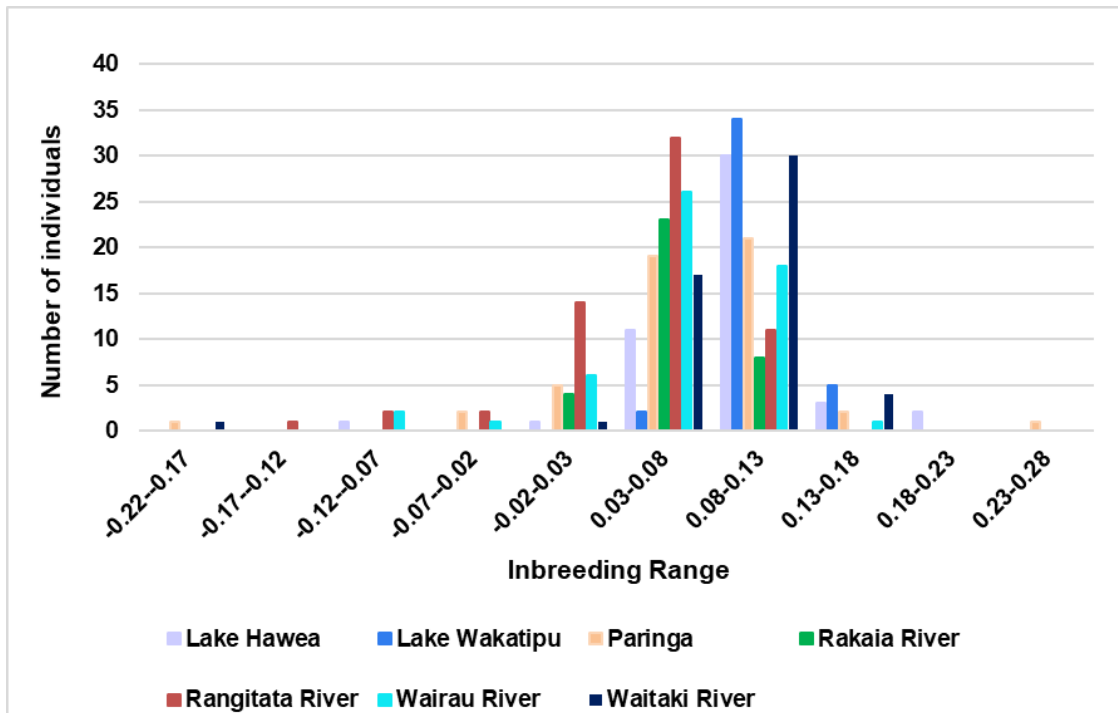
Figure 6.   Frequency distribution of individual inbreeding levels by population (seven main populations only).
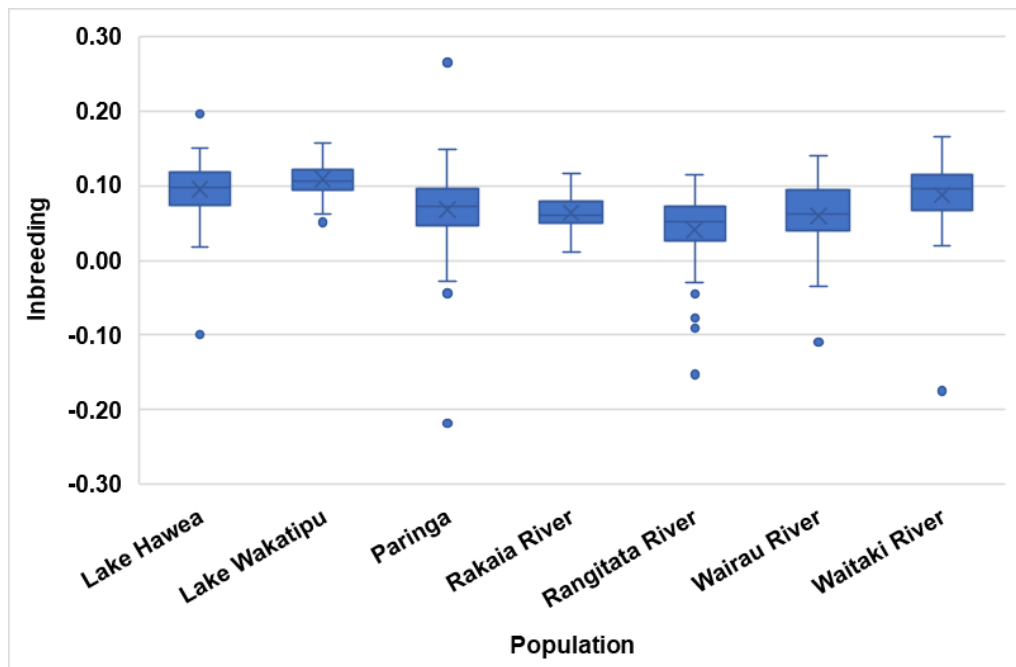

Figure 7.   Box plot of inbreeding estimates for the individuals genotyped in the seven main populations. X = mean, median = line in the box, top and bottom lines of the box = 1st and 3rd quartiles. Minimum and maximum = vertical lines. Outliers = dots above and below the lines.

Mean inbreeding was the lowest in the Rangitata population (4.1 %) and highest in the Lake Wakatipu population (10.8 %). Levels of individual inbreeding are also shown for the four main East Coast river systems grouped by the sample location within the rivers and summary charts are provided in Appendix 2.

Relatedness is a useful measure at the population level (see Figure 2 and Table 2) and provides a measure of how closely individuals within a population are related, and therefore should not ideally be crossed together to avoid inbreeding (based on relatedness above a certain set threshold) if they were available as breeding individuals. For broodstock managers, being able to estimate relatedness between individuals also allows construction of mating designs to control inbreeding if breeding individuals are genotyped. It also helps determine the probability (and risk) that related individuals will mate within a population in the absence of genotyping data.

For the seven main populations the relatedness of all possible combinations of individuals was estimated and the crosses were assigned as being above or below the acceptable relatedness threshold (set at R ≥ 0.125). In the four East Coast river systems surveyed most of the individuals are not highly related so there is a relatively low percentage of crosses between related individuals, assuming that all individuals have an equal chance of mating. In the two landlocked lake populations the opposite is the case, where most individuals are highly related.

Table 5.   Percentage of potential "acceptable" crosses between all individuals within a population where the relatedness between the two individuals is less than R = 0.125.

Population	% of acceptable crosses
Lake Hawea	0.00%
Lake Wakatipu	0.00%
Paringa	57.20%
Rakaia River	84.15%
Wairau River	89.55%
Rangitata River	95.81%
Waitaki River	97.54%

4. DISCUSSION

GBS was successfully used to genotype 383 wild Chinook salmon from 12 South Island populations. Due to the small sample sizes (16 individuals or less) in five of the populations analysed, these populations were only included in the main PCA plot (Figure 2) and Table 2. The rest of the analysis was conducted only on the seven populations with 35 or more individuals. A total of 120,332 SNPs was detected in these seven populations which enabled population structure and genetic diversity within and among these populations to be estimated using different statistical methods.

The results from the different methods were in agreement and are summarised as follows:

1. The two landlocked populations are clearly different from the other sea run populations and from each other. They also have very small effective population sizes (9 to 10 individuals) and many of the genotyped individuals within each population were inbred and related. This is consistent with the isolated history of these stocks and the lack of migration from other rivers into these closed populations. The results suggest these populations were only founded by a small number of individuals or have since gone through one or more genetic bottlenecks that has reduced the number of breeding individuals in each population.

2. The other five main river (sea run) populations were more closely related to each other, but two populations, the West Coast Lake Paringa catchment and Wairau River populations, were separated using PCA analysis (Figure 4), with most of the individuals belonging to these populations forming separate clusters, consistent with their more distant geographic locations from the other main rivers (Rakaia, Rangitata and Waitaki). Similarly, within the three latter populations, most of the Waitaki individuals were separated from the other two by the PCA analysis whereas individuals collected from the Rakaia and Rangitata rivers could not be differentiated by PCA analysis. This was also consistent with the low Fst values between these two populations and their higher mean relatedness (R = 0.026, Table 2).

3. Although the PCA analysis showed clear population clustering based on the location of the sampling for most individuals, some evidence of mixing was identified, such as a few individuals from the Paringa system clustering more closely with the Rakaia and Rangitata River samples. These results are consistent with opportunistic releases by Fish & Game of hatchery-reared yearling salmon smolt from Rakaia River origin into the Paringa catchment for fishery enhancement and management purposes. However, it should be noted that human error, such as mislabelling, incorrect recordkeeping, or sample mixing could result in individuals being allocated to the wrong population. In addition, straying and/or stock mixing (intentional or otherwise) could also result in some individuals clustering with other populations.

4. The ancestral Waitaki River population has the largest effective population size (Ne = 502) and is the most genetically diverse population based on all four indices utilised (Table 4). The effective population size estimated for the Waitaki River in this study was also similar to the estimate of Ne = 598 by Kinnison et al. (2002).

Overall, the present study results are consistent with those from previous studies by Unwin, Kinnison and Quinn (reviewed in Quinn 2018), as some differentiation was detected among the main populations we analysed based on the PCA and Fst results, although the Fst values between the four main East Coast rivers were low. It is possible that the population differences we have detected among the larger rivers could be the result genetic drift over time and isolation by distance.

Population differences may also occur as a result of interannual variation at different locations. For example, age and timing of sexual maturation can lead to genetic differences among salmon year classes spawning in the same rivers. Kinnison et al. (2002) found evidence for temporal variation and this needs to be considered when interpreting the results from a single timepoint at a given location, as was utilized in our study.

The estimates of genetic diversity for the seven main populations have uncovered some populations with reduced diversity plus elevated levels of relatedness and inbreeding. This is demonstrated by the low Ne of the landlocked populations and the Lake Paringa and Wairau River populations. Elevated levels of relatedness were also identified in the landlocked and Lake Paringa populations, and inbreeding levels were high in the landlocked populations. Conversely, the Rangitata and Waitaki Rivers represent more genetically diverse populations, with higher numbers of SNPs, higher heterozygosity and Ne estimates, and lower levels of relatedness. The Rangitata River also has the lowest level of inbreeding, whereas inbreeding appears higher in the Waitaki population complex.

5. RECOMMENDATIONS

1. The GBS results have indicated that New Zealand salmon from seven different populations show signs of genetic differentiation, albeit at relatively lower levels among the sea run populations from the East Coast rivers. If Fish & Game wish to further characterise this and include other main South Island rivers, for example the Clarence, Waiau, Hurunui, Waimakariri and Clutha rivers, more samples will need to be collected. Ideally at least 35 to 50 individuals per population should be sampled during the same season and only adults should be sampled for consistency across studies, unless management and research questions require other life stages to be specifically targeted for sampling.

2. Interannual variations in genetic stock structure can contribute to population differences and/or estimates of genetic diversity. If additional variation or differences are detected this may change the interpretation of the results and subsequent management decisions. Therefore, this needs to be considered, as does ensuring that the life stages that are sampled are consistent and clearly recorded. Therefore, repeat sampling of the main seven populations in this study should be considered to ensure the findings in the study are reproducible and/or if temporal variation needs to be considered. Targeted sampling to characterize the genetic diversity of locally adapted subpopulations displaying novel life history strategies is also advised. For example, those displaying extended (stream-type) freshwater rearing and freshwater lake resident life history strategies.

3. This additional sampling is recommended before considering a population management plan for New Zealand wild salmon. However, given the potential for reduced population sizes, increased relatedness among individuals and higher than ideal inbreeding levels in some populations, managing the populations to avoid a further decline in these parameters is recommended. Until additional information is available, breeding within wild populations, rather than between, and ensuring adequate numbers of males and females are used in any supportive breeding programmes is advised. If breeding individuals could be tagged and genotyped in advance of spawning, mating strategies could be developed to avoid crossing relatives and for maintaining the genetic diversity currently present in the broodstock. This may be particularly helpful for wild populations with higher inbreeding levels, such as the landlocked populations in the Southern Lakes.

6. ACKNOWLEDGEMENTS

7. REFERENCES

Bilton, TP, McEwan, JC, Clarke, SM, Brauning, R, Van Stijn, TC, et al. (2018). Linkage disequilibrium estimation in low coverage high-throughput sequencing data. *Genetics*, 209(2), 289-400. doi:10.1534/genetics.118.300831

Christensen KA, Leong JS, Sakhrani D, Biagi CA, Minkley DR, et al. (2018) Chinook salmon (*Oncorhynchus tshawytscha*) genome and transcriptome. *PLoS ONE* 13(4): e0195461. https://doi.org/10.1371/journal.pone.0195461

Clarke, SM., Henry, HM., Dodds, KG, Jowett, TWD., Manley, TR, et al. (2014). A high throughput single nucleotide polymorphism multiplex assay for parentage assignment in New Zealand sheep. *PLoS One* 9: e93392. doi: 10.1371/journal.pone.0093392

Correa, C. and Moran, P. (2017). Polyphyletic ancestry of expanding Patagonian Chinook salmon populations. Scientific Reports 7: 14338. DOI:10.1038/s41598-017-14465-y

Deans, N., Unwin, M. and Rodway, M. 2004. Sport fishery management. In: Harding J. S.; Mosely, M. P.; Pearson, C. P.; and Sorrell, B. K. (editors), 2004. Freshwaters of New Zealand. New Zealand Hydrological Society Inc. and New Zealand Limnological Society Inc., Christchurch, New Zealand. 764 pp.

Dodds, KG, McEwan, JC, Brauning, R, Anderson, RA, Van Stijn, TC, et al. (2015). Construction of relatedness matrices using genotyping-by-sequencing data. BMC Genomics 16:1047. doi: 10.1186/s12864-015-2252-3

Dodds, KG, Symonds, JE, Bilton, TP, Brauning, R, Ashby et al. (2018a). Use of genotyping-by-sequencing for studying the genetic characteristics of a population. International Symposium of Genetics in Aquaculture XIII, Abstract PG3.

Dodds, KG, Symonds, JE, Brauning, R, McEwan, JC and Clarke, SM (2018b) A depth-adjusted FST calculation for low-depth sequencing data. New Zealand Molecular Mapping Workshop (Mapnet), (abstract).

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6: e19379. doi: 10.1371/journal.pone.001 9379.

Hendry, A. P. and Stearns, S. C. 2004. Evolution illuminated: salmon and their relatives. Oxford University Press, Oxford, UK.

Holmes, R (2018). A review of contemporary salmonid stocking practices in New Zealand. Prepared for Fish & Game New Zealand. Cawthron Report No. 3207. 44 p. plus appendices.

Jellyman, DJ (1987). Possible impact of hydro development on fish and fisheries of the lower Clutha River. New Zealand Freshwater Fisheries Report No. 92, Freshwater Fisheries Centre MAFFish, Christchurch.

Kinnison MT, Bentzen P, Unwin MJ and Quinn TP (2002). Reconstructing recent divergence: evaluating nonequilibrium population structure in New Zealand Chinook salmon. *Mol Ecol* 11: 739–754

Kinnison, MT, Quinn, TP, and Unwin MJ. (2011). Correlated contemporary evolution of life history traits in New Zealand Chinook salmon, *Oncorhynchus tshawytscha*. *Heredity* 106:448-459.

Kinnison MT, Unwin MJ and Quinn TP (1998). Growth and salinity tolerance of juvenile Chinook salmon (*Oncorhynchus tshawytscha*) from two introduced New Zealand populations. *Can. J Zool* 76: 2219–2226.

Kinnison, MT, Unwin MJ and Quinn TP. (2008). Eco-evolutionary vs. habitat contributions to invasion in salmon: experimental evaluation in the wild. *Molecular Ecology* 17:405-414.

McDowall, RM (1990). New Zealand freshwater fishes: a natural history and guide. Heinemann Reed, Auckland.

McDowall R.M. (1994a). The origins of New Zealand's chinook salmon, *Oncorhynchus tshawytscha*. *Marine Fisheries Review*, 56: 1–7.

McDowall R.M. (1994b). Gamekeepers for the Nation: the Story of New Zealand's Acclimatisation Societies, 1861-1990. Canterbury University Press, Christchurch, New Zealand.

Quinn TP, Kinnison MT and Unwin MJ (2001). Evolution of Chinook salmon (*Oncorhynchus tshawytscha*) populations in New Zealand: pattern, rate, and process. *Genetica* 112-113: 493–513.

Quinn TP, Nielsen JL, Gan C, Unwin MJ, Wilmot R et al. (1996). Origin and genetic structure of chinook salmon (*Oncorhynchus tshawytscha*) transplanted from California to New Zealand: allozyme and mtDNA evidence. *Fish Bull* 94: 506–521.

Quinn TP and Unwin MJ (1993). Life history patterns of New Zealand Chinook salmon (*Oncorhynchus tshawytscha*) populations. *Can J Fish Aquat Sci* 50: 1414–1421.

23

Quinn TP, Unwin MJ and Kinnison MT (2000). Evolution of temporal isolation in the wild: genetic divergence in timing of migration and breeding by introduced chinook salmon populations. *Evolution* 54: 1372–1385.

Quinn, TP (2018). The behaviour and ecology of Pacific salmon and trout, Second Edition. University of Washington Press, Seattle. American Fisheries Society, Bethesda, Maryland.

Saint-Pé, K, Leitwein, M, Tissot L, Poulet, N, Guinand, B et al. (2019). Development of a large SNPs resource and a low-density SNP array for brown trout (*Salmo trutta*) population genetics. BMC Genomics 20(1): 582

Symonds, JE, Clarke, SM, King, N, Walker, SP, Blanchard, B et al. (2019). Developing Successful Breeding Programs for New Zealand Aquaculture: A Perspective on Progress and Future Genomic Opportunities. *Frontiers in Genetics* 10: 27

Unwin MJ and Gabrielsson RM (2018). Review of historical hatchery releases of Chinook salmon in New Zealand. Prepared for Contact Energy Ltd. Cawthron Report No. 3111. 39 p.

Unwin MJ, Kinnison MT, Boustead NC and Quinn TP (2003). Genetic control over survival in Pacific salmon (*Oncorhynchus* spp.): experimental evidence between and within populations of New Zealand Chinook salmon (*O. tshawytscha*). *Can J. Fish Aquat Sci* 60: 1–11.

Unwin MJ, Quinn TP, Kinnison MT and Boustead NC (2000). Divergence in juvenile growth and life history in two recently colonized and partially isolated Chinook salmon populations. *J Fish Biol* 57: 943–960.

West, IF and Goode, RH (1987). Aerial counts of spawning Chinook salmon (Oncorhynchus tshawytscha) on the Rakaia River system, Canterbury, New Zealand, 1973-76. *New Zealand Journal of Marine and Freshwater Research* 21: 563-572.

Weir, BS. (1996) Genetic Data Analysis II. Sunderland: Sinauer Associates Inc. 445 p.

Willis, D (2018). Recommendations for addressing declines in Chinook salmon abundance in New Zealand. Miscellaneous report prepared for Fish & Game New Zealand. 14 p.
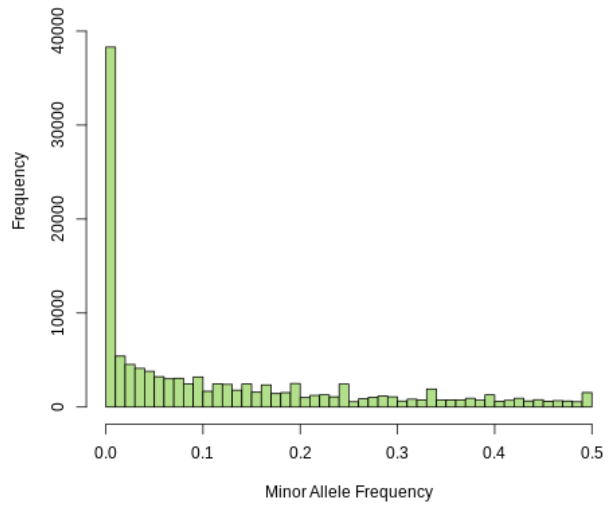
8. APPENDICES

Appendix 1. Minor allele frequency (MAF) distributions for each of the 5 sea-run (SR) populations, separate and combined in SR, and two landlocked populations.
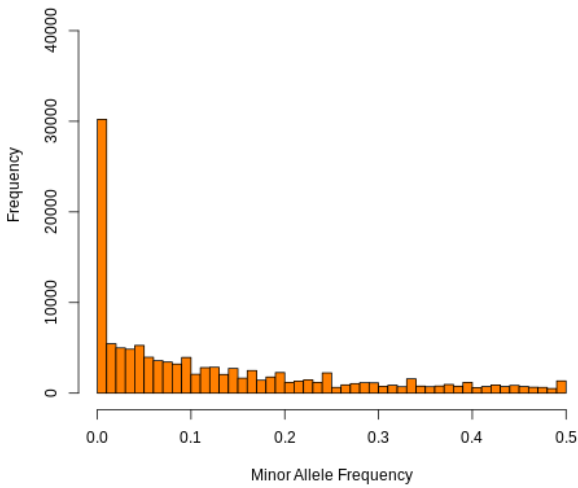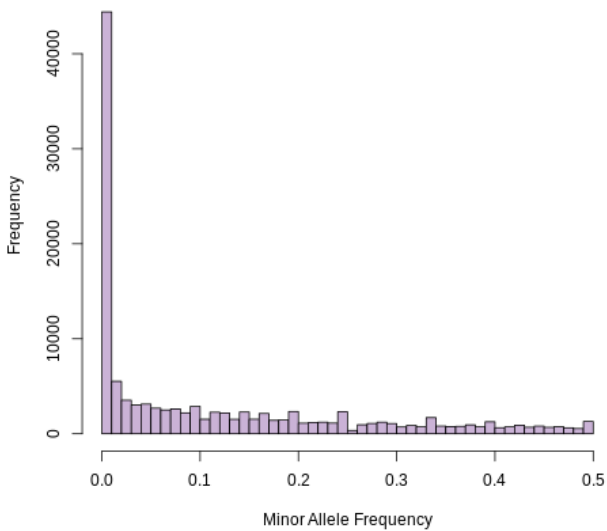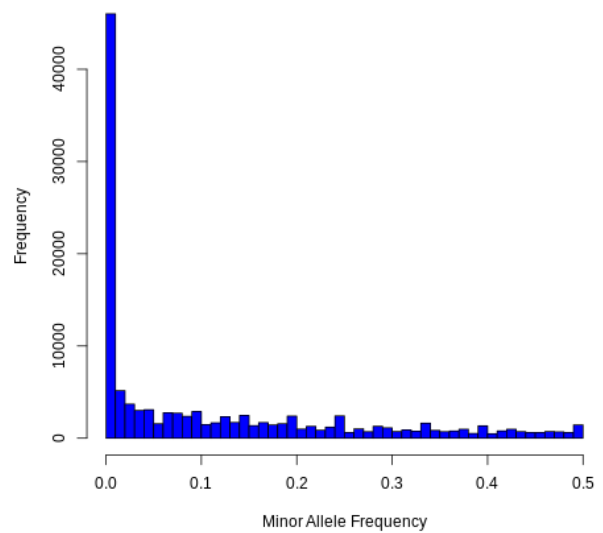
MAF for Rakaia River

MAF for Wairau River

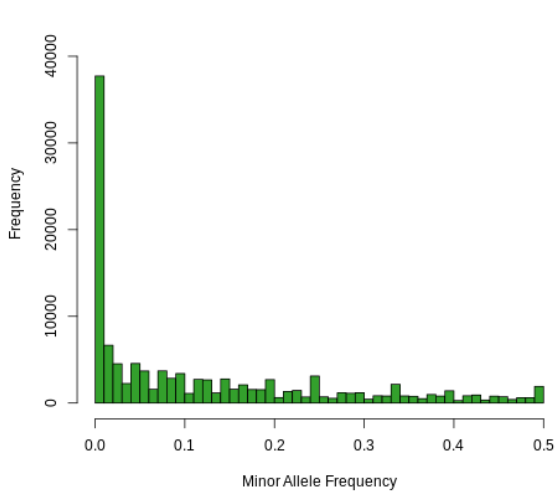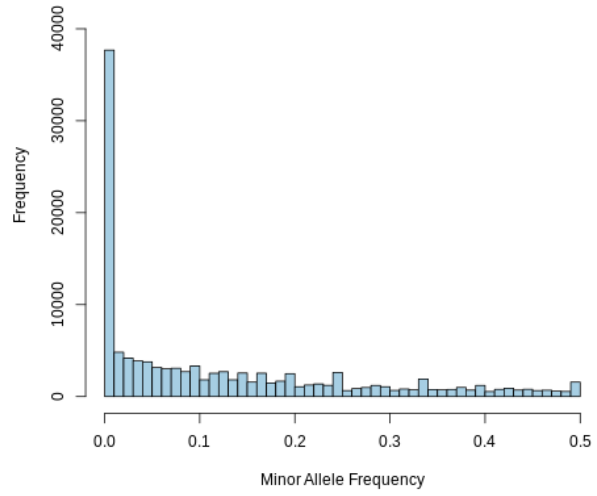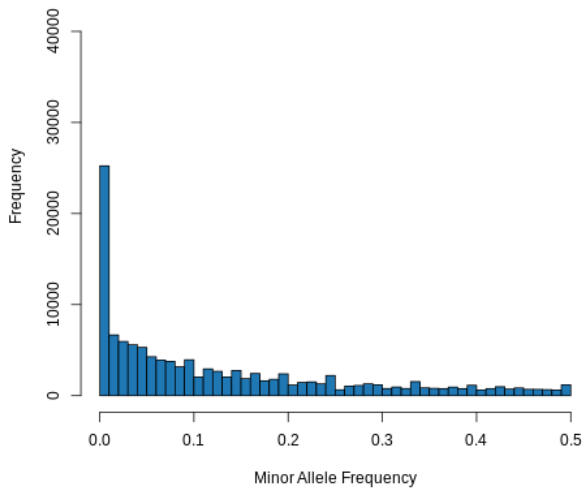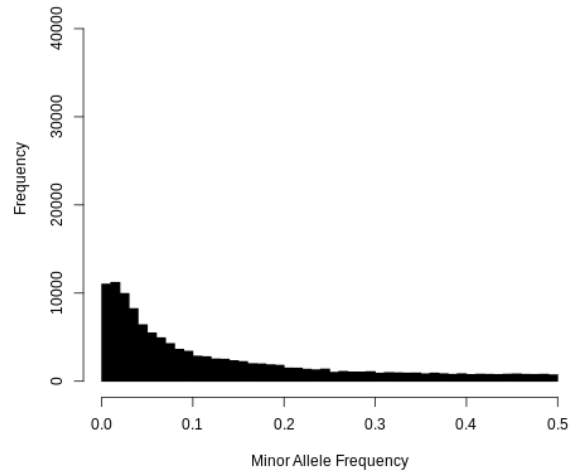**MAF for Waitaki River**

MAF for SR

Appendix 2.   Fixation index (fst values) for population subgroups.

Table A2.1. Pairwise mean Fst values between subgroups within the seven wild Chinook salmon populations.

Pairwise Fst Means	Lake Hawea	Lake Wakatipu	Paringa River	Paringa Windbag	Mellish Stream (Rakaia River, Lake Heron)	Upper Rakaia	Lower Rangitata	Upper Rangitata	Lower Wairau River	Upper Wairau River	Wairau Rainbow	Waitaki down stream	Waitaki up stream	Waitaki Hakataramea
Lake Hawea		0.036	0.028	0.030	0.029	0.024	0.026	0.027	0.017	0.021	0.029	0.023	0.023	0.023
Lake Wakatipu	0.036		0.028	0.029	0.029	0.024	0.025	0.025	0.018	0.022	0.029	0.023	0.023	0.023
Paringa River	0.028	0.028		0.010	0.024	0.021	0.018	0.017	0.022	0.023	0.019	0.019	0.019	0.019
Paringa Windbag	0.030	0.029	0.010		0.021	0.016	0.017	0.016	0.013	0.016	0.019	0.016	0.016	0.016
Mellish Stream (Rakaia River, Lake Heron)	0.029	0.029	0.024	0.021		0.015	0.015	0.014	0.015	0.017	0.017	0.017	0.017	0.016
Upper Rakaia	0.024	0.024	0.021	0.016	0.015		0.011	0.010	0.013	0.014	0.012	0.013	0.014	0.012
Lower Rangitata	0.026	0.025	0.018	0.017	0.015	0.011		0.009	0.011	0.012	0.013	0.013	0.014	0.012
Upper Rangitata	0.027	0.025	0.017	0.016	0.014	0.010	0.009		0.009	0.011	0.013	0.012	0.013	0.012
Lower Wairau River	0.017	0.018	0.022	0.013	0.015	0.013	0.011	0.009		0.017	0.007	0.014	0.013	0.013
Upper Wairau River	0.021	0.022	0.023	0.016	0.017	0.014	0.012	0.011	0.017		0.008	0.015	0.014	0.014
Wairau Rainbow	0.029	0.029	0.019	0.019	0.017	0.012	0.013	0.013	0.007	0.008		0.014	0.015	0.014
Waitaki Downstream	0.023	0.023	0.019	0.016	0.017	0.013	0.013	0.012	0.014	0.015	0.014		0.011	0.012
Waitaki Upstream	0.023	0.023	0.019	0.016	0.017	0.014	0.014	0.013	0.013	0.014	0.015	0.011		0.011
Waitaki Hakataramea	0.023	0.023	0.019	0.016	0.016	0.012	0.012	0.012	0.013	0.014	0.014	0.012	0.011	

Appendix 3.   Individual % inbreeding box plots by location within the main river systems.
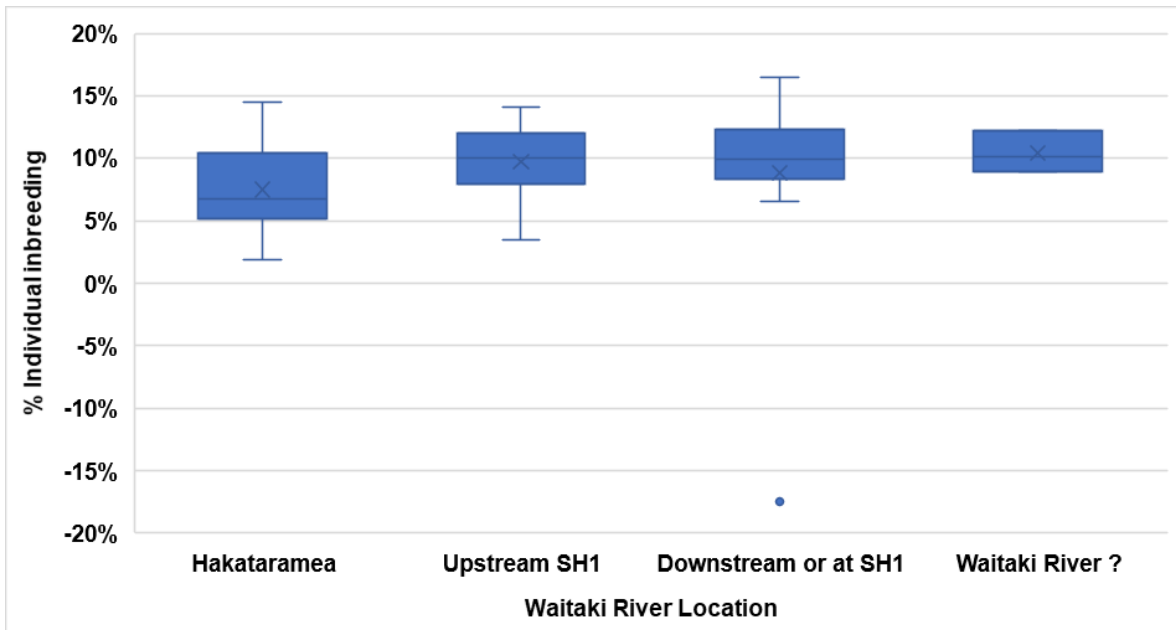
Figure A3.1a.   Box plot of individual inbreeding estimates in the Waitaki River by location. X = mean, median = line in the box, top and bottom lines of the box = 1st and 3rd quartiles. Minimum and maximum = vertical lines. Outliers = dots above and below the lines.
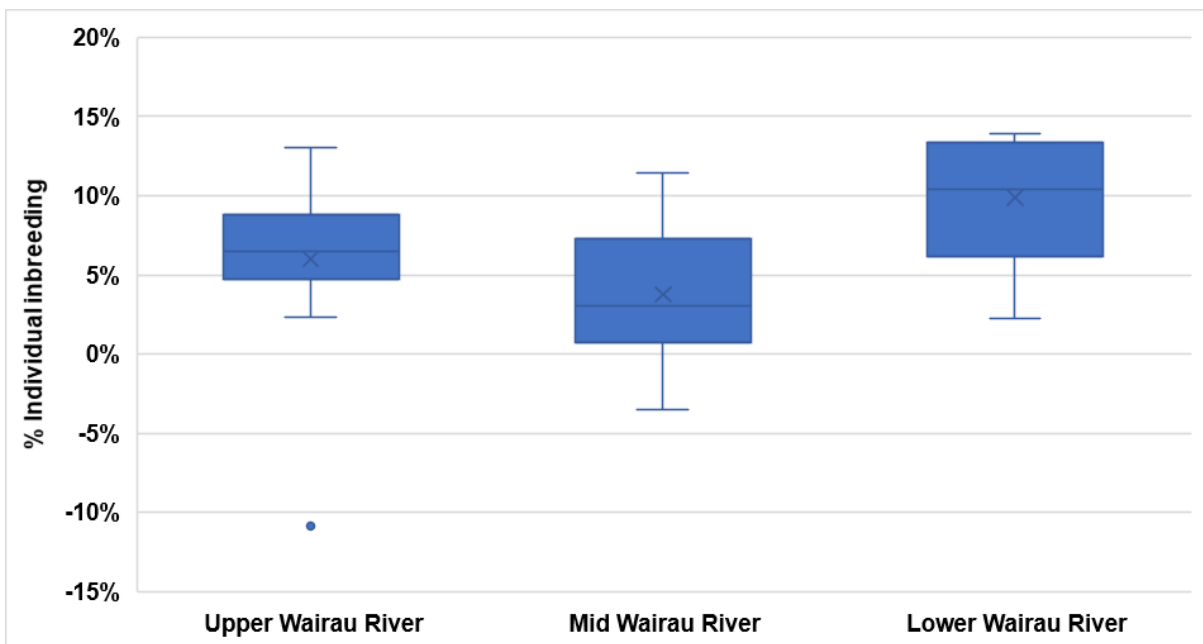
Figure A3.1b.   Box plot of individual inbreeding estimates in the Wairau River by location. X = mean, median = line in the box, top and bottom lines of the box = 1st and 3rd quartiles. Minimum and maximum = vertical lines. Outliers = dots above and below the lines.
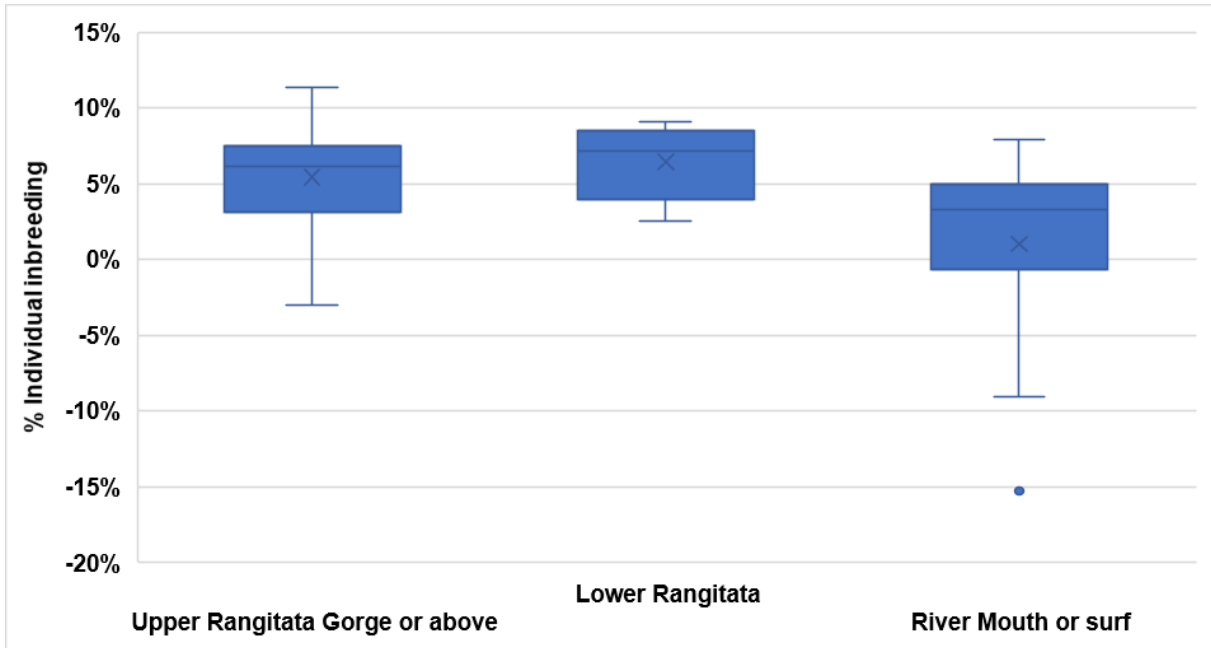
Figure A3.1c.   Box plot of individual inbreeding estimates in the Rangitata River by location. X = mean, median = line in the box, top and bottom lines of the box = 1st and 3rd quartiles. Minimum and maximum = vertical lines. Outliers = dots above and below the lines.

Figure A3.1d.   Box plot of individual inbreeding estimates in the Rakaia River by location. X = mean, median = line in the box, top and bottom lines of the box = 1st and 3rd quartiles. Minimum and maximum = vertical lines. Outliers = dots above and below the lines.

Appendix 4. Research poster presented by Monica Vallender at a University of Otago post graduate student research symposium, 16 September 2019, titled "Genetic diversity of wild Chinook Salmon (*Oncorhynchus tshawytscha*) in South Island, New Zealand".

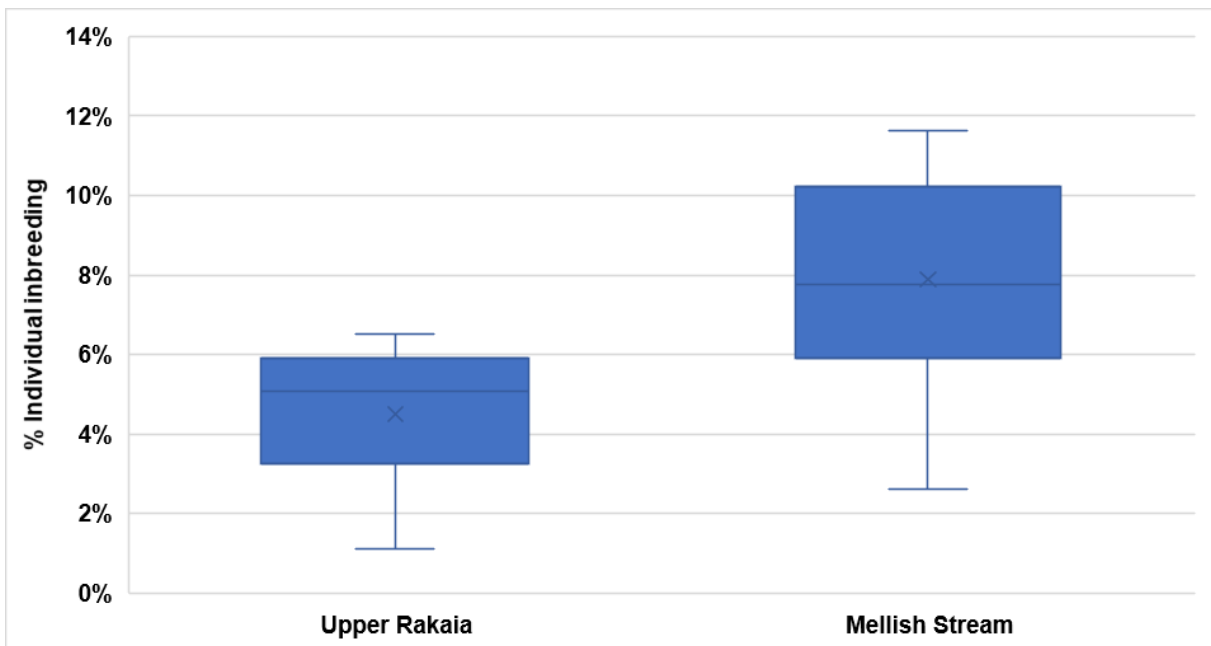# Genetic diversity of wild Chinook Salmon (*Oncorhynchus tshawytscha) in South Island, New Zealand*

Monica Vallender[1,3], Ken Dodds[1], Jane Symonds[2], Rasmus Gabrielsson[2], Rachael Ashby[1], Alana Alexander[2].

Supervisor's: *Neil Gemmell[3] & Shannon Clarke[1].*

[1]AgResearch, Dunedin, New Zealand, [2]Cawthron Institute, Nelson, New Zealand, [3] Department of Anatomy, University of Otago, Dunedin.

Introduction

- Since established in 1909 multiple sub-populations of Chinook have existed throughout the South Island with little genetic drift due to natal homing prior to spawning.[1]

- Historically, NZ Chinook salmon have been managed primarily as wild self-sustaining populations. Although extensive hatchery supplementation aimed at developing a commercial ocean ranching industry did occur at many locations, mainly between 1977 to 1997.[2]

- Salmonids display large interpopulation variation in life history traits, population structure, behaviour and other characteristics, in NZ evidence of phenotypic differences emerged in 1990s, indicating that detectible populations structure has likely evolved.[1]

- Fisheries management now attempts to be population-specific, reflecting the importance of retaining local adaptations. This has become increasingly important as spawning runs in all major Chinook salmon populations have been unusually poor since 2000.[3]

- Managing and maintaining populations sensitive to selection such as salmonids, has become more accessible and realistic since the development of reduced representation sequencing (RRS), a cost-effective approach which can be used without any prior knowledge of the genome.

Aim

- Assess the genetic diversity and structure of the wild NZ Chinook salmon population.

- In collaboration with Fish & Game NZ, investigate the utility of low depth Genotyping-By-Sequencing (GBS)[4] as a cost-effective method to inform stock management.

Methodology

- 420 samples have been collected from the south island of NZ using low-depth restriction enzyme GBS. Outlined in fig.1 (below).

Sequences
Low depth RE-GBS sequencing[4] of 420 samples, PstI-MspI enzyme across 5 lanes of Illumina HiSeq 2500.

Reference Genome: *Canadian O.tshawytscha NCBI's GenBank as accession GCA_002872995.1*[3]

Bioinformatics

De novo approach

UNEAK[14]
Min tag count 5, error tolerance rate 0.03, min.MAF 0.03, max.MAF 0.5, min. call rate 0.1

Reference-based approach

BWA-mem[13] reference alignment (mpileup)
Samtools[12] SNP caller
VCFtools[11] filter
Min. 2 indv. per SNP, remove all singletons + doubletons, min.depth 1 read per SNP.

KGD[4] filter
SNP depth >0.1, sample depth >0.3, HWD -0.05

400 samples remaining - 43,936 SNPs.
Mean sample depth 3.8. Callrate 0.65.

383 samples remaining - 120,811 SNPs .
Mean sample depth 3.6. Callrate 0.64.

Analysis

KGD[4], **KGD-Popgen**[4]
Relatedness statistics (individual, intra- and inter-), PCA, Fst & Pairwise Fst

GUS-LD[5]
Ne
100 SNPs per chromosome, Call rate filter 0.5

Adegenet[6]
DAPC, PCA & F-statistics

fastSTRUCTURE[8] & **NGSadmixture**[9]
Genetic composition of populations

Fastsimcoal2[7]
Demographic inference

ANGSD[10]
Fst

Completed   Future analysis

Figure 1, Outline of Methods. Shown above is the outline of methodology both completed and planned. Each solid rectangle represents a specific analysis undergone, entitled with the software name, use and results from that. Darker blue colour represents completed analysis and lighter blue planned analysis.

Results

Wairau River, Waitaki River, Paringa, Rakaia River, Ocean, Lower Clutha, Lake Mapourika, Rangitata River, Lake Hawea, Hurunui River, Waimakariri River, Lake Wakatipu
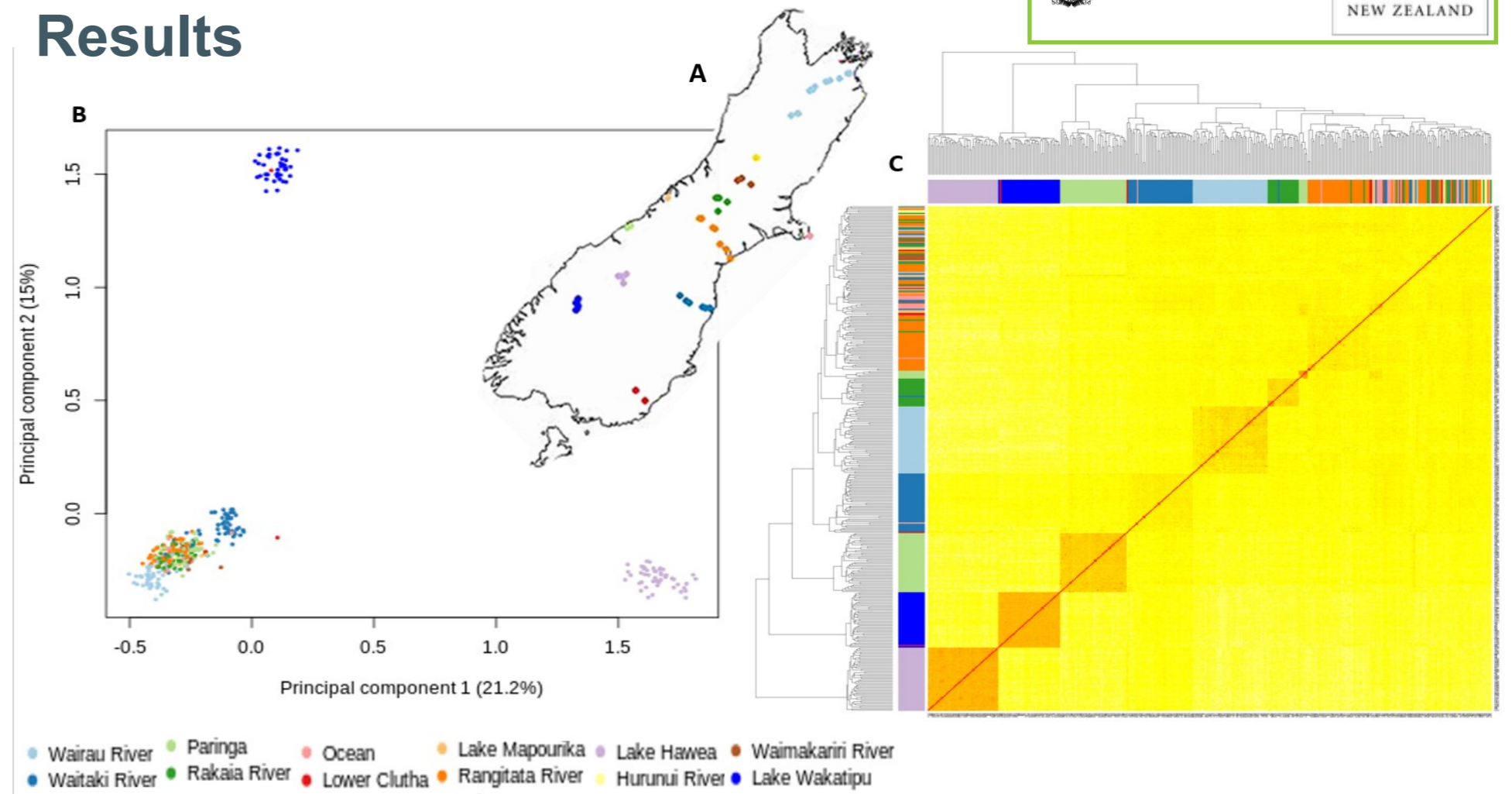
Figure 2: Twelve wild salmon populations across the South Island of New Zealand (A) . The principal component analysis (PCA; B) and genomic relationship matrix (GRM; C) coloured by geographic location (A). The PCA explains 21.2% of the variation in the first component and 15% in the second. For the GRM the darker (and redder) the colour in the matrix, the more related the individuals. Self relatedness is indicated on the diagonal and relatedness between individuals on the off diagonal. The colours in the sidebars of the GRM and the dots on the PCA refer to location of each identified population, colours and labels indicated in the legend in A and correspond to their relative coloured locations in B.
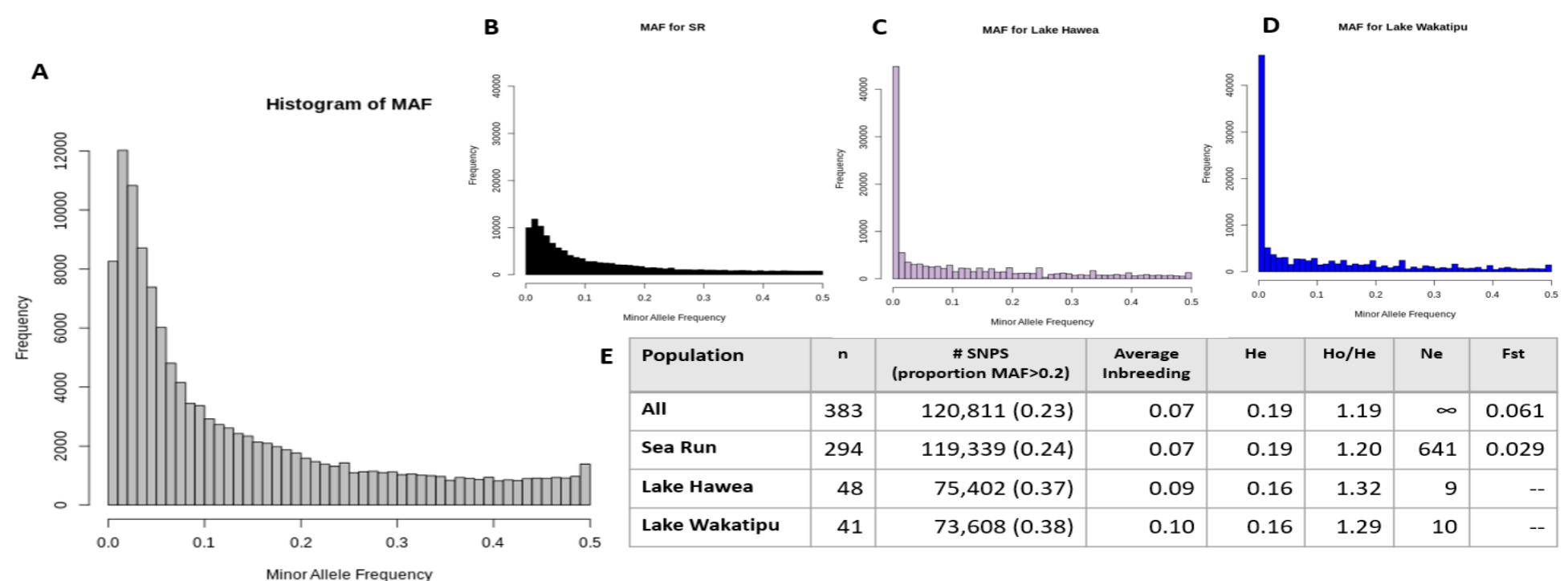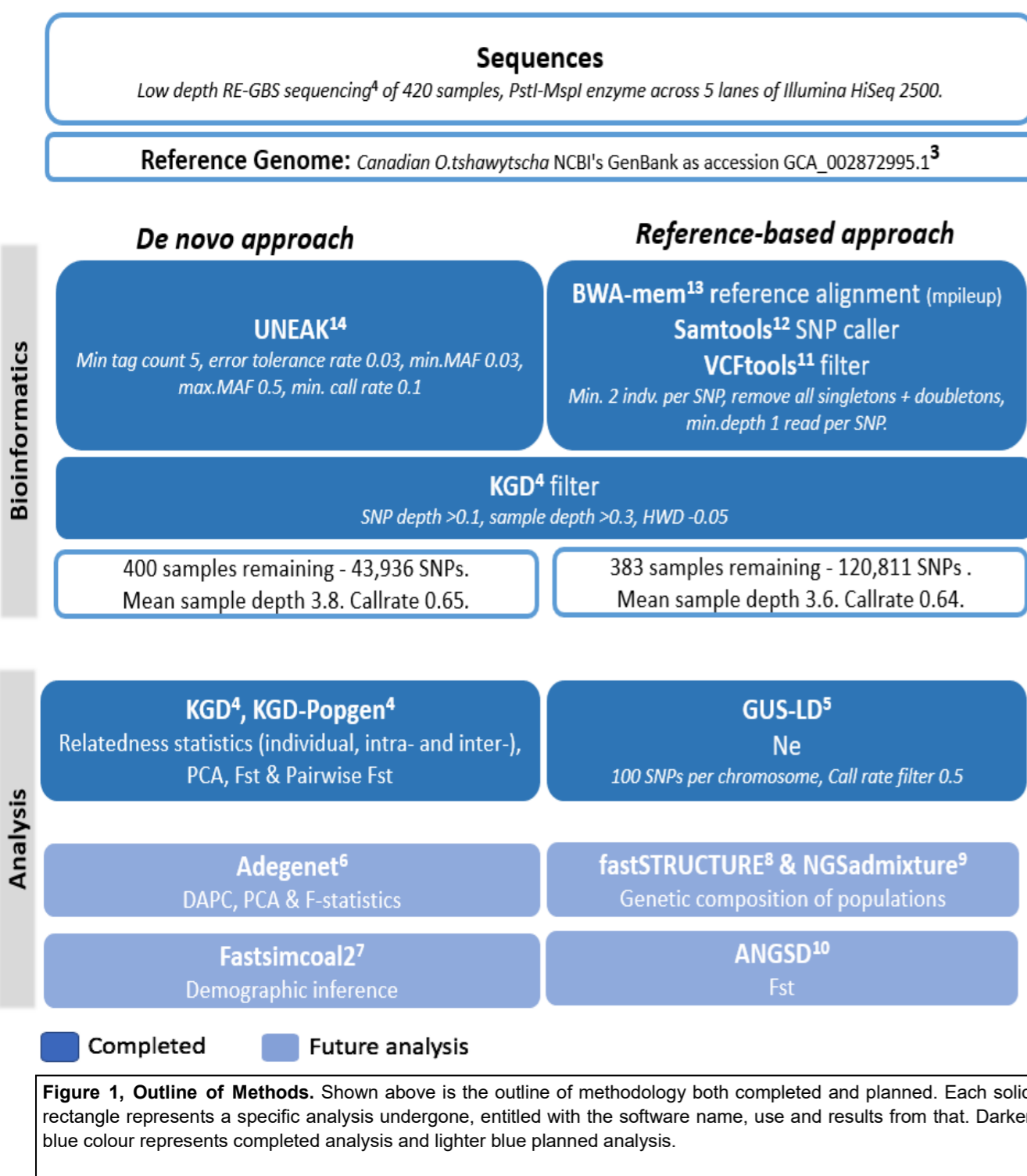

Population	n	# SNPS (proportion MAF>0.2)	Average Inbreeding	He	Ho/He	Ne	Fst
All	383	120,811 (0.23)	0.07	0.19	1.19	∞	0.061
Sea Run	294	119,339 (0.24)	0.07	0.19	1.20	641	0.029
Lake Hawea	48	75,402 (0.37)	0.09	0.16	1.32	9	--
Lake Wakatipu	41	73,608 (0.38)	0.10	0.16	1.29	10	--

Figure 3. Minor allele frequency distribution of combined (All) and by location type, all sea run populations (SR;B), and either land locked population Lake Hawea (C) and Lake Wakatipu (D). Note: The y-axis scale varies between these plots and a MAF of zero indicates a SNP is detected in at least one of the 12 populations but not present (i.e., had only one allele) in the specific group being analysed. A table of summary statistics (E) showing the number of SNPs (MAF > 0) and the proportion of those SNPS which have a MAF>0.2 (considered to be 'truly polymorphic') relative to number of samples (n) for each group. He is the expected heterozygosity for each of these groups. He/Ho is the ratio of expected heterozygosity to observed heterozygosity taking into account depth. Ne is the effective population size calculated using unlinked SNPs, representing the size of a population under the same migration and genetic drift pressures than the measured population and therefore can be indicative of current population size. Fst is the fixation index measure, showing the differences between groups based on conserved allele frequencies indicating that the more 'fixed SNPs' there are between populations, the more different they are likely to be .

Discussion

- This illustrates how the high numbers of SNPs produced by GBS can give a high-power approach to population genetics studies allowing a high quantity of results. Confirming the utility and value in future use of RRS in population genetics studies, providing an informative, cost-effective method of genetic analysis of any species.

- Three genetically distinct groups are reflected in the PCA and GRM which correlate to the respective lifestyle of each population. Supported by an Fst of 6% averaging across all populations and 3% comparing just either lake population to the remaining sea run populations, indicates the land locked populations are successfully breeding and have not diverged under the same influences (environmental or otherwise) than the sea run populations.

- Shallow population structure has been successfully observed between the sea run populations and can be partly explained by the geographical isolation and ocean currents between locations, and known inaccurate homing to close neighbouring rivers on the east cost. The statistical results in fig.3 show high fitness and diversity within the current population, but more research needs to be undertaken to be able to helpfully inform the management of more isolated populations.

Conclusion

Utilising RRS coupled with statistical analysis for low depth GBS has enabled detecting structure and diversity to a fine-scale in wild NZ Chinook salmon. Results confirm a strong potential to be a cost-effective method to inform the population genetics of non-model organisms, and to aid in salmon conservation and stock management.

References

1: Quinn et al (1996). Origin and genetic structure of chinook salmon, Oncorhynchus tshawytscha, transplanted from California to New Zealand: Allozyme and mtDNA evidence. Fishery Bulletin- National Oceanic and Atmospheric Administration. 94. 506-521.
2: Deans, NA; MJ Unwin and M Rodway (2004) Sport Fishery Management Chapter 41 in 'Freshwaters of New Zealand' Jointly published by the New Zealand Hydrological Society and New Zealand Limnological Society.
3: Christensen et al 2018: https://doi.org/10.1371/journal.pone.0195461
4: Dodds et al 2015 https://doi.org/10.1186/s12864-015-2252-3
5: Bilton et al 2018 doi: 10.1534/genetics.118.300831.

6: Jombart 2008:DOI: 10.1093/bioinformatics/btn129
7:Excoffier 2013: https://doi.org/10.1371/journal.pgen.1003905
8: Raj 2014: https://doi.org/10.1534/genetics.114.164350
9: Skotte 2013: DOI 10.1534/genetics.113.154138
10: Korneliussen 2014: DOI: 10.1186/s12859-014-0356-4
11:Danecek 2011:doi: 10.1093/bioinformatics/btr330
12: Li 2009 doi: 10.1093/bioinformatics/btp352
13: Li 2013 arXiv:1303.3997 [q-bio.GN]
14:Glaubitz 2014: https://doi.org/10.1371/journal.pone.0090346