A first investigation into the evolutionary
relationships, population structure and demographic
history of New Zealand Trevally (*Pseudocaranx
georgianus*) and its implications for fisheries
management

A thesis submitted to Victoria University of Wellington in partial fulfillment of the
requirement for the degree of Master of Science in Ecology and Biodiversity

Leah Kemp

Primary Supervisor: Peter Ritchie, Victoria University of Wellington

Secondary Supervisor: Maren Wellenreuther, Plant and Food Research

September 2019

Abstract

Pseudocaranx georgianus is a commercially important fishery in New Zealand. Currently, the management of this fishery assumes that Quota Management Areas comprise single biological stocks of a single species. However, little is known regarding the population structure of New Zealand *P. georgianus* and morphological data suggests that a cryptic *Pseudocaranx* species is included within these fisheries.

Whole genome sequence data was used to assemble and describe the first *P. georgianus* mitogenome. Primers were developed to produce the first genetic sequence data for New Zealand *P. georgianus*. The cytochrome c oxidase subunit I (COI) gene was sequenced for fourteen *P. georgianus* from New Zealand waters. These were compared phylogenetically with existing COI sequence data for *P. georgianus* from Australia and other *Pseudocaranx* species from a world-wide distribution. The hyper-variable control region of 304 *P. georgianus* sampled throughout New Zealand's North Island and 68 *P. georgianus* from three locations in Western Australia were also sequenced. These sequences were used to explore the population structure and demographic history of New Zealand *P. georgianus* using haplotype networks, AMOVA's, genetic diversity measures, Tajima's D, Fu's F and Bayesian migration analyses.

The *P. georgianus* mitogenome is typical of Cartilaginous fish species showing no major gene rearrangements, typical gene region lengths and stop and start codons. While assembling the *P. georgianus* mitogenome, this thesis demonstrates the importance of key methodological choices made when assembling mitogenomes from whole genome sequence data in silico in Geneious version 11.1. The choice of reference mitogenome has the largest influence on the quality of the assembly, impacting the annotation of the final mitogenome and the resolution of uncertain DNA regions. Increasing the number of mapping iterations increased the quality of the assembly but has a limited ability to mitigate the effects of using a poor reference mitogenome. Overall, I demonstrate the need to investigate and report the quality of published mitogenomes.

All *Pseudocaranx* species were monophyletic on the COI gene, supporting the current taxonomy of the *Pseudocaranx* complex. *P. georgianus* from Western Australia and New Zealand's North Island represent a monophyletic clade pending a taxonomic verification that two *Pseudocaranx dentex* sampled

in Australia are in fact *P. georgianus*.

No evidence was found to suggest that either of the New Zealand or Western Australian populations of *P. georgianus* are isolated by distance or clearly structured as distinct stocks. However, some populations of New Zealand *P. georgianus* were genetically distinct, including fish sampled from Raglan and the Bay of Plenty (F_{ST} of 0.02698 (p -value: 0.00901+-0.0091) as well as the North Cape and North Taranaki Bight (F_{ST} : 0.02698, p -value: 0.00901+-0.0091).

Some evidence was found to support the claim that *P. georgianus* along the west coast of New Zealand's North Island is structured and no evidence was found to refute the claim that fish from the Bay of Plenty are the same biological stock as fish from TRE2. Highly divergent control region sequences of fish sampled from Three Kings Islands and the Kermadec Islands suggest that these fish could be a species distinct from *P. georgianus*. Two genetically distinct populations of *P. georgianus* were identified in New Zealand's North Island and Western Australia (F_{ST} : 0.03517, p -value < 0.001), but further research would be required to determine if they are distinct species or populations. One juvenile population sampled in Whangarei had a high level of genetic connectivity with adult *P. georgianus* throughout New Zealand's North Island, likely reflecting the batch spawning and occasional long-distance migration behaviour of *P. georgianus*.

Negative Tajima's D and Fu's F statistics (D : -1.50612, p -value: 0.018; F : -23.54376, p -value: 0.011), unimodal mismatch distributions and skyline plots indicate that the New Zealand *P. georgianus* population has undergone a population expansion, possibly resulting from a geographic range expansion. The Western Australian population may also have undergone a population expansion (D : -1.27903, p -value: 0.086; F : -24.11497, p -value < 0.00001). However, a multimodal mismatch distribution (Harpending's Raggedness index: 0.00454591, p -value: 0.02) indicated that there is some stability in the size of this population.

This thesis is a first genetic investigation into New Zealand *P. georgianus* and has provided important biological insights into this species. Valuable information is revealed which will inform the management of New Zealand *P. georgianus* fisheries as inputs for stock assessment models. Additionally, several future research directions have been revealed which will further extend our knowledge of this taonga. For example, future genetic and taxonomic analyses may reveal a cryptic *Pseudocaranx* species occurring in the Three Kings and Kermadec Islands.

Acknowledgments

I truly wouldn't have been able to undertake, let alone complete this thesis without the help I've received from so many different people.

First and foremost, I would like to thank Richard White for constantly supporting me throughout this journey by providing snacks and walkies to the cafe for coffee. You've been a huge support.

Thank you to my supervisor Peter Ritchie for providing clarity, feedback, continued support as well as the freedom for independent thinking. Another thank you to my secondary supervisor, Maren Wellenreuther for the vital input and feedback on my initial thesis outline and my final thesis.

I would like to thank my parents, Kathryn Hughes and Ian Kemp for always requiring me to develop my own thoughts and teaching me perseverance and tenacity. I'm comforted to know that my mum would have known that I could complete a master's degree even though she wasn't able to see it for herself.

Thank you to my sisters and friends for always supporting me and understanding my several month disappearance into the thesis writing cave.

Another thank you is owed to my lab group, Tom Oosting, Amber Brooks, Yvan Papa and Alex Halliwell for keeping me sane throughout this process, mostly with Friday night beers and jokes at each other's expense that are only occasionally too spicy. On top of this, they guided me through many sticking points in the lab and during the data analysis.

A massive thank you to Moana, the Ministry for Primary Industries fisheries observers and Jennifer Chaplin for being an essential component of this research by providing all the trevally samples used in this thesis.

Figure 1: Screenshot of an R Shiny app displaying a few key results of this thesis. You can interact with a live version at: <https://leahkemp.shinyapps.io/trevallyapp/>.

Key R packages used within R version 3.6.1 to create this thesis and the accompanying app: bookdown (Xie, 2018a), dplyr (Wickham et al., 2019), ggplot2 (Wickham et al., 2018), kableExtra (Zhu, 2019), knitr (Xie, 2018b), leaflet (Cheng et al., 2018), leafletextras (Bhaska and Barret, 2018), mapview (Appelhans et al., 2019), networkD3 (Allaire et al., 2017), phylocanvas (Charlop-Powers, 2017), plotly (Sievert et al., 2018), rmarkdown (Allaire et al., 2018), sf (Pebesma, 2018), sp (Pebesma and Bivand, 2018) and tinytex (Xie, 2019).

See the welly-trevally git repository at <https://github.com/leahkemp/welly-trevally> for the code and documentation associated with this thesis. Additionally, the code for the R Shiny app associated with this thesis can be found at the trevallyApp git repository at <https://github.com/leahkemp/trevallyApp>.

Contents

Abstract	3
Acknowledgments	5
1 General Introduction	17
1.1 <i>Pseudocaranx georgianus</i>	17
1.1.1 Distribution	18
1.1.2 Movement and migration	19
1.1.3 Habitat preferences and reproductive behaviour	21
1.1.4 Sub-stocks and cryptic species	22
1.2 <i>P. georgianus</i> commercial fisheries	22
1.2.1 Stock status	25
1.3 Molecular genetics	27
1.3.1 Phylogenetics	28
1.3.2 Population genetics	28
1.4 Genetics and fisheries management	31
1.5 Research direction	32
2 Assembly and description of the <i>P. georgianus</i> mitogenome	35
2.1 Introduction	35
2.2 Materials and methods	36

2.2.1	<i>P. georgianus</i> mitogenome assembly	36
2.2.2	<i>P. georgianus</i> mitogenome description	38
2.3	Results	38
2.3.1	<i>P. georgianus</i> mitogenome assembly	38
2.3.2	<i>P. georgianus</i> mitogenome description	43
2.4	Discussion	52
2.4.1	Methods for mitogenome assembly	52
2.4.2	<i>P. georgianus</i> mitogenome	55
3	Field and laboratory work	57
3.1	Data collection	57
3.2	Tissue sampling	57
3.3	DNA extraction	60
3.4	Primer design and sequencing	60
3.4.1	COI gene	61
3.4.2	Control region	62
4	Evolutionary relationships of <i>Pseudocaranx</i> species	65
4.1	Introduction	65
4.1.1	Current <i>Pseudocaranx</i> taxonomy	65
4.1.2	<i>Pseudocaranx</i> in New Zealand	66
4.1.3	Historical reliance on morphological characters	66
4.1.4	Previous study	68
4.2	Materials and methods	68
4.3	Results	70
4.4	Discussion	74

5	Population structure and demographic history of <i>P. georgianus</i>	77
5.1	Introduction	77
5.1.1	Population structure of New Zealand <i>P. georgianus</i>	77
5.1.2	Population genetics	78
5.2	Materials and methods	79
5.2.1	Population structure	79
5.2.2	Demographic history	81
5.3	Results	82
5.3.1	Genetic data	82
5.3.2	Pairwise comparisons	83
5.3.3	Test for isolation by distance	84
5.3.4	Test for distinct populations	85
5.3.5	Demographic history	91
5.4	Discussion	100
5.4.1	Population structure and gene flow	100
5.4.2	Demographic history of <i>P. georgianus</i>	108
6	General discussion	111
6.1	Key findings	111
6.1.1	<i>P. georgianus</i> mitogenome	111
6.1.2	Evolutionary relationships of <i>Pseudocaranx</i> species and population genetics of <i>P. georgianus</i>	112
6.1.3	Population structuring of <i>P. georgianus</i> in New Zealand	113
6.2	Combining molecular genetics with morphological data	113
6.3	Population connectivity of <i>P. georgianus</i>	114
6.4	Putting research into practice	114
6.4.1	Inputs for fisheries stock assessment models	114

6.4.2	PCR primers for metabarcoding <i>Pseudocaranx</i> species	115
6.5	Inferences based on two mitochondrial markers	116
6.6	Future research directions	117
Appendix		118
A		119

List of Figures

1.1	Trevally (<i>Pseudocaranx georgianus</i>) illustration by Leah Kemp.	17
1.2	Annual distribution of <i>P. georgianus</i> in New Zealand by life stage as indicated in green.	23
1.3	Seasonal distribution of <i>P. georgianus</i> in New Zealand as indicated in green.	24
1.4	<i>P. georgianus</i> Quota Management Areas and yearly fishing effort for the three major commercially fished Quota Management Areas. The total allowable commercial catch (TACC) is indicated in red.	26
1.5	Three main types of genetic population structures (figure based on Laikre et al., 2005).	30
2.1	Identities of each nucleotide position of whole genome sequence data of one <i>P. georgianus</i> individual (<i>Broodstock trevally 1</i>) mapped to three reference mitogenomes.	40
2.2	Uncertainty in the consensus mitogenome resulting from assembling whole genome data of one <i>P. georgianus</i> individual (<i>Broodstock trevally 1</i>) to three reference mitogenomes.	40
2.3	Proportion of identical sites of the mitogenome assemblies of three <i>P. georgianus</i> individuals to three reference genomes related to the number of mapping iterations.	41
2.4	An example of the annotation conflicts among consensus mitogenomes associated with using three different reference mitogenomes.	42
2.5	The contig depths and percentage identities of whole genome data of one individual (<i>Broodstock trevally 1</i>) mapped to the <i>C. equula 2</i> reference mitogenome using different numbers of mapping iterations.	43
2.6	Unresolved regions along the mitogenome of one individual (<i>Broodstock trevally 1</i>) mapped to <i>C. equula 2</i> using different numbers of mapping iterations.	44

2.7	An example contig scaffold assembled in Geneious version 11.1.5 including nucleotide discrepancies within contigs.	45
2.8	A comparison of contig scaffolds from thirteen <i>P. georgianus</i> individuals mapped to <i>C. equula 2</i> including the range in contig depth and percentage identity at each nucleotide position.	46
2.9	The association between the average number of identical sites and the average pairwise identity along the length of the mitogenome with contig depth of whole genome data of thirteen <i>P. georgianus</i> individuals assembled to <i>C. equula 2</i>	46
2.10	The final annotated <i>P. georgianus</i> mitogenome.	48
2.11	Within species variation at each nucleotide position among thirteen <i>P. georgianus</i> whole mitogenomes (blue), 304 partial control region sequences (black) and 30 partial COI sequences (black) as well as conserved regions and parsimony informative sites.	52
3.1	All sampling locations of <i>P. georgianus</i>	58
4.1	COI primer binding locations and within species variation among thirty <i>P. georgianus</i> individuals from New Zealand (black) and thirteen broodstock <i>P. georgianus</i> (blue). . .	70
4.2	Maximum likelihood <i>Pseudocaranx</i> phylogeny including bootstrap support greater than 50.	72
4.3	Bayesian <i>Pseudocaranx</i> phylogeny including posterior probabilities greater than 0.5. . .	73
4.4	Sampling locations of the four major <i>Pseudocaranx</i> phylogenetic clades including <i>P. georgianus</i> (red), <i>P. dentex</i> (purple), <i>P. dinjerra</i> (yellow) and <i>P. wrighti</i> (green). . . .	74
5.1	Putative sub-stocks of <i>P. georgianus</i> in New Zealand.	80
5.2	The alignment positions of two <i>P. georgianus</i> alignments showing the level of within species variation based on thirteen broodstock <i>P. georgianus</i> (blue) and 304 <i>P. georgianus</i> sampled in New Zealand (black).	84
5.3	Statistical areas from which <i>P. georgianus</i> are genetically distinct (FST's are provided). . .	85
5.4	Relationship between geographic and genetic distance among <i>P. georgianus</i> from New Zealand and Western Australia.	86

5.5	Sampling locations of adult <i>P. georgianus</i> with the same haplotypes as juveniles caught in Whangarei (top) or haplotypes with one (middle) or two (bottom) single nucleotide differences.	87
5.6	Haplotype network and sampling locations of <i>P. georgianus</i> sampled from the Bay of Plenty and TRE2.	89
5.7	Haplotype network and sampling locations of <i>P. georgianus</i> from three putative substocks within TRE7.	90
5.8	Haplotype network and sampling locations of <i>P. georgianus</i> from Three Kings Islands and Kermadec Islands.	91
5.9	Haplotype network and sampling locations of Trevally from three Quota Management Areas in New Zealand.	92
5.10	Haplotype network and sampling locations of juvenile <i>P. georgianus</i> from Whangarei (New Zealand), adult <i>P. georgianus</i> from Western Australia and adult <i>P. georgianus</i> from New Zealand.	93
5.11	Haplotype network and sampling locations of <i>P. georgianus</i> from three locations in Western Australia.	94
5.12	Haplotype network and sampling locations of <i>P. georgianus</i> from New Zealand and Western Australia.	95
5.13	Demographic classes of New Zealand and Western Australian <i>P. georgianus</i> compared to several other New Zealand fishery species.	98
5.14	Mismatch distribution for both populations of <i>P. georgianus</i> including confidence intervals.	98
5.15	Skyline plots showing modelled changes in population sizes and migration rates over time of New Zealand and Western Australian <i>P. georgianus</i> populations including error bars (1.96 x approximate standard error).	99
A.1	Final <i>P. georgianus</i> mitochondrial genome sequence.	122
A.2	Final COI alignment.	123
A.3	Figure continued.	124
A.4	Figure continued.	125
A.5	Final control region alignment (83 of 304 sequences).	126

A.6	Figure continued.	127
A.7	Figure continued.	128
A.8	Final control region alignment of five Trevally sampled from Three Kings and Kermadec Islands.	129
A.9	Posterior probability distributions for all four Migrate model parameters.	130

List of Tables

1.1	Historical fishing effort by Quota Management Area (QMA).	27
2.1	Intergenic regions in the <i>P. georgianus</i> mitogenome.	47
2.2	Description of the genes in the <i>P. georgianus</i> mitogenome.	49
2.3	Description of the tRNA's and rRNA's in the <i>P. georgianus</i> mitogenome.	49
2.4	A comparison of the general features of the mitogenomes of <i>P. georgianus</i> and <i>C. equula</i>	50
2.5	Within species variation along the <i>P. georgianus</i> mitogenome.	51
2.6	Conserved regions in the <i>P. georgianus</i> mitogenome.	51
3.1	A description of the primers designed to target the COI gene of <i>P. georgianus</i>	61
3.2	Components of the PCR reaction used to amplify the partial COI gene of <i>P. georgianus</i>	61
3.3	A description of the primers designed to target the control region of <i>P. georgianus</i>	62
3.4	Components of the PCR reaction used to amplify the partial control region of <i>P. georgianus</i>	62
3.5	Summary information on all COI sequences that were acquired and sequenced for <i>Pseudocaranx</i> species.	63
3.6	Summary information on all <i>P. georgianus</i> samples for which the control region was sequenced.	63
4.1	Information on all COI sequences that were acquired or sequenced for <i>Pseudocaranx</i> species.	75
5.1	Parameters of the final three migration model runs.	83

5.2	Pairwise FST's of <i>P. georgianus</i> among fourteen statistical areas in New Zealand. FST's which are significant are highlighted in yellow.	84
5.3	AMOVA output comparing <i>P. georgianus</i> from TRE2 and Bay of Plenty.	88
5.4	AMOVA output comparing <i>P. georgianus</i> from three putative substocks within TRE7.	88
5.5	AMOVA output of three pairwise comparisons of three putative substocks of <i>P. georgianus</i> within TRE7.	88
5.6	AMOVA output comparing Trevally from three Quota Management Areas (QMA's) in New Zealand.	88
5.7	AMOVA output comparing juveniles from Whangarei to adults in New Zealand and adults in Western Australia.	96
5.8	AMOVA output comparing <i>P. georgianus</i> from three regions in Western Australia.	96
5.9	AMOVA output comparing <i>P. georgianus</i> from New Zealand and Western Australia.	96
5.10	Nucleotide and haplotype diversity of New Zealand and Western Australian populations of <i>P. georgianus</i> based on partial control region sequences.	97
5.11	Acceptance ratios for all parameters of the final migration model run.	97
5.12	MCMC autocorrelation and effective sample sizes for all parameters of the final migration model run.	100
5.13	Information on all <i>P. georgianus</i> samples for which the control region was sequenced.	101

Chapter 1

General Introduction

1.1 *Pseudocaranx georgianus*

The Carangidae family is a large and diverse family of teleost fish which are globally important for commercial and recreational fishing (Froese and Pauly, 2019). The family includes 30 genera of largely marine species such as trevallies, jacks, amberjacks, pompanos, scads, kingfish, pilotfish and queenfishes (Jaafar et al., 2012). The genus *Pseudocaranx* (Bleeker, 1863) includes the five species of trevallies (Dyer and Westneat, 2010; Smith-Vaniz and Jelks, 2006) that occupy an extensive range in temperate regions (Smith-Vaniz and Jelks, 2006). The greatest number of described *Pseudocaranx* species occur in Australasia, where four of the five known species occur (Smith-Vaniz and Jelks, 2006).

In New Zealand, the single species, *Pseudocaranx georgianus* (Cuvier, 1833) is known to occur (which

Figure 1.1: Trevally (*Pseudocaranx georgianus*) illustration by Leah Kemp.

also ranges into southern coastal Australian waters) and is distinguished from a similar species by meristic characters (fifteen caudal vertebrae) and morphological features including a blunt snout, a pronounced hump on the forehead and a large and diffuse opercular spot (Smith-Vaniz and Jelks, 2006). This species is more commonly known as silver trevally, or by the Māori name Araara. *P. georgianus* is a commercially important fishery in Australasia (Langley, 2004). In New Zealand, over 4000 tonnes are allocated annually for commercial fishing (Fisheries Science Group, 2018). The species is also considered to be of high recreational and customary significance with 100 tonnes and one tonne annually allocated for recreational and customary use, respectively (Fisheries Science Group, 2018). The importance of inshore fin fish such as *P. georgianus* are reflected in the National Fisheries Plan which states that “They contribute to our cultural and social traditions, to our nation’s economy and to our sense of overall well-being”.

P. georgianus is both a pelagic and demersal species (Fisheries Science Group, 2018) occurring to a depth of 150 meters (James, 1976). Juvenile *P. georgianus* have a typical embryonic development for fish (James, 1978) and occupy sheltered inshore areas for up to two years until they enter a demersal phase (Fisheries Science Group, 2018). Juvenile *P. georgianus* experience a moderate growth rate in the first few years of life, although the growth rate drops significantly thereafter (James, 1978). Adults live to beyond 40 years and commonly reach 4.5 kilograms and a fork length of 60 centimeters but are known to reach 6-8 kilograms (Fisheries Science Group, 2018). Female fecundity is low until females reach about 40 centimeters fork length (Fisheries Science Group, 2018). It has been reported that the growth rate of *P. georgianus* in New Zealand is faster and the maturation age later than for *P. georgianus* from Australia (Department of Agriculture and Water Resources, 2018).

1.1.1 Distribution

Adult *P. georgianus* are known to occur in inshore and offshore locations in the North and South Island of New Zealand (see Figure 1.2) as well as the southern half of Australia from New South Wales to Perth (Smith-Vaniz and Jelks, 2006). Based on data compiled by the National Institute of Water and Atmospheric Research (NIWA), the greatest density of fish occurs in coastal regions (National Institute of Water and Atmospheric Research, 2019). In the North Island, these high-density regions occur off Ninety Mile beach, off the east coast of North Cape, the Bay of Plenty, Gisborne, South Taranaki Bight and along most of the east coast of the North Island. In the South Island, a higher density of *P. georgianus* is found in Tasman Bay off Nelson and northern Karamaea Bight (see Figure 1.2).

The juvenile distribution of *P. georgianus* in New Zealand is limited to coastal regions, predominantly

around the North Island but reaching as far south as Bruce bay on the West Coast of the South Island and Kaikoura on the East Coast. The spawning distribution of *P. georgianus* is even further constrained to inshore regions of only the North Island (see Figure 1.2). There is limited information on *P. georgianus* spawning regions in the literature, although a putative *P. georgianus* spawning area has been reported in the 0-25m stratum in Urenui Bay extending north of Tirua point by a study that investigated snapper distribution and growth in the North Taranaki Bight (Horn, 1986). In this study, a large number of juvenile *P. georgianus* were caught in two bottom trawl surveys of the area. Since then, juvenile *P. georgianus* have been reported at numerous nursery-like locations around northern New Zealand (Francis et al., 2005) and eggs have been observed along the southern Queensland coast (Neira et al., 2015).

There is some variation in the range and distribution of high density regions of *P. georgianus* associated with season (see Figure 1.3). The high-density region of *P. georgianus* in the Bay of Plenty occurs in all seasons except for summer. In spring, a second region appears off Whakatane in the eastern most region of the Bay of Plenty. In Autumn, the range of the high-density region in the South Taranaki Bight increases and extends further south. The high-density region in the northern parts of Karamea Bight occurs predominantly in summer and over a smaller range in winter but is not present at all in spring and autumn. In summer, the high-density region in the North Taranaki Bight joins with the largest ranging high-density region along the east coast of the North Island over the North Cape and ranging into Great Exhibition Bay in East Northland. Otherwise these regions are separated in spring, autumn and winter. Seasonal changes in the distribution of juvenile and spawning *P. georgianus* in New Zealand has not been reported in the literature.

1.1.2 Movement and migration

An important component of population structure is movement and migration of individuals between populations (Laikre et al., 2005). There is limited information on the movements of New Zealand *P. georgianus*, however field research typically indicates that the species has surprisingly low levels of movement given the migratory nature of Carangidae species (James, 1980). A tagging experiment was undertaken on trawl caught *P. georgianus* in several locations in the Bay of Plenty and Hauraki Gulf. The 130 tagged fish that were recovered over a five year period indicated that *P. georgianus* in this region had limited movement, with 88% moving less than 55 kilometers from their release sites (James, 1980). The greatest distance observed was 246 kilometers from Great Barrier Island to Whakatane (James, 1980). Additionally, a single *P. georgianus* larval fish swam 82 kilometers in a study that forced larval reef fish to swim in a flume chamber without food or rest (Dudley et al., 2000). The literature has not described the levels of *P. georgianus* movement in other regions of New Zealand.

Recently, a tag-recapture study was undertaken on a substantial number of *P. georgianus* (over 6300) from 1007 kilometers of the east coast of Australia and over fourteen years (Fowler et al., 2018). The median movement from the release location to the recapture location was five kilometers and 88.5% of the individuals were recaptured within 50 kilometers of their release location with a time at liberty between one and 427 days. Several individuals were recaptured at their release location. The largest movement recorded was 508 kilometers. Importantly, these tag-recapture studies will likely not capture the extent of fish movements, since it only measures their location at two points in time and is unable to capture non-linear, back-tracing or short term movements.

Acoustic telemetry has been used to investigate the movements of twelve *P. georgianus* in Western Australia for up to one year. This study found that the home ranges of *P. georgianus* in Jurien Bay in Western Australia were varied in a diurnal manner being larger during the day (0.49 to 1.29 squared kilometers) and smaller during the night (0.37 to 1.61 squared kilometers) (Fairclough et al., 2011). The home ranges during the day and night were typically at 1-8 meters depth in reef and seagrass habitats, and occasionally sand or patchy seagrass habitats. One fish moved out of the detectable range which is estimated to be able to detect tags 282 kilometers away.

Further research is available on the movement patterns of close relatives of *P. georgianus*. A long-term monitoring study investigating the habitat usage of *P. dentex* in the Azores Islands in the Northern Atlantic Ocean captured movement patterns over three time scales (Afonso et al., 2009). Using active and passive acoustic tracking and tag-release capture methods, they observed differences in inshore and offshore fish. For three inshore fish, extensive daily movement was observed (between four and seven kilometers). Daily movement was more limited for three offshore fish (400 meters). Over seasonal time scales, inshore fish moved in a pattern similar to their daily movement, whereas offshore fish showed a strong seasonal migration to inshore or other offshore habitats associated with the summer spawning season. Tag-recapture of 58 fish showed long-term movements from the release locations between 100 meters and 52 kilometers. One individual swam at least 30 kilometers in 48 hours.

Carangidae species typically have a migratory nature (James, 1980) and research on *P. georgianus* and other *Pseudocaranx* species suggests that adult and juvenile *P. georgianus* have a strong swimming ability. Despite their apparent capacity for long distance movements, limited movement is often observed with infrequent long distance movements. This may be a result of both an underestimation of *P. georgianus* movements in tag-recapture studies as well habitat preferences associated with life stage and season. Information on the migration and movement patterns of *P. georgianus* provide an indication of their potential for long distance migrations and what level of genetic structure of *P. georgianus* populations is possible. However, it does not indicate whether long distance migrations do occur or whether the migrating individuals reproduce and contribute genetically to resident populations.

1.1.3 Habitat preferences and reproductive behaviour

It appears that factors such as ocean currents, habitat preferences and reproductive behaviour may have an important role in the distribution of adult, juvenile and spawning *P. georgianus*. *P. georgianus* appear to exhibit seasonal preferences in habitat use. Adult *P. georgianus* in New Zealand will move between pelagic and demersal habitats (Fisheries Science Group, 2018). They are thought to be batch spawners laying small batches of eggs over several weeks or months over the summer spawning season (Fisheries Science Group, 2018) as is the case with *P. dentex* (Afonso et al., 2009). *P. georgianus* have been reported by fishermen to move inshore in spring and offshore in summer, although it is unclear if this observation is due to seasonal differences in fishing patterns (Fisheries Science Group, 2018). These observations are supported by Afonso et al. (2008) and Afonso et al. (2009) which observed inshore and offshore populations of *P. dentex* undertake longer range migrations during the spawning season, aggregating around the summits of offshore reefs. Otherwise, these populations are thought to be segregated for most of the year (Afonso et al. (2008), Afonso et al. (2009)). Within a spawning season, a single inshore site was visited by most *P. dentex* individuals tracked in the study (Afonso et al., 2009). This batch spawning and seasonal migration behaviour is thought to increase mating opportunities for *P. dentex* (Afonso et al., 2009) and could do the same for *P. georgianus*.

Pseudocaranx species also exhibit habitat preferences. Juvenile *P. georgianus* are often found inshore in shallow estuaries and harbours in New Zealand (Fisheries Science Group, 2018) and Western Australia (Farmer et al., 2005). A tendency for schools of juvenile *P. dentex* to use sheltered inshore habitats is thought to be a result of feeding preferences and protection from predators (Afonso et al., 2008). This behavioural tendency to seek refuge was highlighted by a study by Castro et al. (1999) observing *P. dentex* behaviour using Fish Aggregation Devices in the Canary Islands (central-east Atlantic). Additionally, modelling of presence-absence data by Brooks (2019) indicates that bathymetry is an important determinant of *P. georgianus* occurrence in New Zealand.

The geographic distribution of *P. dentex* larvae off the coast of Brazil was suggested to be related to factors such as water movements and food availability (Campos et al., 2010). Furthermore, juveniles and sub-adults of *P. chilensis* were possibly transported by currents as larvae to a region over 800 kilometers from where they were previously known to occur (observed on Chanarel Island off the coast of Chile, known to occur on Desventuradas and Juan Fernandez Islands) (Bearez and Villaroel, 2018).

Genetic connectivity of individuals, populations or species can result from either common ancestry or an ongoing reproductive connection of populations. Therefore, to fully understand the taxonomy, population genetic structure and demographic history of *P. georgianus* and understand their relevance to *P. georgianus* biology and fisheries management, the analyses undertaken in this thesis will need to

interpreted in the light of *P. georgianus* movement and migration habits, reproductive behaviour and habitat.

1.1.4 Sub-stocks and cryptic species

There is evidence to suggest sub-stocks of *P. georgianus* and cryptic *Pseudocaranx* species are occurring within the *P. georgianus* fisheries in New Zealand. Time-series analysis of catch at-age data suggest that fish from East Northland to Hauraki Gulf and the Bay of Plenty represent sub-stocks of *P. georgianus* (Fisheries Science Group, 2018). Modelling of stock abundance (Fisheries Science Group, 2018) and observations in a tag-recapture study (James and Stephenson, 1974) suggest that fish from TRE2 and the Bay of Plenty (TRE1) are the same biological stock. Lastly, spatial heterogeneity of age composition and trends in catch per unit effort (CPUE) data suggest three sub-stocks are present within TRE7 on the west coast of New Zealand’s North Island: Ninety Mile Beach, Tauroa Point to North Taranaki Bight and South Taranaki Bight (Fisheries Science Group, 2018).

The taxonomy of the *Pseudocaranx* genus has been historically contentious, largely as a result of their similar morphologies and an underestimation of gene flow between geographically isolated populations (see James and Stephenson (1974) and Smith-Vaniz and Jelks (2006)). In New Zealand, morphological analyses by James and Stephenson (1974) and Smith-Vaniz and Jelks (2006) suggest a species distinct from *P. georgianus* exists in the Kermadec (Raoul) Island and the North Cape of New Zealand, however which species these represent remained inconclusive. Currently, *P. wrighti* is only known to occur in the southern half of Australia (Smith-Vaniz and Jelks (2006)). Importantly, mitochondrial DNA studies in Japan have found that morphologically similar *P. dentex* are genetically distinct (see Masuda et al. (1995) and Yamaoka1992). Furthermore, *Pseudocaranx* species in Australia often have overlapping intra-species variation in taxonomically distinguishing morphological features (see Froese and Pauly (2019), Smith-Vaniz and Jelks (2006)).

1.2 *P. georgianus* commercial fisheries

A rights-based, single species Quota Management System (QMS) was implemented in 1986 to address the legally binding goal to sustainably utilise New Zealand’s fisheries (Cryer et al., 2016). Under this system, a total allowable catch (TAC) is allocated for species stocks defined by Quota Management Areas (QMAs) (Cryer et al., 2016). Much of the TAC is allocated to commercial fishing as the total allowable commercial catch (TACC) with proportions also allocated to customary, recreational use and other fishing-related mortality (Cryer et al., 2016). *P. georgianus* was introduced into the QMS

Figure 1.2: Annual distribution of *P. georgianus* in New Zealand by life stage as indicated in green.

in 1986 and is managed under five quota management areas; TRE1, TRE2, TRE3, TRE7 and TRE10 (Fisheries Science Group, 2018) (see QMA boundaries in Figure 1.4).

The total allowable catch (TACC) of *P. georgianus* varies by QMA, with the current allowances being 1507 tonnes for TRE1, 241 tonnes for TRE2, 22 tonnes for TRE3, 2153 tonnes for TRE7 and 10 tonnes for TRE10. Fishing intensity is highest in TRE7 and TRE1 as shown in Figure 1.4. Landings in TRE1 experienced a general decline after 1986 when TACC limits were implemented. However, since 2014, fishing intensity in TRE1 have nearly reached levels experienced between 1983 and 1986. Fishing in TRE1 has not exceeded the TACC since its introduction. Fishing intensity in TRE2 is comparatively low and has been consistent for the last 30 years. However, landings in TRE2 have exceeded the TACC in 18 of the last 29 years, exceeding the yearly TACC by up to 127 tonnes. Landings in TRE3 are typically well below the TACC and often a yearly catch will be less than 1 tonne. The most intensive fishing occurs in TRE7 along the east coast of New Zealand's North Island. Fishing is often under the TACC, although it has been marginally exceeded in 3 of the last 29 years. Often no *P. georgianus* are taken from TRE10 and a small amount has been taken for only two of the past 29 years. Over the last 31 years, 84.80% of the TACC for all QMA's combined have been fished (see Table 1.1).

P. georgianus is primarily caught along the northern coasts of the North Island, but fish are also caught as far south as the north of the South Island (Fisheries Science Group, 2018). Trawl is the primary fishing method, although purse seine has increased to substantial quantities since the 1970's particularly in the Bay of Plenty. A modest amount of *P. georgianus* is also taken by set net (Fisheries Science Group, 2018).

P. georgianus is also an important commercial fishery in Australia and is fished within the Commonwealth Trawl and Scalefish Hook Sector on the south-eastern Australia (Australian Government

Figure 1.3: Seasonal distribution of *P. georgianus* in New Zealand as indicated in green.

Department of Fisheries, 2019a) and the Norfolk Island Fishery in the Norfolk Islands (Australian Government Department of Fisheries, 2019b). Fishing intensity was historically high in the Commonwealth Trawl and Scalefish Hook Sector, reaching nearly 1600 tonnes in 1990, but fishing intensity has showed a steady overall decrease to just below 55 tonnes in 2017-18 (Department of Agriculture and Water Resources, 2018). Commercial fishing is not permitted within the Norfolk Island Fishery and landings from charter and recreational fishing are comparatively low (Australian Fisheries Management Authority (2010); Australian Government Department of Fisheries (2019b)).

Management of New Zealand fisheries under the Quota Management system assumes that Quota Management Areas (QMA's) represent single biological stocks of a single species. Therefore, it is imperative to identify any sub-stocks or cryptic species currently included in *P. georgianus* fisheries to ensure the sustainable management of this Taonga. Currently, limited data is available to validate or reject claims of *P. georgianus* sub-structure and cryptic *Pseudocaranx* species in New Zealand waters. This highlights the need for genetic analysis of *Pseudocaranx* species in Australasia to resolve morphologically inconclusive species occurring in New Zealand *P. georgianus* fisheries.

1.2.1 Stock status

There is currently no accepted stock assessment of two of the three QMA's undergoing heavy commercial fishing effort (TRE1 and TRE2) but stock assessment for these areas are underway or planned as further data is collected (Fisheries Science Group, 2018). For the most commercially important fishery, TRE7, a stock assessment has been undertaken, but is based only on the core of the fishery from Tauroa point to North Taranaki Bight. Modelling in this stock assessment estimated that female median spawning biomass has declined by roughly half since the 1950's from over 20,000 metric tonnes to around 10,000 metric tonnes (Fisheries Science Group, 2018). This decrease is at times consistent with the rate of increase in the annual catch (Fisheries Science Group, 2018). Spawning biomass was comparatively stable during the 1990's and 2000's at roughly half the level observed in the 1950's, but with a slight recovery observed from 2010 (Fisheries Science Group, 2018). It is unknown how much *P. georgianus* is illegally caught or the impact of other sources of mortality such as the degradation of juvenile habitats (Fisheries Science Group, 2018). Although the New Zealand *P. georgianus* fisheries have and are currently undergoing heavy commercial fishing pressure, the fisheries are considered stable (Fisheries Science Group, 2018) although further data is required from a broader region within TRE7 and stock assessments are required for TRE1 and TRE3 in order to fully understand the status of New Zealand *P. georgianus* fisheries.

Figure 1.4: *P. georgianus* Quota Management Areas and yearly fishing effort for the three major commercially fished Quota Management Areas. The total allowable commercial catch (TACC) is indicated in red.

Table 1.1: Historical fishing effort by Quota Management Area (QMA).

Quota Management Area	Years above TACC	Percentage of TACC caught (1986-2017)	Percentage of TACC caught (2007-2017)
TRE1	0	77.71%	81.12%
TRE2	18	108.89%	105.10%
TRE3	1	11.99%	4.54%
TRE7	3	89.09%	86.16%
TRE10	0	0.64%	0.00%
TRE1, TRE2 and TRE7	-	85.92%	85.39%
All	-	86.01%	84.80%

1.3 Molecular genetics

Genetic studies based on mitochondrial DNA have been an important component of population genetic research for several decades (Ballard and Whitlock, 2004). Mitochondrial markers in genetic research provides distinct biological and practical advantages, causing such markers to be widely used in phylogenetics and population genetics from the 1990's (Ballard and Whitlock, 2004). One of the primary reasons for this interest is that mitochondria do not undergo recombination because of its maternal mode of inheritance (Awise et al., 1987). This provides a genetic sequence which theoretically has an accumulation of mutations as a result of genetic drift without the contribution of mutations from recombination. For phylogenetics and population genetics, this means that mitochondrial markers are considered largely neutral (Awise et al., 1987). However in recent years it has been recognised that natural selection can have some influence on mitochondrial DNA markers (Ballard and Whitlock, 2004).

The mitogenome also has a high mutation rate compared to nuclear markers, particularly the control region (Awise et al., 1987). This has allowed intraspecific analyses at the micro-evolutionary scale (Awise et al., 1987). Mitochondrial DNA markers are also easy to amplify through polymerase chain reaction (Ballard and Whitlock, 2004) making mitochondrial DNA studies more straightforward and cost effective than genomic studies. Overall, mitochondrial DNA markers present an opportunity to undertake a low risk 'first look' at a species of interest that has undergone little or no genetic research.

Today, we are transitioning into the genomics era, as producing whole genome data is becoming more affordable and more sophisticated statistical and computational software is developed (Casillas and Barbadilla, 2017). As we stand on the precipice of the genomics era, time and money constraints still encourage more traditional mitochondrial DNA studies, particularly as a first genetic investigation into a species for which little genetic research has been undertaken.

1.3.1 Phylogenetics

One of the most powerful statistical advances in population genetics over the last few decades was the introduction of the stochastic process model referred to as ‘the coalescent’ (Kingman, 2000). This has enabled researchers to account for randomness inherent in evolution (Rosenberg and Nordborg, 2002) and has produced a ‘tree thinking’ that has transformed many branches of biology (Yang and Rannala, 2012).

These theoretical developments combined with the increased availability of genetic data has progressed the field of phylogenetics from a field based on systematics and taxonomy to a field dominated by molecular genetics (Yang and Rannala, 2012). Constructing phylogenetic trees from genetic polymorphisms allow us to model the evolutionary history of organisms (Rosenberg and Nordborg, 2002) including the persistence of genetic lineages through time as well as species relationships and common ancestry (Yang and Rannala, 2012).

There are several methods of phylogenetic inference, presenting advantages and disadvantages which are discussed in more detail elsewhere (see Yang and Rannala (2012)). However, maximum likelihood and Bayesian inference methods provide advantages over distance matrix and maximum parsimony methods since they can model sequence evolution more realistically (Yang and Rannala, 2012) when appropriate model priors are used (Huelsenbeck and Rannala, 2004).

1.3.2 Population genetics

The field of molecular population genetics arose in the 1960s, born from a theoretical platform of population level concepts of evolution such as Mendelian inheritance concepts of changing allele frequencies in populations and the Darwinian concept that biological evolution of genetically distinct species arise from genetic variation in populations (Casillas and Barbadilla, 2017). For some time, there existed an extensive theoretical foundation for molecular population genetics, however recent technological advances have permitted the field of molecular population genetics to flourish and diversify (Casillas and Barbadilla, 2017). Molecular genetics now has a varied range of applications in ecology, evolution and conservation (Kress et al., 2015).

1.3.2.1 Population structure

There are three main types of genetic population structure, namely no differentiation (panmictic), continuous change (isolation by distance) and distinct populations (distinct stocks) (see Figure 1.5). Based on the null hypothesis of a panmictic population (no genetic differentiation) distinct populations

can be identified using an analysis of molecular variance (AMOVA) (Excoffier et al., 1992). This test compares the level of genetic variation among individuals of a population to the level of genetic variation between populations or sampling locations (Excoffier et al., 1992). The significance of any population differentiation is then established using permutational analyses (Excoffier et al., 1992).

A population structure with continuous change (isolation by distance) can be tested for by looking for a correlation between the pairwise level of genetic differentiation between individuals and the pairwise geographic distance (Meirmans et al., 2012). A positive correlation will result when dispersal distances are smaller than a species' range (Wright, 1943).

A panmictic population (no genetic differentiation) can be tentatively concluded when no evidence is found to suggest a population is structured as distinct stocks or isolated by distance. However, this is not direct evidence of a panmictic population, instead a lack of evidence to reject the null hypothesis of no population structuring.

Another method for describing the population structure of a species is the construction of haplotype networks. Haplotype networks are widely used to visualise the relationships within or between populations or species (Paradis, 2018). A number of methods for constructing haplotype networks exist, including parsimony, distances, maximum likelihood, Bayesian inference, split decomposition or consensus methods (see Holland et al. (2004); Paradis (2018)).

One distance based method of constructing a haplotype network is the minimum spanning haplotypes networks introduced by Bandelt et al. (1999). This extends on the minimum spanning tree algorithm developed by Kruskal (1956) by allowing an integration of many minimum spanning trees rather than providing one minimum spanning tree. This method combines features of the algorithm by Kruskal (1956) for finding minimum spanning trees and the maximum-parsimony heuristic algorithm by Farris (1970; Bandelt et al., 1999). Some methods of constructing haplotype networks will infer unobserved or ancestral haplotypes (Paradis, 2018). However the minimum spanning network does not, providing the benefit of minimising the total length of the network when dealing with large datasets (Bandelt et al., 1999).

Haplotype networks can have advantages over a phylogenetic analyses when it is suspected that there is an ancestral-descendance relationship among the haplotypes compared in the analysis, for example when using a DNA marker with a high mutation rate or when the samples may not all be contemporary (Paradis, 2018).

Figure 1.5: Three main types of genetic population structures (figure based on Laikre et al., 2005).

1.3.2.2 Demographic history

An excess of old mutations and/or a reduction of young mutations on a DNA region can result from several evolutionary forces, including natural selection, population bottlenecks or population expansions (Fu, 1996). Therefore, the presence of evolutionary events in evolutionary lineage can be tested for by testing for departures from the neutral model (Fu, 1996). Fittingly named, the neutral model assumes that a DNA marker is neutral and not undergoing selection (direct or indirect), recombination or that the population is undergoing a size expansion (Tajima, 1989).

Based on the assumption that the number of sites on the DNA sequence is large and the DNA sequences are randomly sampled from a population, departures from the neutral model can be tested for with the Tajima D statistic (Tajima, 1989). Both selection and a recent population bottleneck can produce a negative Tajima's D value, although the types of genetic variation that will result from these evolutionary forces will be different. Natural selection will result in the maintenance of insertions and deletions, whereas a population bottleneck or population expansion will affect all types of DNA polymorphisms, resulting in the maintenance of nucleotide diversity (Tajima, 1989) and therefore haplotype diversity.

The Fu's F statistic is another statistical test developed more recently to test for departures from the neutral model (Fu, 1996). It has the advantage of being more powerful than the Tajima's D statistic at identifying selection or population expansions because it is more powerful for detecting departures from the neutral model (Fu, 1996).

While Tajima's D and Fu's F statistics can indicate whether or not a populations have undergone a

size expansion, mismatch distributions can estimate the size of populations over time. This is because changes in population size leaves characteristic signals in the distribution of nucleotide differences between pairs of individuals (Rogers and Harpending, 1992; Rogers et al., 1996). Therefore comparing the observed distribution and expected distribution of pairwise nucleotide differences can indicate whether a population has experienced a bottleneck, undergone an size expansion (Rogers and Harpending, 1992) or large spatial expansion (Ray et al., 2003).

The levels of migration between populations can be estimated from genetic data by estimating effective population sizes and immigration rates scaled by mutation (Beerli and Felsenstein, 1999), described in this thesis as a ‘migration analysis’. This Bayesian approach integrates all possible relationships of *P. georgianus* based on the coalescent theory (Kingman, 1982a,b, 2000) and migration (Hudson, 1990; Nath and Griffiths, 1993; Nogueira et al., 1990).

1.4 Genetics and fisheries management

Understanding the population structure and taxonomy of fisheries species is an important basis on which fisheries are managed (Laikre et al., 2005; Waples et al., 2008). This understanding allows implementation of fisheries management techniques that reduce the risk of genetic depletion by ensure fishing is undertaken on genetically homogeneous groups (Laikre et al., 2005). Maintaining genetic diversity is crucial to the sustainability of fisheries (Laikre et al., 2005). This is because genetic diversity is the basis of a species ability to adapt to environmental changes and complex interacting pressures (Allendorf and Ryman, 2002). It can also reduce the risk of inbreeding depression which can lead to a reduction in productivity (Allendorf and Ryman, 2002). Biologically informed fisheries management areas (ie. stocks) reduces the chance of over-exploitation of local populations which can result in local extinction or local reductions in genetic diversity (Laikre et al., 2005).

Genetic tools are a comparatively cheap way to understand a species population structure, taxonomic status and genetic diversity, and in turn define fisheries management stocks (Bernatchez et al., 2017). It provides base-level measures that enable the sustainable utilisation of fishery resources (Bernatchez et al., 2017).

Like other New Zealand fisheries, the *P. georgianus* fishery is managed based on total allowable catch of *P. georgianus* stocks. Therefore, the first step in the successful management of this fishery is to understand what *Pseudocaranx* species are occurring in New Zealand’s *P. georgianus* fisheries. The next step is to understand the stock structure of New Zealand *P. georgianus*. Stock structure findings could have important implications for fisheries management by informing the inputs of the stock assessment

model used as a basis for managing this fishery. Understanding the taxonomy and population structure of *P. georgianus* will indicate whether a single or multiple population model is appropriate for New Zealand *P. georgianus* fisheries.

These questions can be addressed with genetic data. Phylogenetic analysis of a DNA marker with a species-level rate of variation can indicate which *Pseudocaranx* species are occurring in New Zealand. Population genetic analyses of a highly variable DNA marker provides fine-scale information on the population structure of *P. georgianus*. Exploring the demographic history of genetically distinct populations identified through population genetic analyses will provide further biological information regarding the population's history and its response to historical events. In turn, this can provide insights into how *P. georgianus* populations will respond to future environmental pressures.

Although some genetic data has been produced for *P. georgianus* sampled from Australian waters, no genetic data is currently available in either GenBank or BOLD systems for New Zealand specimens of *P. georgianus*. Only one population genetic study has been carried out on a *Pseudocaranx* species (see Bearham (2004)) and no such studies have been undertaken on *P. georgianus* in New Zealand.

1.5 Research direction

This thesis will investigate the taxonomy, population structure, population connectivity and the demographic history of *P. georgianus* in New Zealand based on two mitochondrial DNA markers, the COI gene and the control region. *P. georgianus* from Australia will act primarily as a reference population for investigations of New Zealand *P. georgianus*. Other *Pseudocaranx* species will also be investigated to understand the taxonomic standing of *P. georgianus*. First, the mitogenome of *P. georgianus* will be assembled, primarily to act as a practical basis for producing genetic data by allowing primer design and identification of variable and conserved regions in the mitogenome. This research fits within a wider effort that seeks to broaden the application of genomic tools to inform the management and understanding of aquatic organisms (Bernatchez et al., 2017). The specific research questions are:

1. What is the general structure and composition of the *P. georgianus* mitogenome and how does it compare to other fish species?
2. What is the taxonomic status of *P. georgianus* in New Zealand?
3. What is the population structure of *P. georgianus* in New Zealand?
4. What is the level of population connectivity of *P. georgianus* within Australasia?
5. What is the demographic history of New Zealand *P. georgianus*?

6. What are the implications of these findings for the management of New Zealand *P. georgianus* fisheries?
7. How do these findings compare to the findings of other New Zealand fisheries species?

Chapter 2

Assembly and description of the *P. georgianus* mitogenome

2.1 Introduction

The general structure of the mitochondrial genome is typically well conserved among ray-finned and cartilaginous fish and gene rearrangements are uncommon (Sato et al., 2016). These mitogenomes are typically composed of thirteen protein coding genes, twenty-two tRNA's, two rRNA's, one noncoding control region (Aise et al., 1987) and intergenic regions of variable sizes, similar to other vertebrates (Sato et al., 2016). The start and stop codons for protein coding genes are variable, although there are common gene-specific codons among ray-finned and cartilaginous fish (Sato et al., 2016). The most variable structure of fish mitogenomes is the gene length (Sato et al., 2016).

Complete mitogenome sequence data has traditionally been obtained using direct Sanger sequencing and/or mitochondrial DNA re-sequencing (MitoChip) (Ye et al., 2014). With the advent of high-throughput DNA sequencing (e.g. Illumina systems), mitogenome sequences can be obtained by sequencing mitochondrial DNA enriched from nuclear DNA or indirectly through microarray hybridization and long-range PCR-based sequencing (Hahn et al., 2013; Ye et al., 2014). Mitogenome sequences can also be produced indirectly by isolating the mitochondrial sequences from whole genome sequence data in silico. This is done by mapping sequence reads to a reference mitogenome (Ye et al., 2014). For example, this technique has been used to produce the blue cod (*Parapercis colias*) mitogenome (Smith, 2012).

Complete annotated mitogenomes can be used to design species-specific PCR primers which reduces

the occurrence of failed or non-specific primer binding which can be an issue with cross-species designed primers. Annotated sequence data can also be used to verify that primers are successfully amplifying the target region. If there are several mitogenomes available for a species of interest or a group of closely related species, conserved and variable regions can be more easily identified. This has the wider implication of reducing the cost and time spent establishing and optimising PCR protocols and therefore improving the success of population genetic studies.

The mitogenomes of several fish species from a wide range of families have been reported and scores of unpublished mitogenomes are also regularly being uploaded to NCBI GenBank producing an extensive genetic database. A quick search for mitogenomes of ray-finned fish in GenBank returns 12,647 results. Yet for many New Zealand fish species, such as the *Pseudocaranx* genus, there is currently little or no complete mitogenome sequence data available. Constructing the mitogenome of *P. georgianus* is an important first step in the broader aim to study the taxonomy, population structure and demographic history of *P. georgianus* and the wider genus.

2.2 Materials and methods

2.2.1 *P. georgianus* mitogenome assembly

In this chapter, the *Pseudocaranx georgianus* mitogenome is isolated and assembled from whole genome sequence data in Geneious version 11.1.5 (Kearse et al., 2012). Quality checked whole-genome sequence data for thirteen broodstock *P. georgianus* was provided by Plant and Food Research (PFR) in Nelson, New Zealand (referred to as *Broodstock trevally 1* to *Broodstock trevally 13* in this chapter). This data was produced by the Australian Genome Research Facility Ltd. (AGRF) using whole genome Illumina paired-end sequencing on three HiSeq lanes. Between 27 to 36 million reads 125 base pairs in length were produced for each individual. No information was provided on the average depth of coverage of the genomes. The annotated genome is described and compared to a closely related species and the typical features of the mitogenomes of ray-finned and cartilaginous fish. The effect that several methodological choices had on the quality of the contig alignments and consensus *P. georgianus* mitogenomes are also explored.

BLAST searches and a literature search were used to establish species that are closely related to *P. georgianus* for which complete annotated mitogenome is available. The COI gene of a *P. georgianus* individual sampled from Victoria, Australia (GenBank accession: EF609442) was blasted against a nucleotide collection database using a discontinuous megablast that handles more dissimilar sequences like inter-species variation (Biomatters, 2018). This BLAST search queried the National Center for

Biotechnology Information (NCBI) within Geneious version 11.1.5 (Kearse et al., 2012) and returned a list of genetically similar sequences and their ranked pairwise identity scores.

A reference mitogenome that was returned with a high pairwise identity score to *P. georgianus* at the COI gene (*Carangoides equula*, GenBank accession: KM201334) was used to assemble a preliminary *P. georgianus* mitogenome. From this *P. georgianus* mitogenome, species specific primers were designed to target the COI gene of *P. georgianus* in Primer3 version 2.3.7 (Untergasser et al., 2012) within Geneious version 11.1.5 (Kearse et al., 2012). These primers were used to PCR amplify and sequence the COI gene of a few initial New Zealand *P. georgianus* samples caught off the coast of Raglan on the west coast of New Zealand's North Island. One of these initial COI sequences was used in a secondary discontinuous megaBLAST search to identify reference mitogenomes that were closely related to *P. georgianus* from New Zealand waters.

Whole genome sequence data of all thirteen *P. georgianus* individuals were imported into Geneious version 11.1.5 (Kearse et al., 2012) using FASTQ sequence import. During import, the paired reads were set using the appropriate parameters (read technology: Illumina, pair type: paired end - inward pointing, insert size: 125). For all mitogenome assemblies, the whole genome sequence reads were mapped to a reference mitogenome using Geneious mapper, with medium-low sensitivity and no trimming before mapping. From the resulting contig scaffolds, the consensus sequences were annotated by transferring the annotations of the reference genome to the *P. georgianus* consensus mitogenome sequence.

To investigate the effect that the mitogenome assembly methods had on the quality of the mitogenome assemblies and the consensus mitogenomes, the whole genome data of all thirteen individuals was assembled to three closely related reference mitogenomes identified in BLAST and literature searches (*C. equula 1*, *C. equula 2* and *T. japonicus*). Whole genome data of three individuals (*Broodstock trevally 1*, *Broodstock trevally 2* and *Broodstock trevally 3*) was also mapped to one reference genome (*C. equula 2*) using 2, 3, 5, 10, 25, 1000, 10000 and 100000 iterations. Lastly, any differences among individuals mapped to the same reference genome (*C. equula 2*) with the same number of mapping iterations (1000) was compared. The structure and quality of the contig scaffolds and consensus mitogenomes were compared by identifying any differences in their general structure, annotation length or placement, unresolved regions, percentage identities, contig depth and number of identical sites. Uncertainty was calculated by subtracting the pairwise identity at each nucleotide position from one. All pairwise alignments of consensus mitogenomes were created using the Mauve genome alignment plugin (version 1.1.1) (Darling et al., 2010) implemented in Geneious version 11.1.5 (Kearse et al., 2012)

2.2.2 *P. georgianus* mitogenome description

A final *P. georgianus* mitogenome was produced using the best methods identified in section 2.2.1; one *P. georgianus* was mapped to *C. equula* 2 (1000 mapping iterations). The *P. georgianus* mitogenome was then described and compared to the mitogenomes of other ray-finned and cartilaginous fish described by Satoh et al. (2016), as well as to the closely related species *Carangoides equula* (Whitefin Trevally, GenBank accession: NC_025644) (a different individual from that used to assemble the *P. georgianus* mitogenome). The general features of the whole mitogenome structure as well as the protein coding genes, non-coding regions, tRNA's, rRNA's and intergenic regions were identified. The location of the stop and start codons for each gene were identified using typical start and stop codons present in other fish described by Satoh et al. (2016). The level of within-species variation, conserved regions and parsimony informative sites were found using a consensus of mitogenome sequences from all thirteen Broodstock trevally individuals (mapped to *C. equula* 2 with 1000 mapping iterations). The first half of the control region was mapped to sequence data from 304 *P. georgianus* individuals sampled from New Zealand (see Section 2.2.1). Within-species variation was quantified by aligning the data and subtracting the percentage identity at each nucleotide position from 1. Conserved regions and parsimony informative sites were identified in DnaSP version 6.12.3 (Rozas et al., 2017). The final presentation of the *P. georgianus* mitogenome was created using the R Shiny application created by Yu et al. (2018).

2.3 Results

2.3.1 *P. georgianus* mitogenome assembly

The mitogenome of a *Carangoides equula* (Whitefin Trevally) individual sampled from Victoria, Australia (GenBank accession: KX373635, herein referred to as *C. equula* 1) was returned as the mitogenome with the highest pairwise identity to the COI gene (92.1%) of *P. georgianus* in the first BLAST search. This species was also found to be closely related to the *Pseudocaranx* genus in a phylogenetic study of the Carangidae family (Damerou et al., 2018).

A second BLAST search returned a mitogenome of another *C. equula* individual (GenBank accession: KM201334, herein referred to as *C. equula* 2) with a high pairwise identity to the reference (92.4%). The mitogenome with the second highest pairwise identity was a *Trachurus japonicus* individual (GenBank accession: AP003092, herein referred to as *T. japonicus*) with a pairwise identity of 89.7%.

Comparison of reference genomes

When whole genome sequence data of the same *P. georgianus* individual (*Broodstock trevally 1*) was assembled to three different reference mitogenomes, most nucleotide positions had a percentage identity to the reference between 0.9 and 1.0 (see Figure 2.1). When *C. equula 2* and *T. japonicus* were used as reference genomes, the percentage identities ranged between 0.59 and 1.0. However, when *C. equula 1* was used as a reference genome, some nucleotide positions had a low percentage identity to the reference (percentage identity ranged between 0.25 and 1.0). These nucleotide positions with low percentage identities to the reference arose primarily from four poorly resolved regions along the mitogenome. These included one region in 16S rRNA region, one region in the ATP8 gene and two regions in the D-loop (see Figure 2.2). All four of these regions were uncertain when *C. equula 1* was used as a reference genome. When assembled to the *C. equula 2* and *T. japonicus* reference genomes, the unresolved regions in the 16S rRNA region and the first region in the control region were resolved, although the ATP8 region and the second D-loop region remained uncertain (see Figure 2.2). The number of nucleotide positions that are identical to the reference genome (identical sites) was lowest when *C. equula 1* was used as a reference genome and highest when *C. equula 2* was used as a reference (see Figure 2.3).

The overall placement of the annotations on the consensus mitogenomes were congruent among individuals mapped to the same reference genome. However, there was some variation in the placement of annotations by up to thirteen base pairs (See Figure 2.4). Of the three reference genomes, the whole genome sequences assembled equally well to *C. equula 2* and *T. japonicus*. *C. equula 1* performed the poorest as a reference mitogenome, despite being considered taxonomically more closely related to the *Pseudocaranx* genus than *T. japonicus*.

Number of mapping iterations

Increasing the number of mapping iterations used in the mitogenome assemblies reduced the frequency of regions with no coverage, low contig depth and nucleotide positions with a low percentage identity to the reference genome (see Figure 2.5). However, a saturation point appears to be reached where increasing the number of mapping iterations provided no further improvement to lower range contig depths, percentage identities values (see Figure 2.5) or some unresolved regions (see Figure 2.6). This saturation point occurred at a different number of mapping iterations depending on which reference mitogenome was used for the assembly. The highest number of mapping iterations were required to reach this saturation point when the data was assembled to the *C. equula 1* reference genome (around ten mapping iterations). The saturation point for the contig depth was reached after only three

Figure 2.1: Identities of each nucleotide position of whole genome sequence data of one *P. georgianus* individual (*Broodstock trevally 1*) mapped to three reference mitogenomes.

Figure 2.2: Uncertainty in the consensus mitogenome resulting from assembling whole genome data of one *P. georgianus* individual (*Broodstock trevally 1*) to three reference mitogenomes.

Figure 2.3: Proportion of identical sites of the mitogenome assemblies of three *P. georgianus* individuals to three reference genomes related to the number of mapping iterations.

mapping iterations to the *C. equula 2* reference genome, and a saturation for percentage identities was reached after five mapping iterations. For the *T. japonicus* reference genome, the saturation point was reached after five mapping iterations. Once these saturation points were reached, comparatively low contig depth and percentage identities values remained when mapped to the *C. equula 1* reference genome.

There were six regions of uncertainty along the length of the mitogenome (represented as peaks in uncertainty in Figure 2.6) including two regions in the ND2 gene, one region in the ATP6 gene, one region in the ND6 gene and two regions in the control region. Some of these uncertain regions were resolved by increasing the number of mapping iterations (see Figure 2.6), however three of the six regions (one region in the ND2 gene, one region in the ATP6 gene and one region in the control region) were not fully resolved by increasing the number of mapping iterations.

Surprisingly, the number of nucleotide positions that were identical to the reference mitogenome (identical sites) tended to decrease as the number of mapping iterations increased until around five to ten mapping iterations. At this point, usually no additional reads were mapped to the contig scaffold (see Figure 2.3). This may be due the presence of nucleotide discrepancies within contigs (see Figure 2.7) or mitochondrial heteroplasmy being included in the contig scaffold in higher frequency as more contigs are assembled to the reference genome.

Comparison of individuals

The contig depth was highly variable among *P. georgianus* individuals mapped to the *C. equula 2* reference genome, ranging from 67 and 175 (*Broodstock trevally 13*) to 307 and 540 (*Broodstock trevally 5*). In contrast, the percentage identity of the contigs to the reference was consistent among individuals

Figure 2.4: An example of the annotation conflicts among consensus mitogenomes associated with using three different reference mitogenomes.

Figure 2.5: The contig depths and percentage identities of whole genome data of one individual (*Broodstock trevally 1*) mapped to the *C. equula 2* reference mitogenome using different numbers of mapping iterations.

(see Figure 2.8). As the average contig depth increased, the average number of identical sites in the consensus genomes decreased (ranging from 67.2% and 34.8%) (see Figure 2.9). In comparison, the average pairwise identity of the consensus genomes remained around 99% for all average contig depths.

2.3.2 *P. georgianus* mitogenome description

The description of the *P. georgianus* mitogenome is based on the most robust method identified in Section 2.3.1 (*Broodstock trevally 1* mapped to *C. equula 2*, 1,000 mapping iterations). In this assembly, 18,611 of 36,114,914 whole genome sequence reads were assembled to *C. equula 2*. Four mapping iterations were performed before no more whole genome sequences were aligned to the assembly. The consensus mitogenome presented here represents the light strand (L-strand) with a lower proportion of heavier A and G nucleotides (44.6%) than are found on the complimentary strand (55.4%).

Overall structure

The *P. georgianus* mitogenome is 16,578 base pairs in length with a base composition of 27.6% A, 29.7% C, 16.9% G and 25.7% T (see Figure 2.10). The *P. georgianus* mitogenome is slightly more AT rich than the average cartilaginous fish (28.3% A, 28.7% C, 16.6% G and 26.5% T (Sato et al., 2016))

Figure 2.6: Unresolved regions along the mitogenome of one individual (*Broodstock trevally 1*) mapped to *C. equula 2* using different numbers of mapping iterations.

Figure 2.7: An example contig scaffold assembled in Geneious version 11.1.5 including nucleotide discrepancies within contigs.

Figure 2.8: A comparison of contig scaffolds from thirteen *P. georgianus* individuals mapped to *C. equula 2* including the range in contig depth and percentage identity at each nucleotide position.

Figure 2.9: The association between the average number of identical sites and the average pairwise identity along the length of the mitogenome with contig depth of whole genome data of thirteen *P. georgianus* individuals assembled to *C. equula 2*.

and the *C. equula* mitogenome (26.3% A, 30.2% C, 18.1% G, 25.3% T). The *P. georgianus* mitogenome is organised in a typical fashion with no gene rearrangements compared to the mitogenomes of fish and other vertebrates described by (Satoh et al., 2016). The genome consisted of thirteen protein coding genes, twenty-two tRNA's, two rRNA's and one non-coding region (D-loop/control region). There were ten intergenic regions that ranged from one to thirty-six base pairs in length (see Table 2.1) with the largest intergenic region bordering tRNA-Asn and tRNA-Cys. (See Figure A.1 in the Appendix for the full DNA sequence of the *P. georgianus* mitogenome).

Table 2.1: Intergenic regions in the *P. georgianus* mitogenome.

Length	Bordering regions
5	ND2 gene and tRNA-Trp
1	tRNA-Ala and tRNA-Asn
36	tRNA-Asn and tRNA-Cys
1	tRNA-Tyr and COX1 gene
3	tRNA-Ser (1) and tRNA-Asp
7	tRNA-Asp and COX2 gene
1	tRNA-Lys and ATP8 gene
1	tRNA-Arg and ND4L gene
4	tRNA-Ser (2) and tRNA-Leu (2)
4	tRNA-Glu and Cytb gene

Genes

There are thirteen protein coding genes encoded on the *P. georgianus* mitogenome (see Table 2.2). A handful of these genes overlap with neighbouring genes or tRNA's such as ATP8 with ATP6 and ND4 with ND5 (see Table 2.2). The percentage of GC content for a given gene ranges between 43.3% for COX2 and 50.4% for COX3. Two types of start codons were used for protein coding genes, with ATG being the most common start codon in the *P. georgianus* mitogenome as it is for many other fish species (Satoh et al., 2016). Four types of stop codons were present and roughly half of the protein coding genes had incomplete stop codons either one or two base pairs in length. These are likely to be completed with the addition of a poly A tail during RNA processing (Ojala et al., 1981).

tRNA's and rRNA's

The *P. georgianus* mitogenome also encodes two rRNA's on the light strand and twenty-two tRNA's in the light and heavy strands, including two duplicate tRNA's (Ser, Leu) (see Table 2.3). Three pairs of tRNA's overlap (Ile with Gln, Gln with Met and Thr with Pro). The GC content of the tRNA's and rRNA's range from 30.4% for tRNA-Arg to 50.7% for tRNA-Ile. The length of the tRNA's are roughly similar (68-75 base-pairs).

Figure 2.10: The final annotated *P. georgianus* mitogenome.

Table 2.2: Description of the genes in the *P. georgianus* mitogenome.

Gene	Start position	Stop position	Length	Direction	Strand encoded	Overlapping with neighbour	Start codon	Stop codon	GC content
ND1	2885	3859	975	Forward	Light	No	ATG	TAA	49.4%
ND2	4073	5112	1040	Forward	Light	No	ATG	T - -	49.2%
COX1	5507	7057	1551	Forward	Light	N	GTG	TAA	45.7%
COX2	7210	7900	691	Forward	Light	N	ATG	T - -	43.3%
ATP8	7977	8141	165	Forward	Light	Yes	ATG	TAG	46.1%
ATP6	8135	8818	684	Forward	Light	Yes	ATG	TAA	47.1%
COX3	8818	9602	785	Forward	Light	Yes	ATG	TA -	50.4%
ND3	9673	10021	349	Forward	Light	No	ATG	T - -	46.1%
ND4L	10092	10388	297	Forward	Light	Yes	ATG	TAA	51.5%
ND4	10382	11762	1381	Forward	Light	Yes	ATG	T - -	47.4%
ND5	11980	13818	1839	Forward	Light	Yes	ATG	TAA	45.6%
ND6	14336	13815	522	Reverse	Heavy	Yes	ATG	TAA	48.5%
CYTB	14410	15550	1141	Forward	Light	No	ATG	T - -	49.3%

Table 2.3: Description of the tRNA's and rRNA's in the *P. georgianus* mitogenome.

	Start position	Stop position	Length	Direction	Strand encoded	Overlapping with neighbour	GC content
rRNA's							
12S	69	1019	951	Forward	Light	No	47.3%
16S	1092	2810	1719	Forward	Light	No	45.5%
tRNA's							
Phe	1	68	68	Forward	Light	No	50.0%
Val	1020	1091	72	Forward	Light	No	48.6%
Leu (1)	2811	2884	74	Forward	Light	No	50.0%
Ile	3860	3934	75	Forward	Light	Yes	50.7%
Gln	4004	3934	71	Reverse	Heavy	Yes	40.8%
Met	4004	4072	69	Forward	Light	Yes	49.3%
Trp	5118	5188	71	Forward	Light	N	49.3%
Ala	5258	5189	70	Reverse	Heavy	N	40.0%
Asn	5332	5260	73	Reverse	Heavy	N	49.3%
Cys	5435	5369	67	Reverse	Heavy	N	43.3%
Tyr	5505	5436	70	Reverse	Heavy	N	47.1%
Ser (1)	7128	7058	71	Reverse	Heavy	N	47.9%
Asp	7132	7202	71	Forward	Light	N	43.7%
Lys	7901	7975	75	Forward	Light	N	42.7%
Gly	9603	9672	70	Forward	Light	No	37.1%
Arg	10022	10090	69	Forward	Light	No	30.4%
His	11763	11834	72	Forward	Light	No	40.3%
Ser (2)	11835	11902	68	Forward	Light	No	51.5%
Leu (2)	11907	11979	73	Forward	Light	No	45.2%
Glu	14405	14337	69	Reverse	Heavy	No	42.0%
Thr	15551	15622	72	Forward	Light	Yes	51.4%
Pro	15692	15622	71	Reverse	Heavy	Yes	40.8%

Comparison to close relatives

The DNA sequence and gene arrangements of the *P. georgianus* mitogenome are generally well conserved with its close relative (*C. equula*). They have similar overall lengths, start and stop codons and GC contents of genes, tRNA's and rRNA's, but there are some minor differences. For protein-coding genes, *P. georgianus* and *C. equula* share deviations from the most common lengths including a slightly

shorter length ATP8 gene and a slightly longer ATP6 gene and ND2 gene. Start codons for protein-coding genes are typically conserved between *P. georgianus* and *C. equula* although *P. georgianus* does not share the *C. equula* novel start codon for the ATP6 gene (GTG) described by Zou and Li (2016).

Table 2.4: A comparison of the general features of the mitogenomes of *P. georgianus* and *C. equula*.

DNA region	Length			Start codons				Stop codons			GC content	
	P. georgianus	C. equula	Most common	P. georgianus	C. equula	Most common		P. georgianus	C. equula	Most common	P. georgianus	C. equula
Phe	68	68	-	-	-	-	-	-	-	-	50.0%	45.6%
12S	951	950	-	-	-	-	-	-	-	-	47.3%	47.4%
Val	72	72	-	-	-	-	-	-	-	-	48.6%	50.0%
16S	1719	1719	-	-	-	-	-	-	-	-	45.5%	47.0%
Leu (1)	74	74	-	-	-	-	-	-	-	-	50.0%	51.4%
ND1	975	975	975	ATG	ATG	ATG	TAA	TAA	TAA	-	49.4%	52.8%
Ile	75	75	-	-	-	-	-	-	-	-	50.7%	54.7%
Gln	71	71	-	-	-	-	-	-	-	-	40.8%	40.8%
Met	69	69	-	-	-	-	-	-	-	-	49.3%	44.9%
ND2	1040	1047	1046	ATG	ATG	ATG	T - -	TAG	TA -	-	49.2%	53.4%
Trp	71	71	-	-	-	-	-	-	-	-	49.3%	49.3%
Ala	70	70	-	-	-	-	-	-	-	-	40.0%	40.0%
Asn	73	73	-	-	-	-	-	-	-	-	49.3%	50.7%
Cys	67	67	-	-	-	-	-	-	-	-	43.3%	43.3%
Tyr	70	70	-	-	-	-	-	-	-	-	47.1%	45.7%
COX1	1551	1551	1551	GTG	GTG	GTG	TAA	TAA	TAA	-	45.7%	46.3%
Ser (1)	71	71	-	-	-	-	-	-	-	-	47.9%	47.9%
Asp	71	71	-	-	-	-	-	-	-	-	43.7%	43.7%
COX2	691	691	691	ATG	ATG	ATG	T - -	T - -	T - -	-	43.3%	45.3%
Lys	75	74	-	-	-	-	-	-	-	-	42.7%	43.2%
ATP8	165	165	168	ATG	ATG	ATG	TAG	TAA	TAA	-	46.1%	50.9%
ATP6	684	684	683	ATG	GTG	ATG	TAA	TAA	TA -	-	47.1%	48.5%
COX3	785	785	785	ATG	ATG	ATG	TA -	TAA	TA -	-	50.4%	49.0%
Gly	70	70	-	-	-	-	-	-	-	-	37.1%	38.6%
ND3	349	349	349	ATG	ATG	ATG	T - -	TAG	T - -	-	46.1%	48.4%
Arg	69	69	-	-	-	-	-	-	-	-	30.4%	36.2%
ND4L	297	297	297	ATG	ATG	ATG	TAA	TAA	TAA	-	51.5%	55.2%
ND4	1381	1381	1381	ATG	ATG	ATG	T - -	T - -	T - -	-	47.4%	49.9%
His	72	72	-	-	-	-	-	-	-	-	40.3%	41.7%
Ser (2)	68	68	-	-	-	-	-	-	-	-	51.5%	55.9%
Leu (2)	73	73	-	-	-	-	-	-	-	-	45.2%	45.2%
ND5	1839	1839	1839	ATG	ATG	ATG	TAA	TAG	TAA	-	45.6%	48.3%
ND6	522	522	552	ATG	ATG	ATG	TAA	TAA	TAG	-	48.5%	53.8%
Glu	69	69	-	-	-	-	-	-	-	-	42.0%	40.6%
CYTb	1141	1141	1141	ATG	ATG	ATG	T - -	T - -	T - -	-	49.3%	50.0%
Thr	72	72	-	-	-	-	-	-	-	-	51.4%	51.4%
Pro	71	71	-	-	-	-	-	-	-	-	40.8%	40.8%
D-loop	886	889	-	-	-	-	-	-	-	-	38.4%	39.7%

Variable and conserved regions

The within-species genetic variation of the *P. georgianus* mitogenome was 2.1%, which was comprised of 356 single nucleotide polymorphisms (SNP's). The most variable region was the control region (see Figure 2.11) with 100 SNP's (11.2% of the region) and 41 parsimony informative sites (4.6% of the region) (see Table 2.5). The ND2 gene was the second most variable (4.2% of the region) with 1.7% of region consisting of parsimony informative sites. rRNA's (12S and 16S) and tRNA's were the least variable and containing several conserved regions (see Figure 2.11). One conserved region was identified in the first half of the control region, although the conservation value was low (conservation: 0.757, p -value < 0.0001) (see Figure 2.11).

Table 2.5: Within species variation along the *P. georgianus* mitogenome.

DNA region	Length	Single nucleotide polymorphisms		Parsimony informative sites	
		Number	Proportion of region	Number	Proportion of region
12S	951	5	0.5%	0	0.0%
16S	1719	9	0.5%	1	0.1%
ND1	975	23	2.4%	9	0.9%
ND2	1040	44	4.2%	18	1.7%
COX1	1551	15	1.0%	5	0.3%
COX2	691	7	1.0%	0	0.0%
ATP8	165	1	0.6%	0	0.0%
ATP6	684	14	2.0%	4	0.6%
COX3	785	9	1.1%	2	0.3%
ND3	349	5	1.4%	1	0.3%
ND4L	297	5	1.7%	2	0.7%
ND4	1381	25	1.8%	7	0.5%
ND5	1839	36	2.0%	14	0.8%
ND6	522	13	2.5%	5	1.0%
CYTB	1141	37	3.2%	11	1.0%
D-loop	886	100	11.2%	41	4.6%

Table 2.6: Conserved regions in the *P. georgianus* mitogenome.

DNA region	Start	End	Length	Conservation value	Significance
tRNA-Phe, 12s rRNA	1	222	221	1.000	0.0086
12s rRNA	224	669	445	1.000	0.0001
12s rRNA, tRNA-Val, 16s rRNA	798	1213	424	1.000	0.0001
16s rRNA	1218	1857	639	1.000	0.0000
16s rRNA	1958	2232	274	1.000	0.0027
16s rRNA, tRNA-Leu (1)	2329	2963	634	1.000	0.0000
tRNA-Ile, tRNA-Gln, tRNA-Met	3864	4059	195	1.000	0.0151
tRNA-Trp, tRNA-Ala, tRNA-Asn	5145	5354	209	1.000	0.0111
tRNA-Cys, tRNA-Tyr, COX1	5386	5849	436	1.000	0.0000
COX1	5851	6041	190	1.000	0.0168
COX1	6172	6551	379	1.000	0.0003
COX1	6823	6971	148	1.000	0.0414
COX1, tRNA-Ser (1), tRNA-Asp, COX2	6979	7381	402	1.000	0.0002
COX2	7647	7831	184	1.000	0.0191
COX2, tRNA-Lys, ATP8	7833	8118	285	1.000	0.0022
COX3	9081	9250	169	1.000	0.0264
COX3, tRNA-Gly, ND3	9369	9719	350	1.000	0.0005
ND3, tRNA-Arg	9918	10077	159	1.000	0.0327
ND4L, ND4	10283	10434	151	1.000	0.0388
ND4	10507	10727	220	1.000	0.0088
ND4	11521	11717	196	1.000	0.0147
ND4, tRNA-His, tRNA-Ser (2), tRNA-Leu (2), ND5	11719	12010	291	1.000	0.0019
ND5	12069	12289	220	1.000	0.0088
ND5	12408	12559	151	1.000	0.0388
ND6	14121	14272	151	1.000	0.0388
tRNA-Glu, CYTB	14353	14524	171	1.000	0.0253
CYTB, tRNA-Thr, tRNA-Pro	15507	15703	196	1.000	0.0147
Control region	15738	16073	335	0.757	0.0000

Figure 2.11: Within species variation at each nucleotide position among thirteen *P. georgianus* whole mitogenomes (blue), 304 partial control region sequences (black) and 30 partial COI sequences (black) as well as conserved regions and parsimony informative sites.

2.4 Discussion

2.4.1 Methods for mitogenome assembly

An exploration into the key choices made when assembling the *P. georgianus* mitogenome from whole genome sequence data revealed some subtleties in how the assembly algorithm handles real life data.

Reference mitogenome

When assembling the *P. georgianus* mitogenome, the choice of reference genome had an important impact on the mitogenome assembly and the consensus mitogenomes. This included the annotation of gene and spacer regions, the identity of the whole genome data to the reference mitogenomes and resolution of uncertain regions.

Importantly, two reference mitogenomes from the same species (*Carangoides equula*) performed differently as reference genomes. In fact, a reference genome of a species from a more distantly related genus (*T. japonicus*) performed better as a reference genome than a genome from a genus that is considered more closely related to the *Pseudocaranx* genus (*C. equula 1*). It is unknown if this difference is due to incorrect taxonomic labelling of the reference genomes, within-species variation or differences in the quality of the reference mitogenomes. Nonetheless, it highlights the importance of establishing how a mitogenome is performing as a reference during the assembly process. More specifically, a reference mitogenome from a taxonomically closely related species does not guarantee it will perform as well as a reference mitogenome.

In the case where only a portion of the mitogenome requires assembly, how genetically closely related the reference genome will need to be will depend on the mutation rate of the region. Less sequencing

reads were mapped to highly variable regions such as the control region and in turn more sequencing reads were mapped to conserved regions such as 12S rRNA and 16S rRNA. This means that highly variable regions will require assembly to a closer relative than more conserved regions.

Number of mapping iterations

Increasing the number of mapping iterations past the so called ‘saturation point’ minimized the number of regions with no contig coverage, low coverage depth or low identity to the reference. Reaching this saturation point produced a more complete and robust consensus sequence as more whole genome sequence data is contributed to the contig scaffold. The number of mapping iterations required to reach this saturation point was intuitively associated with the similarity of the whole genome data to the reference mitogenome (the overall genetic similarity of the two individuals). The number of identical sites does not appear to be affected by the number of mapping iterations, instead there is a stronger influence from the average contig depth. The difference suggests that the low proportion of identical sites in some individuals could be caused by sequencing errors or contamination of sequences in during high-throughput sequencing. For the map to reference algorithm (Geneious version 11.1.5 (Kearse et al., 2012)) there does not appear to be a downside associated with using many mapping iterations such as 100 or 10,000. This is because the mapping iterations are discontinued once no additional sequence reads are aligned to the reference. When using this approach, caution is advised because using large numbers of mapping iterations could introduce mistakes or bias in the consensus sequence as more sequencing errors are included in the assembly.

The number of mapping iterations required to produce the most robust mitogenome was intuitively related to the similarity of the reference genome to the whole genome data. Less similar reference genomes required more mapping iterations to reach the highest quality contig scaffold. A reduction in quality in the contig scaffold and consensus mitogenome sequence, which was produced when using a more distantly related reference, could not be mitigated by simply increasing the number of mapping iterations. The individuals with a high contig depth also tended to have a larger spread of contig depths along the mitogenome, indicating that some regions of the mitogenome were receiving a disproportionately higher number of sequencing reads in the contig scaffolds. This indicates that the number of mapping iterations required to produce accurate mitogenome sequence data is variable among whole genome sequences of different *P. georgianus* individuals and among DNA regions. More specifically, an individual for which a high number of mapping reads have been produced may require fewer mapping iterations because more conserved regions such as 12S rRNA and 16S rRNA anchor the first parts of the assembly process.

Different *P. georgianus* individuals

Despite the pairwise identity along the length of the mitogenomes being highly consistent among individuals mapped to the same reference genome, the contig depth and proportion of identical sites was highly variable among the individuals that were sequenced. These observations appear related, since a negative trend between the average contig depth and proportion of identical sites was independent to how many mapping iterations were used. The decrease in the proportion of identical sites with increased coverage could be a result of an uneven number of sequencing reads across the genome and therefore variability in the number of reads mapped to the reference.

Recommendations

1. Use a high quality reference mitogenome of a close relative. Although it is tempting to rely on a high number of mapping iterations when using a reference genome that is dissimilar or distantly related to the whole genome data, there is a limit to what a large number of mapping iterations can do to resolve uncertainties in the contig scaffold and the consensus mitogenome. In the case where only a portion of the mitogenome requires assembly, the genetic similarity of the reference genome can be tailored to suit the mutation rate of the region. To ensure the genetic similarity of the reference genome, it is advisable to conduct BLAST searches of several regions of the mitogenome including highly variable regions such as the control region. This will verify whether the reference genome is genetically similar as opposed to simply taxonomically similar to the whole genome data.
2. Ensure enough mapping iterations are used. Doing so will reduce and in some cases eliminate low coverage and low percentage identity regions. The number of mapping iterations required to reach this saturation point will depend on the data set. For example, if the reference is highly similar to the whole genome data, fewer iterations will likely be required and in turn, a highly dissimilar data set will require more mapping iterations.
3. Investigate the quality of the assembly. Mapping the whole genome data to a few different reference genomes, comparing the percentage identity along the length of the genome and looking for annotation conflicts will aid in establishing a robust reference mitogenome.
4. Report data confidence. The quality of the contig scaffolds and consensus mitogenomes produced in this study were highly variable. A lower quality assembly could carry errors into downstream analyses. Reporting confidence in the data by plotting and reporting contig depth, percentage

identity along the length of the mitogenome as well as omitting poorly resolved regions would minimize uncertainties in the data resulting from sequencing or assembly errors.

This study was by no means an exhaustive investigation into the methods for assembling mitogenomes from whole genome sequence data, although it addresses a few central method choices that are made during the assembly process. It acts as a reminder to keep a genetic perspective when dealing with genetic data, to understand as well as report how the chosen methods affect and works for or against a given data set.

2.4.2 *P. georgianus* mitogenome

The *P. georgianus* mitogenome is typical of ray-finned and cartilaginous fish (see Satoh et al. (2016)), with no gene rearrangements identified, typical gene region lengths and gene stop and start codons. However, these results should be verified with direct mitochondrial sequencing methods to ensure the assemblies produced in this chapter are not simply reflecting the gene arrangements of the reference mitogenomes. For future research, this *P. georgianus* mitogenome assembly has highlighted several possible target gene-regions for population genetic and taxonomic studies. The COI gene of *P. georgianus* was moderately variable and is a common “bar coding” target used to delimit boundaries among different fish species (Ward et al. (2009)). With a good database of fish COI sequences, this makes it a good target for taxonomic investigation of the *Pseudocaranx* genus. The control region was the most variable and can be used for fine-scale population structure of *P. georgianus*. Furthermore, the *P. georgianus* mitogenome produced here provides the opportunity for it to be used as a reference mitogenome for the assembly of other fish species.

Chapter 3

Field and laboratory work

3.1 Data collection

Genetic data for species in the *Pseudocaranx* genus was sourced from The National Center of Biotechnology Information (NCBI) GenBank database and the Barcode of Life data System (BOLD Systems). Sequences that were not publicly available were provided by Dr William White from the Australian National Fish Collection at the Commonwealth Scientific and Industrial Research Organisation (CSIRO) and Dr Dirk Steinke from the University of Guelph, Ontario. The partial COI gene of forty-five specimens from the *Pseudocaranx* genus from a broad geographic range were obtained. This included sequences from thirty *P. georgianus*, eight *P. dentex*, sixteen *P. wrighti* and five *P. dinjerra* fish.

National Aquatic Biodiversity Information System (NABIS) map layer data for *P. georgianus* Quota Management Areas and species distribution (annual, seasonal, juvenile, spawning distributions) was provided by Julianne Sellars from the Ministry for Primary Industries (MPI).

3.2 Tissue sampling

In total, 1339 *P. georgianus* tissue samples were obtained from several locations around the North Island of New Zealand as well as seven from the Three Kings Islands and four from the Kermadec Islands. A further 87 *P. georgianus* tissues samples were obtained from three locations in Western Australian waters (see Figure 3.1).

Throughout sampling, all samples were handled with scissors and forceps sterilized with laboratory grade Bastion 70% isopropyl alcohol between each individual. All tissue samples were either preserved

Figure 3.1: All sampling locations of *P. georgianus*.

immediately, or within six hours of sampling or acquisition and were preserved and stored in separate tubes. Sample coding systems for museum voucher specimens or sub-sampled tissue were also recorded for each relevant sample. All tissues were held in a 4°C freezer within a PC2 laboratory at Victoria University of Wellington. The remaining tissues were placed in an -80°C freezer for long-term storage. Tissues sampled during his study were taken from the caudal fin and preserved in a high-salt DMSO solution (20% dimethyl sulfoxide (DMSO), 0.5M Ethylenediaminetetraacetic acid (EDTA) and double distilled water (ddH₂O) (UV treated, passed through a filtration system and autoclaved), sodium chloride (NaCl) saturated). All sub-sampled tissues from Museum collections and previous studies were muscle tissue preserved in ethanol, therefore all sub-samples of these tissues were preserved in 99% ethanol.

The bulk of the caudal fin tissue samples were collected from fish ready for processing at Moana New Zealand warehouse, which is located at 1-3 Bell Avenue, Mt Wellington, Auckland. A few additional tissues were sampled from their Wellington warehouse located at 17/21 Lorne St, Wellington. These factories process commercial *P. georgianus* landings from around New Zealand's North Island. Sampling involved cutting a two centimeter portion of the caudal fin from up to 100 fish from a particular fishing vessel. Prior to sampling, fish were chilled for one to three days on board the fishing vessels and

in the factory. While in the factory, the tissue samples were placed in individual zip-lock bags and kept on ice. Within three hours, each fin clip was transferred to individual tubes filled with 1.3 milliliters of high-salt DMSO preservative. These tubes were then frozen at -4°C within one to three days of sampling. The sampling location was typically provided verbally as a statistical area or a picture of the route that the fishing vessel took during the multi-day fishing trip. From this information, GPS coordinates of the sampling location were estimated and will be accurate to the level of statistical area.

Many tissue samples were also collected by Ministry for Primary Industries (MPI) fisheries observers who were assigned to commercial fishing vessels. Sampling boxes containing tubes filled with the high-salt DMSO preservative solution, sampling instructions, sample record sheet, scissors, forceps, gloves and 70% isopropanol wipes were mailed to fisheries observers. The sampling boxes were mailed back to the MPI office in Wellington. The sampling location was recorded as GPS coordinates at the start of the trawling tow.

A small number of *P. georgianus* tissue samples were also provided by fishers in Whangarei.

All *P. georgianus* samples caught in Western Australian waters were sub-sampled from muscle tissue held at the Center for Fish and Fisheries Research (Murdoch University, Western Australia) by Dr Jennifer Chaplin. These tissue samples were preserved long-term at Murdoch University in 70% ethanol. Sub-sampling was undertaken on alternate days from other genetic work in the lab to limit cross-contamination. Sub-sampled tissue was placed in 99% ethanol for a few days prior to transportation. The ethanol was removed and the preserved tissue mailed to Victoria University of Wellington. Upon receipt of the tissue samples, the tissues were re-preserved in 99% ethanol.

Muscle tissue was also sub-sampled from voucher specimens held in the fish collection at The Museum of New Zealand Te Papa Tongarewa (Te Papa) (169 Tory Street, Wellington) that were fixed in 98% ethanol. The tissue was sub-sampled by a museum curator and transferred into tubes containing 99% ethanol.

Eighteen whole juvenile *P. georgianus* caught in Whangarei as a part of a mixed species catch were provided by Dr Darren Parsons from the Auckland office of the National Institute of Water and Atmospheric Research (NIWA). The tissue samples were frozen for several months at NIWA. When they were delivered to Victoria University of Wellington, the whole fish specimens were preserved in a high salt DMSO preservative.

3.3 DNA extraction

A high-salt extraction protocol was used to isolate DNA from selected samples. Using flame-sterilized utensils, excess salts or ethanol were removed or scraped from a small portion of tissue. The tissue was finely chopped and placed in a solution for cell lysis, comprised of DNA extraction buffer, sodium dodecyl sulfate (SDS) and Proteinase K. These were gently agitated for at least two hours in an orbital mixer set to 50°C. Proteins were precipitated using NaCl, then removed along with the supernatant. The DNA was precipitated with 100% isopropyl alcohol, pelleted in a centrifuge and then washed with 70% ethanol. The pelleted DNA was dried, re-suspended in a Tris-EDTA (TE) buffer and kept frozen at -4°C until required for PCR amplification (the full DNA extraction methodology can be found in Appendix A).

To ensure large un-degraded genomic DNA was being extracted, initial DNA extractions were run through a 1% agarose gel, stained with Ethidium Bromide and visualized with a Uvitec Essential V6 UV imaging box. The DNA quantity and quality of all DNA extractions were measured on an Implen NanoPhotometer NP80 NanoDrop. A ratio of 260/280 and 260/230 readings above 1.8 were considered acceptable for PCR amplification. The median readings were 1360.8ng/ μ L for DNA quantity, 2.121 for 260/280 readings and 2.224 for 260/230 readings. RNA digestion was not used in the DNA extraction process, therefore NanoDrop readings higher than 1.8 were expected. DNA extractions were repeated on tissue samples until they met the quality and quantity standard. The concentration of the DNA template was standardised across samples by diluted with them with TE buffer to a final concentration of between 100-200ng/ μ l. The extracted genomic DNA was used to PCR amplify the COI gene and control region.

3.4 Primer design and sequencing

From the consensus *P. georgianus* mitogenome assembled in Chapter 2), species-specific primers were designed to target the COI gene and the control region of selected samples. The primers were designed in Primer 3 version 2.3.7 (Untergasser et al., 2012) within Geneious version 11.1.5 (Kearse et al., 2012) and synthesised by Invitrogen.

PCR amplification was undertaken in a sterilised fume hood in a dedicated PCR room within a PC2 laboratory. Initial PCR's were conducted to establish a set of optimised PCR conditions for the custom primers. This involved varying the annealing temperature, MgCl concentration and DNA template concentration until a single PCR-product of high yield was produced. The final PCR products were purified with ExoSAP-ITTM PCR Product Cleanup Reagent (incubated at 37°C for 15 minutes

to remove excess primers and dNTP's, then 80°C for 15 minutes to inactivate the Exo-SAP-IT™ enzyme).

DNA sequencing of COI gene as well as initial and low yield control region PCR products was carried out using a ABI 3730 Sequencer at Massey University Sequencing and Genotyping Services in Palmerston North, New Zealand. Although the majority of the control region sequencing was carried out by Macrogen in Seoul, the Republic of Korea.

3.4.1 COI gene

The COI gene of fourteen *P. georgianus* fish sampled from New Zealand's North Island were PCR amplified and sequenced using the final PCR protocol below. The final components of the PCR reaction are described in Table 3.2) and summary information on all final *P. georgianus* COI sequences can be found in Table 3.5 (see Chapter 4 for more detailed information on the COI sequences).

1. An initial soak of 95°C for 120 seconds
2. 35 cycles of:
 - Denaturation (95°C for 60 seconds)
 - Annealing (69.7°C for 60 seconds)
 - Extension (72°C for 60 seconds)
3. A final extension step of 72°C for 10 minutes.

Table 3.1: A description of the primers designed to target the COI gene of *P. georgianus*.

Primer name	Direction	Primer sequence (5' to 3')	GC content
TRE_tRNA_Tyr_F	Forward	GGG CTA CAA TCC ACC GCT TA	55%
TRE_mid_COI_R	Reverse	ACA CCA TTC CCA TGT ACC CG	55%

Table 3.2: Components of the PCR reaction used to amplify the partial COI gene of *P. georgianus*.

PCR component	Concentration	Volume (µl)
PCR H ₂ O	-	14.4
NH ₄ Reaction Buffer (Bioline)	10x	2.5
MgCl (Bioline)	50mM	1.5
BSA	10mg/ul	1.5
Forward primer (TRE_tRNA_Pro_F)	10uM	1.0
Reverse Primer (TRE_control_R_2)	10uM	1.0
dNTP's	5uM	2.0
BIOTAQ™ DNA polymerase	-	0.1
DNA template	100-200ng/ul	1.0
TOTAL	-	25.0

3.4.2 Control region

The control region of 389 *P. georgianus* fish sampled from New Zealand's North Island and Western Australia were PCR amplified and sequenced using the final PCR protocol below. The final components of the PCR reaction are described in Table 3.4) and summary information on all final *P. georgianus* control region sequences can be found in Table 3.6 (see Chapter 5 for more detailed information on the control region sequences).

1. An initial soak of 95°C for 120 seconds
2. 35 cycles of:
 - Denaturation (95°C for 60 seconds)
 - Annealing (68.4°C for 60 seconds)
 - Extension (72°C for 60 seconds)
3. A final extension step of 72°C for 10 minutes

Table 3.3: A description of the primers designed to target the control region of *P. georgianus*.

Primer name	Direction	Primer sequence (5' to 3')	GC content
TRE_tRNA_Pro_F	Forward	ACT GCT CAG AAA AAG GAG ACT CT	43%
TRE_control_R_2	Reverse	AGT GAT TCT AGG AGT TTA GGG GG	47%

Table 3.4: Components of the PCR reaction used to amplify the partial control region of *P. georgianus*.

PCR component	Concentration	Volume (ul)
PCR H ₂ O	-	13.9
NH ₄ Reaction Buffer (Bioline)	10x	2.5
MgCl (Bioline)	50mM	2.0
BSA	10mg/ul	1.5
Forward primer (TRE_tRNA_Pro_F)	10uM	1.0
Reverse Primer (TRE_control_R_2)	10uM	1.0
dNTP's	5uM	2.0
BIOTAQ™ DNA polymerase	-	0.1
DNA template	100-200ng/ul	1.0
TOTAL	-	25.0

Table 3.5: Summary information on all COI sequences that were acquired and sequenced for *Pseudocaranx* species.

Species	Sampling location	Source	Number of sequences
<i>P. dentex</i>	French Polynesia	Genbank	1
<i>P. dentex</i>	Israel	BOLD	1
<i>P. dentex</i>	Queensland	Genbank	1
<i>P. dentex</i>	South Africa	BOLD	2
<i>P. dentex</i>	United States of America	BOLD	1
<i>P. dentex</i>	Victoria	BOLD	1
<i>P. dentex</i>	Western Australia	BOLD	1
<i>P. dinjerra</i>	Western Australia	Genbank	5
<i>P. georgianus</i>	New Zealand	Sequenced	14
<i>P. georgianus</i>	Victoria	Dirk Steinke	2
<i>P. georgianus</i>	Western Australia	Dirk Steinke	1
<i>P. georgianus</i>	Western Australia	Genbank	10
<i>P. georgianus</i>	Western Australia	William White (CSIRO)	3
<i>P. wrighti</i>	South Australia	Dirk Steinke	3
<i>P. wrighti</i>	Western Australia	BOLD	1
<i>P. wrighti</i>	Western Australia	Dirk Steinke	1
<i>P. wrighti</i>	Western Australia	Genbank	4
<i>P. wrighti</i>	Western Australia	William White (CSIRO)	7

Table 3.6: Summary information on all *P. georgianus* samples for which the control region was sequenced.

Sampling country	Sampling QMA	Sample source	Number of sequences
New Zealand	TRE1	Fishers Whangarei	6
New Zealand	TRE1	Darren Parsons	13
New Zealand	TRE1	Moana Auckland	78
New Zealand	TRE1	MPI	55
New Zealand	TRE10	Te Papa	4
New Zealand	TRE2	Moana Auckland	58
New Zealand	TRE2	Te Papa	1
New Zealand	TRE7	Moana Auckland	20
New Zealand	TRE7	Moana Wellington	4
New Zealand	TRE7	MPI	61
New Zealand	TRE7	Te Papa	23
Australia	-	Murdoch University	68

Chapter 4

Evolutionary relationships of *Pseudocaranx* species

4.1 Introduction

4.1.1 Current *Pseudocaranx* taxonomy

Delimiting species in the *Pseudocaranx* genus has proved difficult, resulting in several taxonomic revisions (James and Stephenson, 1974). From initially a single species, *Pseudocaranx dentex*, several species have subsequently been described that cover a wide geographic distribution. Located around the Juan Fernández Islands on the west coast of Chile, *Pseudocaranx chilensis* was split from *Pseudocaranx dentex* by Dyer and Westneat (2010). In Tosa bay, Japan, genetic and morphological variation suggest that two species exist within *P. dentex* which are both distinct from *P. wrighti* and *P. chilensis* based on morphology and geography (Masuda et al. (1995); Yamaoka et al. (1992)). In Australia, James and Stephenson (1974) observed specimens from southern Australia with variable vertebrae numbers that were subsequently identified by Smith-Vaniz and Jelks (2006) as *P. wrighti* and *Pseudocaranx georgianus*. That study also identified a new species (*Pseudocaranx dinjerra*) from Western Australia and discussed specimens that likely represent another species occurring off southern Queensland and Lord Howe Island which they state, “definitely are neither *P. dinjerra* nor *P. wrighti*” (p. 97).

The current taxonomy describes five species in the *Pseudocaranx* genus; *Pseudocaranx dentex* (Bloch & Schneider, 1801), *Pseudocaranx chilensis* (Guichenot, 1848), *Pseudocaranx wrighti* (Whitley, 1931), *Pseudocaranx georgianus* (Cuvier, 1833) and *Pseudocaranx dinjerra* (Smith-Vaniz and Jelks, 2006).

4.1.2 *Pseudocaranx* in New Zealand

Currently, *P. georgianus* is the only *Pseudocaranx* species known to occur in New Zealand waters. However, taxonomic studies have indicated that a cryptic species that may occur in New Zealand based on morphology. James and Stephenson (1974) observed specimens in Norfolk Island, Kermadec (Raoul) Island and the North Cape of New Zealand with fewer caudal vertebrae (fourteen) than *P. georgianus* from other regions of New Zealand. They did not consider this morphological difference enough to warrant considering it a different species. However, after re-examining the same specimens, Smith-Vaniz and Jelks (2006) believe James and Stephenson (1974) underestimated other morphological characters that differentiate these specimens from *P. georgianus* from the rest of New Zealand. Smith-Vaniz and Jelks (2006) claim these fish are “definitely neither *P. dinjerra* nor *P. wrighti*” (p.97).

Despite the specific claim by Smith-Vaniz and Jelks (2006) that the specimens they observed in the Kermadec (Raoul) Island and the North Cape of New Zealand are not *P. wrighti*, Duffy and Ah Yong (2015) reports *P. wrighti* occurring in the both of these regions in a study recording the taxa occurring within the Kermadec Island Marine Reserve. Interestingly, their taxonomic classification was based partly on Smith-Vaniz and Jelks (2006).

Other sources do not describe *P. wrighti* occurring in these regions. They are reported to occur only in the southern half of Australia from Western Australia to New South Wales (Froese and Pauly (2019), Smith-Vaniz and Jelks (2006)).

Preliminary genetic work has also found fish from south-eastern Australia which are distinct from fish from northern New Zealand (Rowling and Raines, 2000). However, it is unknown if these fish are distinct from *P. georgianus* from other regions in New Zealand or are the same species from the North Cape and the Kermadec Islands described by James and Stephenson (1974) and Smith-Vaniz and Jelks (2006).

4.1.3 Historical reliance on morphological characters

Although many morphological features have been used to delimit species in this genus, such as head shape and the number of gill rakers, anal fin rays, dorsal rays, anal spines and scutes (see James and Stephenson (1974); Smith-Vaniz and Jelks (2006)), the total number of vertebrae and caudal vertebrae have important distinguishing features for *Pseudocaranx* species. The total vertebrae number was the only morphological feature that was measured which showed a clear difference between two genetically distinct *Pseudocaranx* groups in Japan (*P. dentex* (Type A) and another undescribed species (Type B)) (Yamaoka et al., 1992). Distinction of wild specimens of these two groups (Type A and Type B)

using caudal vertebrae number was also supported by Masuda et al. (1995).

Caudal vertebrae number was the only character with no intra-species variation for Australian species of *Pseudocaranx*, unlike counts of three types of fin rays and three types of gill rakers in Smith-Vaniz and Jelks (2006) for which frequency distributions overlap among species and caudal vertebrae count was one of the features used to describe many of the currently recognised *Pseudocaranx* species (Smith-Vaniz and Jelks, 2006). Vertebrae number was also one of the morphological features used to combine five species described at the time to *Caranx georgianus* (later revised to *Pseudocaranx georgianus*) (James and Stephenson, 1974). Because of its comparative lack of intra-species variation, caudal vertebrae number has historically been an important component of delimiting *Pseudocaranx* species.

Importantly, Masuda et al. (1995) found that a quarter of artificially reared *P. dentex* specimens had a greater number of vertebrae (25) than wild caught specimens (24) despite being genetically the same species. They postulated that this could have been triggered by the lower water temperatures in the artificial environment and speculated that differences in water temperature during the larval stage in a wild environment could affect this trait. It was also observed that *P. georgianus* in New Zealand (at the time described as *Caranx georgianus*) showed a regular increase in the proportion of fish with 24 vertebrae rather than 25 vertebrae with decreases in latitude (James and Stephenson, 1974), where waters are generally cooler. It is unknown whether these represent genetically distinct species, or within species variation in vertebrae number possibly as a result of environmental temperature.

Other morphological characteristics have been used as the basis for separating species. Principal component analyses have found subtle groupings of *P. dinjerra*, *P. georgianus* and *P.sp. "dentex"* (a putative species off Queensland and Lord Howe Island) based on snout and upper jaw lengths, straight lateral line and anal-fin base lengths (Smith-Vaniz and Jelks, 2006). Yamaoka et al. (1992) found general differences among two genetic types of *P. dentex* found in Japan such as a the number of anal-rays, visibility of vertical stripes on the lateral side of juveniles and body size, although with the exception of vertebrae number, no consistent differences in morphology were found, similar findings were reported by Masuda et al. (1995). For museum specimens of *Pseudocaranx* species sampled from a broad geographic range in Western Australian waters, James and Stephenson (1974) failed to find morphological variation beyond vertebrae number to support a previous separation of species in this region.

Although there is variation in morphological characters within the *Pseudocaranx* genus that may be associated with species level taxonomy, this variation is not obviously defined between species. The range of intra-species variation of most of these characters are reported to overlap between species (see James and Stephenson (1974); Masuda et al. (1995); Smith-Vaniz and Jelks (2006); Yamaoka et al. (1992)). In fact, even though the average number of dorsal-fin rays, anal-fin rays and pectoral-

fin rays differ among species, many *P. wrighti*, *P. dinjerra*, *P. georgianus* individuals will show the same number of these characteristics. Researchers have delimited several species in the *Pseudocaranx* genus, but the taxonomy remains for the large part unresolved. Because of the lack of clear species-level delineation of morphological characters and the possible contribution of environmental variables to character variability, the current *Pseudocaranx* taxonomy, which is based entirely on morphology, requires phylogenetic verification. This will allow us to verify current *Pseudocaranx* taxonomy and identify morphologically cryptic species.

4.1.4 Previous study

Two phylogenetic studies of the wider Carangidae family have been undertaken (Damerou et al., 2018; Santini and Carnevale, 2015) although they include a small number of *Pseudocaranx* species from a limited geographic range. Santini and Carnevale (2015) include only a single *P. dentex* specimen sampled from South of Cape Everard in Victoria (which is likely to be in fact *P. georgianus* given its sampling location) and a *P. wrighti* specimen sampled from Rottnest Island in Western Australia. Damerou et al. (2018) includes a single *P. dentex* fish sampled from an archipelago of Cape Verde in the North Atlantic Ocean.

At the genus level, only two studies have used genetic markers to inform the taxonomy of the *Pseudocaranx* species. Yamaoka et al. (1992) revealed a cryptic species based on 28 isozyme and protein loci of 94 *P. dentex* specimens from Japan, which was later supported by an analysis of the 16S-rRNA DNA sequences of 24 fish (Masuda et al., 1995). Genetic differences between *P. georgianus* (labelled as *P. dentex* at the time) and *P. wrighti* within southern Australian waters based on control region sequences of 59 fish have also been reported (Bearham, 2004).

A comprehensive phylogenetic study of all *Pseudocaranx* species found in Australasia has not yet been undertaken and the relationships of *P. georgianus* from New Zealand has not been investigated. Importantly, the current management of New Zealand *P. georgianus* fisheries rely on the assumption that *P. georgianus* within Quota Management Areas are composed only of *P. georgianus*.

4.2 Materials and methods

The partial COI sequences of four species in the *Pseudocaranx* genus (*P. dentex*, *P. dinjerra*, *P. georgianus* and *P. wrighti*) were either procured or sequenced in this thesis. Sequences were obtained from The National Center of Biotechnology Information (NCBI) GenBank database, the Barcode of Life data System (BOLD Systems) or private sources. Additional COI sequences were produced using

DNA extracted from *P. georgianus* tissue samples sampled in this thesis and primers designed in this thesis (see Chapter 3 for further information). Sequencing of these PCR products was carried out using a ABI 3730 Sequencer at Massey University Sequencing and Genotyping Services in Palmerston North, New Zealand

The editing of the DNA sequences and alignment was carried out in Geneious version 11.1.5 (Kearse et al., 2012). The *Pseudocaranx* sequences and one outgroup (*Seriola lalandi*, GenBank accession: KM877619) were aligned using the Geneious alignment tool with the default settings (Alignment type: global alignment with free end gaps, Cost matrix: 65% similarity, Gap open penalty: 12, Gap extension penalty: 3, Refinement iterations: 2). The alignments were manually edited to removed base pair calling noise and areas where the primer had been incorporated into the sequence. The chromatogram files of the DNA sequences were visually checked for signal strength and unambiguous base calling. The final alignment covered a 602 base pair region of the COI gene, excluding the first 96 and last 853 base-pairs of the gene (see Figure 4.1 and see Figure A.2 in the Appendix for the final COI gene alignment).

A nucleotide substitution model appropriate for the COI gene of this genus was estimated by using jModelTest version 2.1.10 (Darriba et al., 2012, Guindon and Gascuel (2003)). The final COI alignment was used to test 88 models including those with equal-unequal base frequencies (+F), with-without a proportion of invariable sites (+I) and with-without rate heterogeneity among sites (+G). Eleven substitution schemes were explored using a Maximum Likelihood base tree for likelihood calculations. The model with the highest Bayesian Information Criterion (BIC) support was HKY+G and was used for phylogenetic analyses.

Phylogenetic inference of the *Pseudocaranx* genus was based on both maximum likelihood and Bayesian inference. Both phylogenies were created using plugins within Geneious version 11.1.5 (Kearse et al., 2012) (PHYML plugin version 2.2.4 (Guindon et al., 2010) and MrBayes plugin version 2.2.4 (Huelsenbeck and Ronquist, 2001)). The appropriate transition/transversion ratio and gamma distribution parameter was specified in the maximum likelihood phylogeny and branch support was given by 100 bootstrap replicates. The gene genealogies for the Bayesian phylogeny were explored using four MCMC chains 1,100,000 in length and 0.2 chain temperature. There was a burn in of 100,000 samples after which every 200th gene genealogy was sampled.

Figure 4.1: COI primer binding locations and within species variation among thirty *P. georgianus* individuals from New Zealand (black) and thirteen broodstock *P. georgianus* (blue).

4.3 Results

In total, partial COI sequences for eight *P. dentex*, five *P. dinjerra*, thirty *P. georgianus* and sixteen *P. wrighti* individuals were either procured or sequenced (see Table 4.1).

The best fit nucleotide substitution model was a HKY+G model (freqA = 0.2255, freqC = 0.2990, freqG = 0.1809, freqT = 0.2946, transition/transversion ratio = 7.4435, gamma shape = 0.201).

Both phylogenies had congruent overall topologies, although the branch support was variable between the two inference types (see Figures 4.2 and 4.3). Typically, the maximum likelihood phylogeny gave lower support values (MLsupport) for branches compared to the Bayesian phylogeny (Bsupport). Despite this, both phylogenies returned 4 overall clades:

1. *P. georgianus*/*P. dentex* complex (clade 1)
2. *P. dentex* complex (clade 2)
3. *P. dinjerra* clade (clade 3)
4. *P. wrighti* clade (clade 4)

See Figure 4.4 and Table 4.1 for the sampling locations of *Pseudocaranx* species from these four phylogenetic clades.

Both phylogenies congruently returned *P. wrighti* as a monophyletic clade with high support (MLsupport: 70, Bsupport: 1). This *P. wrighti* clade was distinguished from the *P. georgianus*/*P. dentex*/*P. dinjerra* complex comprised of three sister clades (clade 1, 2 and 3) by the Bayesian phylogeny (Bsupport: 1), although the support for the maximum likelihood phylogeny was low (MLsupport: 54).

One of the sister clades in the *P. georgianus*/*P. dentex*/*P. dinjerra* complex was a paraphyletic group including all *P. georgianus* individuals sampled throughout Australia and New Zealand and two *P. dentex* individuals sampled from Rottnest Island in Western Australia and Cape Everard in Victoria (*P. dentex* 02 and *P. dentex* 06). This clade again had high support from the Bayesian phylogeny (Bsupport: 0.93) but low support from the maximum likelihood phylogeny (MLsupport: 54). The rest of the *P. dentex* individuals sampled from a broad range of locations (Israel, South Africa, Queensland, French Polynesia and the United States of America) (*P. dentex* 01, *P. dentex* 03, *P. dentex* 04, *P. dentex* 05, *P. dentex* 07 and *P. dentex* 08) represented a second monophyletic sister clade with high support (MLsupport: 78, Bsupport: 1). The third sister clade included *P. dinjerra* individuals sampled from the Abrolhos Islands in Western Australia, with high support in the Bayesian phylogeny (Bsupport: 1) and the maximum likelihood phylogeny (MLsupport: 98).

Within the four overall clades, some individuals were more closely related to each other than to the rest of the individuals in their clade. Within clade 1, there were two grouping of *P. georgianus* individuals:

1. Three *P. georgianus* individuals from Victoria in Australia (*P. georgianus* 07), Gisborne in New Zealand (*P. georgianus* 16) and Geographe Bay in Western Australia (*P. georgianus* 24) are grouped together with high support in the Bayesian phylogeny (Bsupport: 0.95).
2. Three *P. georgianus* individuals from Raglan in New Zealand (*P. georgianus* 12), Rottnest Island in Western Australia (*P. georgianus* 20) and Geraldton in Western Australia (*P. georgianus* 29) are grouped together with low support (Bsupport: 0.66).

Both *P. georgianus* groupings were not supported in the maximum likelihood phylogeny. Within clade 2, the *P. dentex* individual from Israel was the most distantly related to the rest of the *P. dentex* within this clade, with the rest of the clade sampled from South Africa, Queensland, French Polynesia and the United States of America, grouped together with high support in both phylogenies (MLsupport: 78, Bsupport: 0.95). Within clade 3, three pairs of *P. wrighti* were grouped:

1. *P. wrighti* 04 from Rottnest Island and *P. wrighti* 15 from Warnbro Sound in Western Australia (MLsupport: 70, Bsupport: 0.97).
2. *P. wrighti* 07 from Victoria and *P. wrighti* 12 from Warnbro Sound in Western Australia.
3. *P. wrighti* 02 from Rottnest Island and *P. wrighti* 16 from Cockburn Sound in Western Australia (MLsupport: 61, Bsupport: 0.88).

Figure 4.2: Maximum likelihood *Pseudocaranx* phylogeny including bootstrap support greater than 50.

Figure 4.3: Bayesian *Pseudocaranx* phylogeny including posterior probabilities greater than 0.5.

Figure 4.4: Sampling locations of the four major *Pseudocaranx* phylogenetic clades including *P. georgianus* (red), *P. dentex* (purple), *P. dinjerra* (yellow) and *P. wrighti* (green).

4.4 Discussion

This phylogenetic analysis as a whole supported the current taxonomy of the *Pseudocaranx* genus, with clear genetic support for the grouping of *P. wrighti* and *P. dinjerra*. Although, the taxonomic distinction of *P. dentex* and *P. georgianus* was not as clearly distinguished. Despite most *P. dentex* individuals being placed in a monophyletic group, two *P. dentex* individuals were placed within a well-supported group along with all *P. georgianus* individuals from Australasia. The grouping of these *P. dentex* individuals within clade 1 could be due them being *P. georgianus* individuals which are taxonomically mislabeled. Since these *P. dentex* specimens were sampled in 1994 (*P. dentex* 06) and 2006 (*P. dentex* 02), it is possible that they are *P. georgianus* specimens that have not been relabeled since the taxonomic revision of previously described *P. dentex* to *P. georgianus* by Smith-Vaniz and Jelks (2006). To confirm this, the specimens could be taxonomically inspected if these fish remain in the CSIRO, Australian National Fish Collection. If these specimens were confirmed to be *P. georgianus* it would mean that the current taxonomic classifications of all four of the species are genetically supported by the COI gene. Additionally, no evidence was found on the COI gene of a single *P. georgianus* fish caught off the North Cape of New Zealand (*P. georgianus* 15) to support the

Table 4.1: Information on all COI sequences that were acquired or sequenced for *Pseudocaranx* species.

Label	Sampling location		Source	GenBank accession	BOLD ID	Museum ID
	Region	Area				
<i>P. dentex</i> 01	Israel	Nitzanim	BOLD	-	BIM281-13	P. 14752
<i>P. dentex</i> 02	Western Australia	North of Rottnest Island	BOLD	-	FOAF169-07	CSIRO H 6346-07
<i>P. dentex</i> 03	South Africa	Techobanine	BOLD	KF489713	DSFSG564-11	-
<i>P. dentex</i> 04	Queensland	Caloundra	Genbank	MK101242	-	-
<i>P. dentex</i> 05	French Polynesia	Mangareva Island	Genbank	MK658523	-	-
<i>P. dentex</i> 06	Victoria	South of Cape Everard	BOLD	EF609442	FOAC429-05	CSIRO H 3766-03
<i>P. dentex</i> 07	South Africa	-	BOLD	KF489712	DSFSG649-11	-
<i>P. dentex</i> 08	United States of America	Illinois	BOLD	-	OCARH994-12	-
<i>P. dinjerra</i> 01	Western Australia	Aborolhos Islands	Genbank	MK101237	-	-
<i>P. dinjerra</i> 02	Western Australia	Aborolhos Islands	Genbank	MK101238	-	-
<i>P. dinjerra</i> 03	Western Australia	Aborolhos Islands	Genbank	MK101239	-	-
<i>P. dinjerra</i> 04	Western Australia	Aborolhos Islands	Genbank	MK101240	-	-
<i>P. dinjerra</i> 05	Western Australia	Aborolhos Islands	Genbank	MK101241	-	-
<i>P. georgianus</i> 01	New Zealand	Whangamata	Sequenced	-	-	-
<i>P. georgianus</i> 02	New Zealand	Whangamata	Sequenced	-	-	-
<i>P. georgianus</i> 03	New Zealand	Whangamata	Sequenced	-	-	-
<i>P. georgianus</i> 04	New Zealand	Whangamata	Sequenced	-	-	-
<i>P. georgianus</i> 05	Victoria	Port Phillip Bay	Dirk Steinke	-	FMVIC823-08	A25443-006
<i>P. georgianus</i> 06	Western Australia	North East of Rottnest Island	Dirk Steinke	-	FMVIC824-08	A29358-020
<i>P. georgianus</i> 07	Victoria	-	Dirk Steinke	-	FMVIC822-08	A25443-006
<i>P. georgianus</i> 08	New Zealand	Raglan	Sequenced	-	-	-
<i>P. georgianus</i> 09	New Zealand	Raglan	Sequenced	-	-	-
<i>P. georgianus</i> 10	New Zealand	Raglan	Sequenced	-	-	-
<i>P. georgianus</i> 11	New Zealand	Raglan	Sequenced	-	-	-
<i>P. georgianus</i> 12	New Zealand	Raglan	Sequenced	-	-	-
<i>P. georgianus</i> 13	New Zealand	Raglan	Sequenced	-	-	-
<i>P. georgianus</i> 14	New Zealand	Bay of Plenty	Sequenced	-	-	-
<i>P. georgianus</i> 15	New Zealand	North Cape	Sequenced	-	-	-
<i>P. georgianus</i> 16	New Zealand	Gisborne	Sequenced	-	-	-
<i>P. georgianus</i> 17	New Zealand	Wellington region	Sequenced	-	-	-
<i>P. georgianus</i> 18	Western Australia	North of Rottnest Island	William White (CSIRO)	JN313226	FOAN113-11	CSIRO H6346-30
<i>P. georgianus</i> 19	Western Australia	North West of Rottnest Island	William White (CSIRO)	JN313242	FOAN134-11	CSIRO H6350-17
<i>P. georgianus</i> 20	Western Australia	North West of Rottnest Island	William White (CSIRO)	JN313243	FOAN135-11	CSIRO H6350-18
<i>P. georgianus</i> 21	Western Australia	Port Gregory	Genbank	MK101227	-	-
<i>P. georgianus</i> 22	Western Australia	Geraldton	Genbank	MK101228	-	-
<i>P. georgianus</i> 23	Western Australia	Dongara	Genbank	MK101229	-	-
<i>P. georgianus</i> 24	Western Australia	Geographe Bay	Genbank	MK101230	-	-
<i>P. georgianus</i> 25	Western Australia	Cape Naturaliste	Genbank	MK101231	-	-
<i>P. georgianus</i> 26	Western Australia	Geographe Bay	Genbank	MK101232	-	-
<i>P. georgianus</i> 27	Western Australia	Port Gregory	Genbank	MK101233	-	-
<i>P. georgianus</i> 28	Western Australia	Cape Naturaliste	Genbank	MK101234	-	-
<i>P. georgianus</i> 29	Western Australia	Geraldton	Genbank	MK101235	-	-
<i>P. georgianus</i> 30	Western Australia	Aborolhos Islands	Genbank	MK101236	-	-
<i>P. wrighti</i> 01	Western Australia	North of Rottnest Island	William White (CSIRO)	-	FOAC425-05	-
<i>P. wrighti</i> 02	Western Australia	North of Rottnest Island	William White (CSIRO)	-	FOAC426-05	-
<i>P. wrighti</i> 03	Western Australia	North of Rottnest Island	William White (CSIRO)	-	FOAC427-05	-
<i>P. wrighti</i> 04	Western Australia	North of Rottnest Island	William White (CSIRO)	-	FOAC428-05	-
<i>P. wrighti</i> 05	South Australia	Spencer Gulf, Western Shoal	Dirk Steinke	-	FMVIC825-08	A29204-003
<i>P. wrighti</i> 06	South Australia	Spencer Gulf, Western Shoal	Dirk Steinke	-	FMVIC826-08	A29214-002
<i>P. wrighti</i> 07	South Australia	Spencer Gulf, Middlebank	Dirk Steinke	-	FMVIC827-08	A29222-001
<i>P. wrighti</i> 08	Western Australia	North East of Rottnest Island	BOLD	EF609443	FOAC424-05	CSIRO H 4625-01
<i>P. wrighti</i> 09	Western Australia	South of Rottnest Island	William White (CSIRO)	JN313227	FOAN117-11	CSIRO H 6351-01
<i>P. wrighti</i> 10	Western Australia	South of Rottnest Island	William White (CSIRO)	JN313228	FOAN118-11	CSIRO H 6351-02
<i>P. wrighti</i> 11	Western Australia	South of Rottnest Island	William White (CSIRO)	JN313229	FOAN119-11	CSIRO H 6351-02
<i>P. wrighti</i> 12	Western Australia	Warnbro Sound	Genbank	MK101243	-	-
<i>P. wrighti</i> 13	Western Australia	Warnbro Sound	Genbank	MK101244	-	-
<i>P. wrighti</i> 14	Western Australia	Warnbro Sound	Genbank	MK101245	-	-
<i>P. wrighti</i> 15	Western Australia	Warnbro Sound	Genbank	MK101246	-	-
<i>P. wrighti</i> 16	Western Australia	Cockburn Sound	Dirk Steinke	-	FMVIC828-08	A29351-013

claim that a cryptic species may be occurring in this region (see Smith-Vaniz and Jelks (2006)).

All fish in this study were sampled from locations where the species is known to occur, providing no evidence to revise current species boundaries and distribution. The species sampled from the largest geographic range was *P. georgianus* and fish from this species were sampled from around the North Island of New Zealand and southern Australia. *P. wrighti* is known to occur from Western Australia, southern Australia to New South Wales Froese and Pauly (2019). This study only identified *P. wrighti* from areas within this geographic range, supporting the current species boundaries. *P. dinjerra* were sampled from only one location from the limited range where it is known to occur, on the west

coast of Western Australia Froese and Pauly (2019). *P. dinjerra* occurrence data is limited and this species may occur in a wider geographic range. In order to validate the edges of *Pseudocaranx* species' boundaries, taxonomic and genetic work should be undertaken on individuals sampled from a broader geographic range. This would be particularly important for species where fish were sampled from a limited geographic range in this study such as *P. wrighti* or for which limited species occurrence data is available such as *P. dinjerra*.

P. dentex has the largest geographic range, occurring in many regions worldwide (Froese and Pauly, 2019). A greater number of *Pseudocaranx* species have been identified within Australasia than any other area of the world. This likely reflects the fact that taxonomic work has been focused on *Pseudocaranx* species in Australasia. Where taxonomic or genetic work has been undertaken, cryptic species have been identified. For example, in Japan (Masuda et al. (1995); Yamaoka et al. (1992)), Australasia (James and Stephenson (1974); Smith-Vaniz and Jelks (2006)) and Chile (Dyer and Westneat, 2010). In this chapter, one *P. dentex* individual sampled from Israel was genetically distinct from other *P. dentex* from South Africa, French Polynesia and the USA. Given that *P. dentex* occurs over a very large geographic range, it is likely that there are other cryptic species within *P. dentex*.

In some cases, individuals within a species were closely related despite large distances between the locations they were sampled. *P. georgianus* individuals from the east coast of New Zealand and the west coast of Australia were closely related. Similarly, *P. wrighti* from Victoria and Western Australia were closely related. This genetic connection within species over long distances could indicate either gene flow between remote regions from migrants or a common ancestral history. Teasing apart these variables could be achieved through population genetic or tag-recapture studies based in *Pseudocaranx* biology.

Chapter 5

Population structure and demographic history of *P.* *georgianus*

5.1 Introduction

5.1.1 Population structure of New Zealand *P. georgianus*

P. georgianus is the only *Pseudocaranx* species currently known to occur in New Zealand waters. However, taxonomic studies have indicated a cryptic species that may occur in New Zealand based on the morphology of specimens in Norfolk Island, Kermadec (Raoul) Island and the North Cape of New Zealand (see Smith-Vaniz and Jelks (2006); James and Stephenson (1974)). At present, the literature is unclear on what species these fish represent (see Section 4.1).

Putative sub-stocks of *P. georgianus* in New Zealand have also been reported. Based on the most recent plenary document from the Ministry for Primary Industries (MPI), stock assessments have been conducted for three of the five *P. georgianus* Quota Management Areas (QMA's), TRE1, TRE2 and TRE7 where commercial fishing is focused (Fisheries Science Group, 2018). Within TRE1, two biological stocks (East Northland to Hauraki Gulf and Bay of Plenty, see Figure 5.1) are believed to exist based on time-series of catch at-age data. However, standardised bottom trawl catch per unit effort (CPUE) data and aerial sighting data produced conflicting results in age-based total catch-history assessment models, which meant no revisions were made to the stock assessments (Fisheries

Science Group, 2018).

Although TRE1 continues to be treated as one stock, an age based total catch-history assessment model became the basis for the Northern Inshore Working Group's recommendation to undertake further assessments of these areas (Fisheries Science Group, 2018). Modelling results found that stock abundance in TRE2 fluctuates with movement to and from this QMA, indicating that fish in TRE2 are likely to be part of the same biological stock as the Bay of Plenty (within TRE1) (Fisheries Science Group, 2018). Based on tag-recapture data by James (1978) of many fish, movement of *P. georgianus* was observed between these two regions. Stock assessments for TRE2 were also recommended by Fisheries Science Group (2018) to be conducted in parallel with TRE1 stock assessments.

Spatial heterogeneity of age composition and trends in catch per unit effort (CPUE) data of *P. georgianus* from TRE7 suggest the presence of three biological stocks; Ninety Mile Beach, the core area of the fishery from Tauroa Point to North Taranaki Bight and South Taranaki Bight (Fisheries Science Group, 2018). A stock assessment for TRE7 was updated in 2015, although it was limited to only one part of one of these putative sub-stocks (Kaipara-Manukau-North Taranaki Bight, see Figure 5.1) (Langley et al., 2015). These spatial stratifications were not included in stock models because the catch-at-age data was limited (Fisheries Science Group, 2018). No further information is available regarding the population structure of *P. georgianus* in New Zealand.

5.1.2 Population genetics

Understanding the population structure and taxonomy of fisheries species is an important basis for fisheries management (Laikre et al., 2005; Waples et al., 2008). It minimises the risk of genetic depletion and inbreeding depression which is crucial to the ability for species to adapt to environmental pressures (Allendorf and Ryman (2002); Laikre et al. (2005)) (see Section 1.4). Currently, the management of New Zealand *P. georgianus* fisheries relies on the assumption that *P. georgianus* within TRE1, TRE2 and TRE7 comprise single biological stocks of a single species.

Population genetic analyses can be used to investigate the population structure of species by constructing haplotype networks (Paradis, 2018), testing for distinct stocks (Excoffier et al., 1992) and testing for isolation by distance (Meirmans et al. (2012); Wright (1943)). It can also be used to understand the demographic history of a population or species. This will indicate how a population has responded to historical events, and in turn, provide information on how the population will respond to future environmental pressures (see Section 1.3.2).

The control region is the most variable region of the *P. georgianus* mitogenome (see Section 2.3.2) and is therefore a useful DNA target for intraspecific analyses at the micro-evolutionary scale (Avisé et al.,

1987) (see Section 1.3). No studies have yet investigated the population genetics of New Zealand *P. georgianus*. However, such a study will provide insights into the population structure and demographic history of *P. georgianus* and will contribute to a robust management of *P. georgianus* fisheries.

5.2 Materials and methods

Partial control region sequences of *P. georgianus* sampled in New Zealand and Western Australia were produced using DNA extracted from *P. georgianus* tissue samples sampled in this thesis and primers designed in this thesis (see Chapter 3 for further information). Sequencing of low yield control region PCR products was carried out using an ABI 3730 Sequencer at Massey University Sequencing and Genotyping Services in Palmerston North, New Zealand. Although majority of the control region sequencing was carried out by Macrogen in Seoul, the Republic of Korea.

Data preparation was undertaken in Geneious version 11.1.5 (Kearse et al., 2012). The chromatogram files of the DNA sequences were visually checked for signal strength and clear base calling. Poor quality sequences were excluded from further analyses such as sequences with a low signal to background signal. Sequence alignments were constructed using the Geneious alignment tool with the default settings (Alignment type: global alignment with free end gaps, cost matrix: 65% similarity, gap open penalty: 12, gap extension penalty: 3, refinement iterations: 2) and manually trimmed to remove base pair calling noise and primer binding locations.

Two final NEXUS alignments were produced. The first included all sequences from New Zealand's North Island and Western Australia. A second alignment of five sequences for *P. georgianus* sampled from offshore New Zealand islands was created. These sequences were aligned separately since they were highly divergent and difficult to align to the control region sequences of *P. georgianus* sampled from the rest of New Zealand or Western Australia.

5.2.1 Population structure

The main alignment of control region sequences of *P. georgianus* sampled from New Zealand's North Island and Western Australia was used to make pairwise comparisons were made for each statistical area in New Zealand that was sampled. Pairwise comparisons were made using the pairwise difference distance and significance was established with 110 permutations.

The main alignment was also sub-set at the following levels to compare populations at different scales using haplotype networks and AMOVA's:

Figure 5.1: Putative sub-stocks of *P. georgianus* in New Zealand.

1. TRE2 and Bay of Plenty
2. Ninety Mile Beach (putative sub-stock 1), North Taranaki bight to Taurau (putative sub-stock 2) and South Taranaki Bight (putative sub-stock 3)
3. Quota Management Areas (TRE1, TRE2, TRE7)
4. New Zealand and Western Australia
5. Geraldton, East Augusta and Poison Creek
6. Juveniles, adults from New Zealand and adults from Western Australia

For the comparison of the New Zealand and Western Australian *P. georgianus* populations, a random sub-sample of the New Zealand population was taken to ensure the sample sizes were comparable.

Because the control region has a high mutation rate and samples were taken over nine years, haplotype networks were chosen to be constructed over a phylogeny since some of the sequences could be ancestral to other sequences (see Paradis (2018)). Furthermore, there was large amount of variation in how different algorithms such as estimated unobserved haplotypes, so a minimum spanning network algorithm was used to avoid estimating unobserved haplotypes while providing advantages over a minimum spanning tree method (see section 1.3.2.1).

Therefore, the presence of distinct populations was tested for by creating minimum spanning haplotypes networks (epsilon value of 0) for all population groupings in PopART version 1.7 (Leigh and Bryant, 2015). Tests for population differentiation was also done in PopART version 1.7 (Leigh and Bryant, 2015) using analysis of molecular variance (AMOVA). The presence of continuous genetic change in the populations (isolation by distance) was tested by correlating geographic distance with genetic distance between fourteen statistical areas in New Zealand's North Island and three locations in Western Australia. The shortest geographic distance between localities was measured along the coastline using the measurement tool in Google Maps.

5.2.2 Demographic history

The population structure identified in Section 5.2.1 was used to investigate the demographic history of *P. georgianus*. The control region sequence data was grouped into two significantly genetically distinct populations from two locations; New Zealand's North Island and Western Australia. The full NEXUS alignment was imported into DNA Sequence Polymorphism version 6.12.3 (Rozas et al., 2017) to identify all distinct *P. georgianus* haplotypes and create Arlequin input files.

Basic demographic measures were calculated in Arlequin version 3.5.2.2 (Excoffier and Lischer, 2010). The level and type of genetic variation within and between populations was calculated using pairwise

differences and haplotype frequency counts. Tajima's D and Fu's F statistics were calculated using 1000 simulated steps. Mismatch distributions were based on the demographic expansion model and calculated with pairwise differences and 100 bootstrap replicates.

Mutation-scaled population sizes and migration rates over time were calculated in Migrate version 4.4.3 (Beerli (1998); Beerli and Felsenstein (1999)) on a remote high-performance cluster. Because the number of control region sequences for *P. georgianus* sampled in New Zealand were much higher than those from Western Australia, New Zealand *P. georgianus* sequences were sub-sampled in each model run using a random seed. Three migration models were explored:

1. Free migration in both directions between the two populations ($\text{AUS} \Leftrightarrow \text{NZ}$)
2. Migration in only one direction from Western Australia to New Zealand ($\text{AUS} \Rightarrow \text{NZ}$)
3. Migration in only one direction from New Zealand to Western Australia ($\text{NZ} \Rightarrow \text{AUS}$)

These three models were run several times using different prior distributions and MCMC settings. Iterative model runs were undertaken until the whole posterior distribution of the mutation-scaled population sizes and immigration rates were sampled, the acceptance ratio of each parameter was roughly half, an adequate MCMC sample size was obtained and the autocorrelation of MCMC sampling was limited. The parameters of the final three model runs are outlined in Table 5.1. The three final model runs were compared to find the migration model with the highest support. The demographic history of both New Zealand and Western Australian populations of *P. georgianus* were estimated from the highest supported model.

5.3 Results

5.3.1 Genetic data

The first NEXUS alignment covered a 640 base-pair region of the control region (see Figure 5.2). In this alignment, no base pairs were manually edited because all sequences were high quality. This alignment was composed of control region sequences of 384 *P. georgianus*, 68 of which were sampled from three locations in Western Australia and 316 from off the coast of the North Island of New Zealand. The New Zealand sequences included thirteen juvenile *P. georgianus* caught from one location in Whangarei. Of the 303 adult *P. georgianus* from New Zealand, 139 were sampled from TRE1, 58 from TRE2 and 106 from TRE7. These samples were taken from fourteen statistical areas around the North Island of New Zealand. The 68 samples from Western Australia were comprised of 27 *P. georgianus* from Poison Creek, thirteen from Geraldton and 28 off East Augusta (see Table 5.13).

Table 5.1: Parameters of the final three migration model runs.

Parameter		Value
Data options		
Transition - transversion ratio		6.9285
Base-pair frequency		Calculated from data
Sequencing error rate		0
Parameter start setting		
Parameter start values		F_{ST} -type calculation
Mutation rate		Constant
F_{ST} model type		Theta
MCMC search strategies		
Bayes update frequency	0.4 (0.6 for free migration model)	
Bayes proposals (theta)	Metropolis-Hastings sampler	
Bayes proposals (migration)	Metropolis-Hastings sampler	
Bayes prior distribution (theta)	Uniform (min = 0.15, max = 0.70)	
Bayes prior distribution (migration)	Uniform (min = 0.00, max = 4000)	
Long chains		1
Number of updates not recorded	3000 (6000 for free migration model)	
Number of sampled updates		50,000
Burn-in		20,000
Target acceptance ratio		0.44
Lowest allowed migration value		0.001
Complete run replicates		4

The second alignment covered a 585 base-pair region of the control region (see Figure 5.2) of three *P. georgianus* from the Kermadec Islands and two *P. georgianus* from Three Kings Islands. The strength of the signal of these sequences were generally lower and uncertain bases were edited conservatively. See the Appendix for both partial control region alignments (Figure A.5 and A.8).

5.3.2 Pairwise comparisons

The F_{ST} values among the majority of the sampled statistical areas in New Zealand were not significantly different (see Table: 5.2). However, *P. georgianus* from five pairs of statistical areas were significantly different from one another (see Figure: 5.3). However, the genetic distance among them was typically small and the populations reside in regions isolated from one another, such as the West and East of New Zealand's North Island. Some moderate genetic distances were identified (F_{ST} : 0.18474, p -value: 0.01802+-0.0121 between statistical area 040 and 009) (F_{ST} : 0.10996, p -value: 0.00901+-0.0091 between statistical area 039 and 009) but were based on a very small number of samples so the relevance of these findings cannot be validated (see Figure: 5.3).

As demonstrated in Figure: 5.3, some genetically distinct populations were identified with small genetic distances. *P. georgianus* from statistical area 041 off Raglan on the West Coast and 009 in the Bay of Plenty on the East Coast were significantly genetically distinct with an F_{ST} of 0.02698 (p -value: 0.00901+-0.0091). However, statistical area 047 off the North Cape and 041 in the North Taranaki

Figure 5.2: The alignment positions of two *P. georgianus* alignments showing the level of within species variation based on thirteen broodstock *P. georgianus* (blue) and 304 *P. georgianus* sampled in New Zealand (black).

Bight (F_{ST} : 0.02698, p -value: 0.00901+/-0.0091) are both located on the West Coast of New Zealand's North Island.

Table 5.2: Pairwise F_{ST} 's of *P. georgianus* among fourteen statistical areas in New Zealand. F_{ST} 's which are significant are highlighted in yellow.

Statistical Area	002	003	008	009	010	011	013	016	039	040	041	042	046	047
002	-	-	-	-	-	-	-	-	-	-	-	-	-	-
003	0.00031	-	-	-	-	-	-	-	-	-	-	-	-	-
008	-0.00829	-0.00006	-	-	-	-	-	-	-	-	-	-	-	-
009	0.00226	0.00511	-0.00277	-	-	-	-	-	-	-	-	-	-	-
010	-0.00294	-0.00516	-0.00908	0.02235	-	-	-	-	-	-	-	-	-	-
011	-0.01035	-0.00085	-0.01543	-0.00496	-0.00293	-	-	-	-	-	-	-	-	-
013	0.00098	0.00546	-0.00736	0.00141	-0.00109	-0.00020	-	-	-	-	-	-	-	-
016	-0.11890	-0.08608	-0.16346	-0.04707	-0.12322	-0.14786	-0.12545	-	-	-	-	-	-	-
039	0.03878	0.05844	0.03833	0.10996	0.05128	0.04707	0.04411	-0.04167	-	-	-	-	-	-
040	0.04139	0.03653	0.03953	0.18474	-0.01391	0.09031	0.03875	0.23077	0.08861	-	-	-	-	-
041	-0.00099	0.00973	-0.00099	0.03387	-0.00219	0.00220	0.00749	-0.11034	0.03223	0.00253	-	-	-	-
042	-0.00835	-0.01070	-0.01619	-0.00282	-0.01073	-0.01091	-0.01252	-0.10419	0.03377	0.03377	-0.00445	-	-	-
046	-0.02558	-0.02429	-0.02958	0.0129	-0.04694	-0.01422	-0.02812	-0.06875	-0.01454	-0.04032	-0.03627	-0.03615	-	-
047	0.00594	-0.00373	-0.00083	-0.01393	0.01609	-0.00213	0.00051	-0.02969	0.05640	0.13669	0.02698	-0.01115	-0.01683	-

5.3.3 Test for isolation by distance

Scatterplots and linear regression models did not find evidence to suggest that geographic distance between samples predicts genetic distance of *P. georgianus* in either New Zealand's North Island, or Western Australia (see Figure: 5.4).

Three of the thirteen juveniles had the same haplotype as adult *P. georgianus* sampled from three locations in the East Coast of the North Island of New Zealand; Off Mangonui in Northland, Te Kaha in the Bay of Plenty and Hicks bay in East Cape (See Figure 5.5). There were a further fifteen adult individuals with one SNP difference to a juvenile from Whangarei in the North-West and North-East of

Figure 5.3: Statistical areas from which *P. georgianus* are genetically distinct (F_{ST} 's are provided).

the North Island of New Zealand. A further eighteen adults had haplotypes with two SNP differences to a juvenile, ranging as far south as the Wellington Region. Adult *P. georgianus* with the same haplotype as a juvenile were found close to the juvenile sampling location in the north-east of New Zealand's North Island. The number and geographic range increased with increasing genetic differences from a juvenile.

5.3.4 Test for distinct populations

A haplotype network of *P. georgianus* from the Bay of Plenty and TRE2 show these fish are highly genetically connected (see Figure 5.6) and an AMOVA did not find evidence to suggest they are distinct stocks (see Table 5.3).

Within TRE7, the haplotype lineages among the three putative sub-stocks did not appear to be distinct (see Figure 5.7). However, an AMOVA analysis found evidence to suggest that at least one pair of putative sub-stocks were significantly genetically distinct at the 10% significance level (F_{ST} : 0.01004, p -value: 0.059) (see Table 5.4). Pairwise comparisons identified that the first putative sub-stock (Ninety Mile Beach) was distinct from the third putative sub-stock (South Taranaki Bight) (F_{ST} : 0.03019,

Figure 5.4: Relationship between geographic and genetic distance among *P. georgianus* from New Zealand and Western Australia.

p -value: 0.039) (see Table: 5.5).

The control region sequences of *P. georgianus* from the Three Kings Islands and Kermadec Islands were highly divergent from the control region sequences of *P. georgianus* sampled around the North Island of New Zealand and Western Australia. Each of the five fish were distinct haplotypes. There were only five to eight single nucleotide differences between haplotypes from these regions. However, one fish sampled from the Three Kings Islands was highly divergent, with 25 single nucleotide differences from the its most similar haplotype from the Kermadec Islands (see Figure 5.8).

There was some subtle grouping of haplotypes of *P. georgianus* in New Zealand, however the groupings were not associated with QMA's (see Figure 5.9). Instead, the haplotypes sampled from each of these three QMA's were highly genetically linked. An AMOVA did not find evidence to suggest *P. georgianus* haplotypes from TRE1, TRE2 and TRE7 are genetically distinct (p -value: 0.498, see Table 5.6).

Haplotypes of juvenile *P. georgianus* from Whangarei were similar to the haplotypes of adult *P. georgianus* sampled from New Zealand's North Island. However, a few juvenile *P. georgianus* had haplotypes similar to adult *P. georgianus* from Western Australia (see Figure 5.10). No evidence was found to suggest that this juvenile population was genetically different from the adult *P. georgianus* sampled in New Zealand's North Island (F_{ST} : 0.00032, p -value = 0.328), however they were significantly different from *P. georgianus* sampled from Western Australia (F_{ST} : 0.03735, p -value < 0.001) (see Table 5.7).

Within the Western Australian population of *P. georgianus*, there was a high level of haplotype connectivity and no obvious grouping of haplotypes associated with any of the three sampling locations (see Figure 5.11). This was supported by an AMOVA that did not find evidence to suggest a ge-

Figure 5.5: Sampling locations of adult *P. georgianus* with the same haplotypes as juveniles caught in Whangarei (top) or haplotypes with one (middle) or two (bottom) single nucleotide differences.

netic difference between *P. georgianus* from these locations (see Table 5.8). Interestingly, one pair of identical haplotypes were sampled from the two most distant sampling locations in Western Australia (Geraldton and Poison Creek, a minimum of roughly 1450 kilometers of coastline).

Several *P. georgianus* haplotypes in New Zealand and Western Australia were linked, although the haplotypes displayed some general grouping associated with these two sampling regions (see Figure 5.8).

Figure 5.6: Haplotype network and sampling locations of *P. georgianus* sampled from the Bay of Plenty and TRE2.

Figure 5.7: Haplotype network and sampling locations of *P. georgianus* from three putative substocks within TRE7.

Figure 5.8: Haplotype network and sampling locations of *P. georgianus* from Three Kings Islands and Kermadec Islands.

5.3.5 Demographic history

For the New Zealand and Western Australian populations of *P. georgianus*, there was a high level of haplotype diversity (0.9992 and 0.991 respectively) (see Table 5.10) compared to other New Zealand fishery species such as Blue Cod (*Parapercis colias*) (Gebbies (2014); Smith (2012)), Kahawai (*Arripis trutta*) (Hodgson, 2011). However, *P. georgianus* had a similar level of haplotype diversity to Spotty wrass (*Notolabrus celidotus*) (Scott (2010); Wilcox (2015)), Tarakihi (*Nemadactylus macropterus*) and Australasian Snapper (*Pagrus auratus*). Both the New Zealand and Western Australian populations of *P. georgianus* also had a high level of nucleotide diversity (0.0343 and 0.0296 respectively), higher than all the above-mentioned species.

Most *P. georgianus* were a distinct haplotype and there was no haplotype sharing between New Zealand and Western Australian populations. Within New Zealand, identical or similar haplotypes were often sampled from broadly different geographic regions around New Zealand's North Island. For example, four *P. georgianus* sampled from TRE1, TRE2 and TRE7 had identical haplotypes.

The high level of both nucleotide and haplotype diversity placed both *P. georgianus* populations in the fourth demographic class (see Figure 5.13) along with New Zealand populations of Tarakihi

Figure 5.9: Haplotype network and sampling locations of Trevally from three Quota Management Areas in New Zealand.

Figure 5.10: Haplotype network and sampling locations of juvenile *P. georgianus* from Whangarei (New Zealand), adult *P. georgianus* from Western Australia and adult *P. georgianus* from New Zealand.

Figure 5.11: Haplotype network and sampling locations of *P. georgianus* from three locations in Western Australia.

Figure 5.12: Haplotype network and sampling locations of *P. georgianus* from New Zealand and Western Australia.

Table 5.7: AMOVA output comparing juveniles from Whangarei to adults in New Zealand and adults in Western Australia.

	Degrees of freedom	Sum of squares	Sigma squared	Variation	FST	Significance
Juveniles and adults in New Zealand						
Among populations	1	241.352	0.071	0.03236	0.00032	0.328
Within populations	315	68950.907	218.892	99.96764	-	-
Total	316	69192.259	218.963	-	-	-
Juveniles and adults in Western Australia						
Among populations	1	687.286	6.437	3.73524	0.03735	< 0.001
Within populations	79	13105.578	165.893	96.26476	-	-
Total	80	3792.864	172.330	-	-	-

Table 5.8: AMOVA output comparing *P. georgianus* from three regions in Western Australia.

	Degrees of freedom	Sum of squares	Sigma squared	Variation	FST	Significance
Among populations	2	281.164	-0.757	-0.45742	-0.00457	0.647
Within populations	65	10811.644	166.333	100.45742	-	-
Total	67	11092.809	165.576	-	-	-

Table 5.9: AMOVA output comparing *P. georgianus* from New Zealand and Western Australia.

	Degrees of freedom	Sum of squares	Sigma squared	Variation (%)	FST	Significance
Among populations	1	1432.912	8.768	3.5169	0.03517	<0.001
Within populations	134	32230.926	240.529	96.4831	-	-
Total	135	33663.838	249.297	-	-	-

(*Nemadactylus macropterus*) and Australasian Snapper (*Pagrus auratus*). Overall, the genetic diversity was lower for *P. georgianus* sampled from Western Australia, possibly as a result of the smaller sample size and number of locations sampled from this population.

The genetic diversity was composed predominantly of substitutions (40% for the New Zealand population and 27% for the Western Australian population) and transitions (34.06% for the New Zealand population and 20.31% for the Western Australian population) (see Table 5.10). Fewer nucleotide positions were transitions and even fewer were insertions and deletions (indels).

The New Zealand population of *P. georgianus* had a negative Tajima's D (D : -1.50612, p -value: 0.01800) and Fu's F (F : -23.54376, p -value: 0.01100). Similarly, the Western Australian population of *P. georgianus* had a negative Tajima's D (D : -1.27903, p -value: 0.086) and Fu's F (F : -24.11497, p -value < 0.00001). These were all significant at the 5% significance level, except for the Fu's F statistic for the Western Australian population.

The observed and expected mismatch distributions for New Zealand *P. georgianus* population were not significantly different (p -value: 0.97) or the Western Australian population (p -value: 0.06). Under the sudden expansion model, no evidence was found to suggest a ragged multimodal mismatch distribution for New Zealand *P. georgianus*. In contrast, the mismatch distribution of Western Australian *P.*

georgianus was significantly ragged (Harpending's Raggedness index: 0.00454591, p -value: 0.02).

When the support for the free migration model ($\text{AUS} \Leftrightarrow \text{NZ}$) was compared to the $\text{NZ} \Rightarrow \text{AUS}$ (85.51668) and $\text{AUS} \Rightarrow \text{NZ}$ models (79.48104), a positive log Bayes factor was returned indicating the most highly supported model was the $\text{AUS} \Leftrightarrow \text{NZ}$ model. For MCMC sampling, the acceptance ratios of this model was roughly half (see Table 5.11) and there was a large sample size for all model parameters (see Table 5.12). The autocorrelation between MCMC sampled updates for theta1 and theta2 was minimal (see Table 5.12). However, there was some autocorrelation between the sampled updates of both migration rates (0.73492 and 0.80200) that was not fully resolved by lowering the update frequency of the MCMC sampling. The distributions of the sampled posterior probabilities of all four estimated parameters (population size of New Zealand and Western Australia over time (theta 1 and 2) and migration each direction over time (migration AUS to NZ and migration NZ to AUS)) can be found in the Appendix (see Figure A.9).

The skyline plots produced in the migration analysis indicate that both populations have undergone size expansions at similar rates (see Figure 5.15). It also indicates that both populations had a historical peak in bi-directional migration rates. However, these migration rates slowed more recently coinciding with the population expansions. Today, the migration rates between the two populations is estimated to be nearly zero.

Table 5.10: Nucleotide and haplotype diversity of New Zealand and Western Australian populations of *P. georgianus* based on partial control region sequences.

Population	Sample size	Overall diversity				Types of genetic diversity (by percentage of sites)			
		Nucleotide diversity	Haplotype diversity	Polymorphic sites	Average number of differences between haplotypes	Transitions	Transversions	Substitutions	Insertions and Deletions (indels)
New Zealand	304	0.034308 +/- 0.016773	0.9992	41.88%	21.888809 +/- 9.674026	34.06%	12.50%	40.00%	3.59%
Western Australia	68	0.029591 +/- 0.014716	0.9991	22.81%	18.73094 +/- 8.402724	20.31%	2.50%	21.72%	1.72%

Table 5.11: Acceptance ratios for all parameters of the final migration model run.

Parameter	Accepted changes	Ratio
theta1	72948227/180015806	0.40523
theta2	87054958/180014346	0.48360
migration2-1	86810388/179986353	0.48232
migration1-2	72461286/179978625	0.40261
Genealogies	4893231/480004870	0.01019

Figure 5.13: Demographic classes of New Zealand and Western Australian *P. georgianus* compared to several other New Zealand fishery species.

Figure 5.14: Mismatch distribution for both populations of *P. georgianus* including confidence intervals.

Figure 5.15: Skyline plots showing modelled changes in population sizes and migration rates over time of New Zealand and Western Australian *P. georgianus* populations including error bars (1.96 x approximate standard error).

Table 5.12: MCMC autocorrelation and effective sample sizes for all parameters of the final migration model run.

Parameter	Autocorrelation	Effective sample size
theta1	0.33483	100009.44
theta2	0.36133	94568.10
migration2-1	0.73492	32739.71
migration1-2	0.80200	23756.45
Ln(Prob(D G))	0.88393	12321.71

5.4 Discussion

5.4.1 Population structure and gene flow

Within New Zealand

Overall, *P. georgianus* in New Zealand was not structured clearly as distinct populations, nor were they isolated by distance. Instead, the New Zealand population of *P. georgianus* was largely panmictic with a high level of genetic connectivity of individuals among broad geographic regions around New Zealand's North Island. For example, for the three major commercial Quota Management Areas (TRE1, TRE2 and TRE7), no evidence was found to suggest that *P. georgianus* are genetically distinct among these QMA's.

In saying this, the comparisons of *P. georgianus* at a smaller geographic scale identified some populations that were genetically distinct. These were often sampled from geographically remote regions such as the Bay of Plenty on the East Coast and North/South Taranaki Bight on the West Coast. However, one pair of genetically distinct populations was identified on New Zealand's West Coast (North Taranaki Bight and North Cape to Tauroa point). Despite the significant genetic differentiation of these populations and the geographic remoteness of the populations, the genetic distance between them was small to moderate (F_{ST} between 0.02 and 0.18). However, the sample of these some population comparisons were small, so their biological relevance is unknown.

The proposition that *P. georgianus* in the Bay of Plenty is the same biological stock as TRE2 by Fisheries Science Group (2018) and James and Stephenson (1974) was supported by a lack of genetic differentiation in haplotype and AMOVA analyses between fish from these regions. Extending beyond this, the control region sequence data indicates that this lack of stock differentiation could extend further than the above-mentioned regions. For example, there was no evidence to suggest that other regions within TRE1 ranging up to East Northland are genetically distinct from TRE2. Therefore, no evidence was found to dispute the decision to combine TRE1 and TRE2 in future stock assessments (see Fisheries Science Group (2018)).

Table 5.13: Information on all *P. georgianus* samples for which the control region was sequenced.

Sampling location			Sample source	Tissue type	Sampling date	Exact GPS known	Number of samples
Country	QMA	Statistical area					
New Zealand	TRE1	area 002	Moana Auckland	Caudal fin	03-09-18	No	17
New Zealand	TRE1	area 002	MPI	Caudal fin	30-08-18	Yes	10
New Zealand	TRE1	area 002	MPI	Caudal fin	31-08-18	Yes	11
New Zealand	TRE1	area 003	Darren Parsons	Whole fish	08-03-18	No	13
New Zealand	TRE1	area 003	Moana Auckland	Caudal fin	05-08-18	No	33
New Zealand	TRE1	area 008	Moana Auckland	Caudal fin	04-09-18	No	28
New Zealand	TRE1	area 009	Fishers Whangarei	Muscle	26-07-17	Yes	6
New Zealand	TRE1	area 009	MPI	Caudal fin	20-09-18	Yes	9
New Zealand	TRE1	area 010	MPI	Caudal fin	15-09-18	Yes	3
New Zealand	TRE1	area 010	MPI	Caudal fin	16-09-18	Yes	3
New Zealand	TRE1	area 010	MPI	Caudal fin	21-09-18	Yes	8
New Zealand	TRE1	area 010	MPI	Caudal fin	22-09-18	Yes	11
New Zealand	TRE2	area 011	Moana Auckland	Caudal fin	04-11-18	No	28
New Zealand	TRE2	area 013	Moana Auckland	Caudal fin	05-09-18	No	11
New Zealand	TRE2	area 013	Moana Auckland	Caudal fin	05-11-18	No	2
New Zealand	TRE2	area 013	Moana Auckland	Caudal fin	06-11-18	No	17
New Zealand	TRE2	area 016	Te Papa	Muscle	Apr 2009	Yes	1
New Zealand	TRE7	area 039	Moana Auckland	Caudal fin	02-02-19	No	4
New Zealand	TRE7	area 040	Te Papa	Muscle	04-12-09	Yes	2
New Zealand	TRE7	area 041	MPI	Caudal fin	15-02-19	Yes	9
New Zealand	TRE7	area 041	MPI	Caudal fin	16-11-18	Yes	5
New Zealand	TRE7	area 041	MPI	Caudal fin	19-09-18	Yes	5
New Zealand	TRE7	area 041	MPI	Caudal fin	23-10-18	Yes	6
New Zealand	TRE7	area 041	MPI	Caudal fin	24-10-18	Yes	5
New Zealand	TRE7	area 041	Te Papa	Muscle	19-12-09	Yes	2
New Zealand	TRE7	area 041	Te Papa	Muscle	23-12-09	Yes	3
New Zealand	TRE7	area 041	Te Papa	Muscle	06-10-09	Yes	1
New Zealand	TRE7	area 042	MPI	Caudal fin	10-09-18	Yes	1
New Zealand	TRE7	area 042	MPI	Caudal fin	11-08-18	Yes	3
New Zealand	TRE7	area 042	MPI	Caudal fin	11-09-18	Yes	2
New Zealand	TRE7	area 042	MPI	Caudal fin	11-11-18	Yes	2
New Zealand	TRE7	area 042	MPI	Caudal fin	12-08-18	Yes	4
New Zealand	TRE7	area 042	MPI	Caudal fin	14-11-18	Yes	1
New Zealand	TRE7	area 042	MPI	Caudal fin	15-11-18	Yes	1
New Zealand	TRE7	area 042	MPI	Caudal fin	16-11-18	Yes	1
New Zealand	TRE7	area 042	MPI	Caudal fin	19-09-18	Yes	1
New Zealand	TRE7	area 042	MPI	Caudal fin	19-10-18	Yes	1
New Zealand	TRE7	area 042	MPI	Caudal fin	20-09-18	Yes	3
New Zealand	TRE7	area 042	MPI	Caudal fin	20-10-18	Yes	1
New Zealand	TRE7	area 042	MPI	Caudal fin	21-10-18	Yes	1
New Zealand	TRE7	area 042	MPI	Caudal fin	29-09-18	Yes	1
New Zealand	TRE7	area 042	MPI	Caudal fin	31-07-18	Yes	4
New Zealand	TRE7	area 042	MPI	Caudal fin	05-08-18	Yes	1
New Zealand	TRE7	area 042	MPI	Caudal fin	06-08-18	Yes	1
New Zealand	TRE7	area 042	MPI	Caudal fin	08-10-18	Yes	2
New Zealand	TRE7	area 046	Te Papa	Muscle	18-02-11	Yes	2
New Zealand	TRE7	area 046	Te Papa	Muscle	22-02-11	Yes	1
New Zealand	TRE7	area 046	Te Papa	Muscle	24-02-11	Yes	3
New Zealand	TRE7	area 047	Moana Auckland	Caudal fin	29-07-18	Yes	20
New Zealand	TRE7	area 047	Te Papa	Muscle	13-04-13	Yes	1
New Zealand	TRE7	area 047	Te Papa	Muscle	14-04-13	Yes	1
New Zealand	TRE7	area 047	Te Papa	Muscle	17-04-13	Yes	1
New Zealand	TRE7	area 047	Te Papa	Muscle	19-04-13	Yes	1
New Zealand	TRE7	area 047	Te Papa	Muscle	22-02-11	Yes	1
New Zealand	TRE7	area 047	Te Papa	Muscle	23-02-11	Yes	4
New Zealand	TRE10	area 092	Te Papa	Muscle	27-10-16	Yes	1
New Zealand	TRE10	area 092	Te Papa	Muscle	07-05-11	Yes	1
New Zealand	TRE10	area 092	Te Papa	Muscle	08-05-11	Yes	1
New Zealand	TRE10	area 093	Te Papa	Muscle	23-05-11	Yes	1
Australia	-	-	Murdoch University	Muscle	10-03-12	Yes	27
Australia	-	-	Murdoch University	Muscle	17-02-12	Yes	10
Australia	-	-	Murdoch University	Muscle	19-02-12	Yes	18
Australia	-	-	Murdoch University	Muscle	09-08-12	Yes	13

There was some inconsistency in the genetic support for the existence of three putative sub-stocks within TRE7 on New Zealand east coast (see Fisheries Science Group (2018)). An AMOVA anal-

ysis revealed that at least one pair of these putative sub-stocks is genetically distinct. Subsequent pairwise AMOVA's suggest that the genetically distinct populations are the Ninety Mile Beach (putative sub-stock 1) and South Taranaki Bight (putative sub-stock 3) populations. In contrast, pairwise comparisons of the statistical areas on New Zealand's west coast identified that fish from the North Taranaki Bight (putative sub-stock 2) are genetically distinct from fish from the North Cape to Tauroa point (a slight extension of the range of putative sub-stock 1). This difference in the findings could be a result of the small number of samples (six) obtained for *P. georgianus* in the third putative sub-stock (South Taranaki Bight), the differences among the two statistical methods used (AMOVA and F_{ST} comparisons) or the geographic range of fish included for comparisons

Nonetheless, the presence of stock-structuring of *P. georgianus* along the west coast of New Zealand's North Island was supported by the genetic data. However, it remains unclear how *P. georgianus* are structured in this region. Observations of spatial structuring of age composition in these three putative sub-stocks (Fisheries Science Group, 2018) may simply not cause genetic structuring or be a result of genetic structuring. Instead, these observations could be a result of behavioural factors that may not have bearing on the genetic population structure of *P. georgianus* within TRE7. Alternatively, the genetic data produced in this study may not have been able to clearly reject the null hypothesis of no population structuring due to sample size or the statistical methods used.

In order to investigate the latter, the number of control region sequences for *P. georgianus* from the South Taranaki Bight population could be bolstered. Additionally, producing genetic data for fish sampled from regions missed in this study such as the east coast of the Auckland region and Northland would help inform where the boundaries of any possible sub-stocks occur. Taking a genetic approach in future studies may enable a distinction of spatial stratifications that are or are not genetically based.

Interestingly, greater genetic distances were found within New Zealand than between New Zealand and Western Australia. For example, the genetic distance of two pairs of populations in New Zealand (Wellington Region and the Bay of Plenty (F_{ST} : 0.10996) and South Taranaki Bight and the Bay of Plenty (F_{ST} : 0.08474)) was greater than the genetic distance of the New Zealand and Western Australian populations (F_{ST} : 0.03517). This is in spite of the fact that the Western Australian population is a minimum of 4,600 kilometers away from the New Zealand population compared to a maximum of 1200 kilometers (along the coastline) between the genetically distinct populations within New Zealand.

In regards to the claim that a cryptic *Pseudocaranx* species (possibly *P. wrighti*) is occurring in the North Cape of New Zealand (see Duffy and Ahyong (2015) and Smith-Vaniz and Jelks (2006)), these fish were found to be distinct from *P. georgianus* from statistical area 041 off Raglan (F_{ST} : 0.02698, p -value: 0.00901+-0.0091) in pairwise comparisons of statistical areas. However, there was no evidence

to suggest that *P. georgianus* from the North Cape were genetically distinct from *P. georgianus* from the rest of New Zealand North Island. Previous research may be referring to the highly divergent population of fish sampled in this study from the Three Kings and the Kermadec Islands (discussed further below). However, this study did not identify this or any other cryptic species occurring close to the coast of New Zealand's North Cape.

It is important to emphasize that finding evidence of population structuring using the AMOVA approach is highly dependent on how the samples are grouped for population comparisons. There was limited prior knowledge of the population structure of *P. georgianus* in New Zealand to inform how samples should be grouped. Therefore, the population groupings used in this study relied upon fisheries management boundaries and putative sub-stocks identified by prior research. This could contribute to the inconsistencies in findings for population structuring of *P. georgianus* in New Zealand. In this case, large scale AMOVA comparisons such as comparing Quota Management Areas may not find evidence for population structuring but smaller scale comparisons might.

The population of juveniles used in this study were caught from one location within Whangarei at a single point in time. Surprisingly, there was no obvious genetic distinction of these juveniles from adult *P. georgianus* throughout New Zealand's North Island. These juveniles and adults had a similar level of within-group genetic variability and percentage identity. Additionally, they were both genetically distinct from *P. georgianus* in Western Australia with a similar genetic distance (F_{ST}). Firstly, this implies that this juvenile population received genetic input from only New Zealand *P. georgianus*. Secondly, the similar level of genetic variability as well as a high level of genetic connectivity to adult *P. georgianus* from broad regions in New Zealand North Island could reflect their batch spawning behaviour. This alludes to a high level of adult movement for spawning, resulting in this juvenile population.

However, further research would be required to determine the level of connectivity or differentiation of other juvenile populations to adult *P. georgianus* populations, particularly juveniles on the west coast of New Zealand. Such work could also provide information on the spawning locations and movement patterns of adult *P. georgianus*. It would clarify the results of this thesis by informing if there is an ongoing reproductive contact between *P. georgianus* in New Zealand and Australia and whether these populations represent distinct species (discussed further below).

P. georgianus are thought to undertake movements associated with the spawning season (Fisheries Science Group (2018)). This seasonal movement could obscure the population structure of New Zealand *P. georgianus* identified in this study since samples were taken over all four seasons. Further investigations of the fine-scale population genetic structure of *P. georgianus*, such as the three putative sub-stocks along the west coast of New Zealand's North Island, could benefit from accounting for sea-

son. For example, serial-sampling of populations over several seasons would indicate whether season influences the estimates of population genetic structure. This would also provide further information on the movements of New Zealand *P. georgianus*.

This study found no evidence for isolation by distance among adult *P. georgianus* in New Zealand. However, the genetic distance (number of SNP's) appears to increase with increased geographic distance from the juvenile population in Whangarei. In other words, there may be isolation by distance between the juvenile population in Whangarei and adults in New Zealand's North Island. This could reflect adults returning to the same spawning regions annually despite occupying many other regions of New Zealand for non-spawning purposes.

Although the literature indicates that *P. georgianus* often show limited movement, it also indicates that a few individuals will undergo long distance migrations (Fowler et al. (2018), James (1980)). This behaviour could maintain a genetic connection between geographically remote populations, particularly since only a few migrating individuals are required to prevent population genetic divergence ($\sim$ one migrant per generation) (Slatkin, 1987). This could mean that there is population structuring of *P. georgianus* in New Zealand that could be obscured by a few long-distance migrants. On the other hand, it highlights the genetic benefit of long-distance migrations in maintaining genetic diversity in New Zealand *P. georgianus* fisheries. This is an important biological component of the ability for *P. georgianus* populations to remain robust when undergoing commercial fishing pressure.

Offshore Islands

Interestingly, five control region sequences of *P. georgianus* sampled from the Three Kings and the Kermadec Islands were distinct from the control region sequences of *P. georgianus* from New Zealand's North Island as well as Western Australia. It appears that this genetic difference could indicate they are distinct species. These fish are genetically distinct despite being sampled as little as 56 kilometers from *P. georgianus* in Northern New Zealand. This is a distance which *P. georgianus* would certainly be able to traverse, given reports of migrations over hundreds of kilometers (see Fowler et al. (2018), James (1980)). If the fish from the offshore islands were the same species as New Zealand *P. georgianus*, a greater level of genetic similarity between these fish would be expected. This is pertinent considering the level of genetic similarity of *P. georgianus* from much more geographically remote regions such as New Zealand's North Island and Western Australia.

This has important implications for fisheries management. Although the TRE1 fishery includes the Three Kings Islands, at least for the fish sampled in this study, catch was often taken closer to the coastal regions of the North Cape for which this study found no evidence to suggest they are not

P. georgianus. This suggests that commercial fishing in TRE1 may not be landing fish from this genetically distinct population. Going forward, measures should be put in place to ensure this cryptic species occurring in the Three Kings Islands are not included in the total allowable catch (TAC) of TRE1 and are accounted for in TRE1 stock assessments.

Further research would be required to establish what species these genetically distinct fish are and what their geographic range is. Firstly, sequencing of the COI gene could be undertaken, allowing for phylogenetic comparison to the COI sequences available for four *Pseudocaranx* species (*P. georgianus*, *P. wrighti*, *P. dentex* and *P. dinjerra*). Since the COI gene appears to be consistent with taxonomic studies of the *Pseudocaranx* genus (see Chapter 4), this gene should be capable of identifying this cryptic species. Secondly, the control region of other taxonomically verified *P. wrighti* could be produced and tested for a genetic similarity to the fish sampled from Three Kings and the Kermadec Islands. This would verify whether these fish are *P. wrighti* as suggested by Duffy and Ahyong (2015) or are not *P. wrighti* as suggested by Smith-Vaniz and Jelks (2006). Importantly, the control region has been shown to be capable of distinguishing *P. georgianus* and *P. wrighti* by Bearham (2004). Unfortunately, no control region sequences are currently publicly available for *P. wrighti*, although short control region sequences of *P. wrighti* have been produced by Bearham (2004) which could be used if the sequences are accessible.

Supporting these genetic approaches with taxonomic analysis of these offshore island specimens will provide a taxonomic description of a possible undescribed species, further supporting a genetic classification of this cryptic species and provide more morphological data for *Pseudocaranx* species. The Three Kings Islands and Kermadec Island samples sequenced in this study are held in the Te-Papa museum collection as museum specimens. Muscle tissue has been preserved, allowing the possibility of further genetic sequencing of these specimens. Furthermore, the whole fish has been preserved allowing the potential for further morphological analyses. It is important to note that the PCR yield using the control region primers designed in this study was typically lower for fish sampled from the Kermadec and Three Kings Islands. This could be a result of their variability from other *P. georgianus* meaning these primers do not bind as optimally to the control region of fish from these regions. If further sequencing was undertaken on these fish, control region primers may need to be redesigned in order to optimize the PCR reactions.

Among countries

A significant genetic differentiation was found between *P. georgianus* from New Zealand's North Island and *P. georgianus* from Western Australia (F_{ST} : 0.03517, p -value <0.001). The genetic differentiation

was also reflected in one juvenile population sampled in Whangarei which had a similar level of genetic differentiation to adult *P. georgianus* in Western Australia (F_{ST} : 0.03735, p -value < 0.001). Despite the genetic differentiation of both adult and juvenile *P. georgianus* from New Zealand and adults in Western Australia, the genetic distance between these populations was small and many haplotypes from these populations were similar.

The focus then becomes how to interpret this genetic similarity between two geographically remote populations. This genetic similarity could be a result of ongoing gene flow between these two populations. Alternatively, it could result from a shared common ancestry where the two populations may not experience ongoing gene flow, but not enough time has passed since their separation from a common ancestor for full coalescence of the lineages to occur. Understanding the contribution of these factors to the level of current day observed genetic similarity between these populations is important to understand whether these populations represent distinct species and whether there is ongoing gene flow between *P. georgianus* fisheries in New Zealand and Australia.

On the COI gene (see Section 4.3) and the control region, there was no clear differentiation between *P. georgianus* from New Zealand and Western Australia. Despite the COI gene being less variable than the control region, a strong genetic similarity was still found between geographically remote *P. georgianus*. For example, fish from Gisborne on the East Coast of New Zealand and Victoria and Geographe Bay in Western Australia were closely related with high statistical support. This genetic similarity and the lack of clear separation of control region haplotypes of *P. georgianus* in New Zealand and Western Australia highlights the possibility of these populations being the same species still experiencing gene flow.

On the other hand, the best fit Bayesian migration model estimated that there was a peak of bi-directional migration between the two populations. However, it also estimated that there has been a reduction in this migration to almost zero recently. This suggests that there is no current migration occurring between the two populations. Furthermore, historical migration among the two populations could explain the genetic similarity of the two populations. A relatively recent history of migration between the two populations could have prevented a full coalescence of the two populations. This would cause a low level of genetic differentiation (small F_{ST} value) even if the populations have been separated for some time.

To further interpret the genetic connection of *P. georgianus* from these two geographically remote populations, we should frame it within the typical movement behaviour of *P. georgianus* and their capacity for long distance migrations. The sheer geographic distance between these two populations (a minimum of 4,600 kilometers) suggests that the genetic similarity of these populations has arisen from common ancestry of these populations (ie. no ongoing migration or gene flow between the populations).

Although little is provided in the literature of *P. georgianus* movements, particularly *P. georgianus* in New Zealand, no studies have reported *P. georgianus* undergoing long-distance migrations of this scale. The maximum distance recorded for the only two tag-recapture studies was 246 kilometers for New Zealand *P. georgianus* (James, 1980) and 508 kilometers for Australian *P. georgianus* (Fowler et al., 2018). In the past, taxonomic analyses of *P. georgianus* have overestimated the importance of geographic remoteness (James and Stephenson, 1974). In turn, their migration capability has been underestimated, resulting in what James and Stephenson (1974) argues was an inappropriate taxonomic separation of geographically remote populations. There's no doubt that these populations are geographically remote and distinct populations. However, the answer for whether these populations represent distinct species is less clear.

What appears more likely is that these populations are the same species that are isolated by distance over a larger geographic scale. Although this study did not find evidence for isolation by distance among the adult *P. georgianus* of either population, *P. georgianus* are distributed in far broader regions than this study was able to capture. They are known to occur from Western and Southern Australian waters to the East Coast of New Zealand to around New Zealand's South Island (Smith-Vaniz and Jelks, 2006). There also occur in several intermediary locations between New Zealand and Western Australia such as Lord Howe Island, Norfolk Island and the East Coast of Australia (Smith-Vaniz and Jelks, 2006). Future genetic studies capturing *P. georgianus* from this broader geographic range may inform if *P. georgianus* is in fact isolated by distance. This may also inform if isolation by distance is a factor in the genetic similarity of geographically remote populations such as the New Zealand and Western Australian populations of *P. georgianus*.

Given that this thesis found some evidence to suggest that there is some population structuring of New Zealand *P. georgianus*, it's unlikely that a lack of population structure would be concluded by future population genetic research on this species. However, in order to confidently exclude this possibility, a large number of samples from a broad range of locations, analysis of more several genetic loci and other data beyond genetic data would be required (Laikre et al., 2005).

Western Australia

Within Western Australia, no population structuring was identified among the three sampling regions, despite their range over a considerable geographic distance (roughly 630km from Geraldton to East Augusta, and a further 801 kilometers between East Augusta and Poison Creek). This is in conflict with findings by Bearham (2004) who found a significant genetic difference in the control region between *P. georgianus* from Geraldton and those sampled from offshore Perth ($F_{ST} = 0.01493$, $p\text{-value} = 0.0176$)

and New South Wales ($\phi_{ST} = 0.00968$, p -value < 0.001). *P. georgianus* samples were not taken or compared from Perth or New South Wales populations, however fish from Geraldton were not found to be genetically distinct from *P. georgianus* from East Augusta or Poison Creek which are geographically more remote than Geraldton and Perth (roughly 400 kilometers away).

5.4.2 Demographic history of *P. georgianus*

The skyline plots indicated that there was an expansion in the size of both populations coinciding with a reduction in bi-directional migration between the two populations. Both populations also had a high ratio of haplotype and nucleotide diversity composed primarily of single nucleotide polymorphisms. However, the results of the Tajima's D statistic, Fu's F statistic and mismatch distributions were less clear. For the New Zealand population, a significantly negative Tajima's D and Fu's statistics combined with a high ratio of haplotype and nucleotide diversity and low number of insertions and deletions pointed to a population expansion. Furthermore, no evidence was found to suggest the mismatch distribution for the New Zealand *P. georgianus* population was multimodal.

For the Western Australian population, the negative Tajima's D statistic was not significant, but the Fu's F statistic was significantly negative. This difference could be due to the higher sensitivity of the Fu's F statistic. In contrast to the New Zealand population, evidence was found for a multimodal mismatch distribution for the Western Australian population. It appears that the Western Australian population may have also undergone a historical population expansion given the high level of haplotype to nucleotide diversity and low number of insertions and deletions. However, its population size could be more stable than the New Zealand population given the multimodal mismatch distribution.

Simulation studies have shown that a unimodal mismatch distribution can result from large spatial expansions if there is enough migration between sub-populations (Ray et al., 2003). The migration analysis estimated that such historical migration has occurred between the New Zealand and Western Australian populations. Therefore, these demographic signals may be due to a range expansion of an ancestral *P. georgianus* population. In particular, this may be the case for the New Zealand *P. georgianus* population since all demographic measures pointed to a population expansion. To give one example, the availability of new habitat could cause a range expansion and an associated increase in the size of the New Zealand population, or both populations. This would have coincided with a reduction in migration and gene flow between the populations. However, it is important to mention that there was some autocorrelation between MCMC sampled estimates of migration rates, meaning that some caution should be taken when making inferences from this migration analysis.

Because no mutation rate was available for *P. georgianus* and none was able to be acquired for any

closely related Carangidae species, the mutation scaled estimates of population sizes and migration rates were left scaled by mutation to avoid wildly inaccurate time estimates. However, if a mutation rate was able to be obtained for *P. georgianus* or a close relative, it could be used to estimate the timing of important historical changes in population size and migration rates. This could be then compared to geologic or climate events to create hypotheses on the causes for these demographic signals. This would be particularly interesting since several other New Zealand fishery species show genetic signals of population size expansions (eg. Spotty wrass (*Notolabrus celidotus*) [Scott (2010); Wilcox2015], Tarakihi (*Nemadactylus macropterus*) and Australasian Snapper (*Pagrus auratus*)) hypothesised to be a result of increases in usable habitats after the last ice age.

Chapter 6

General discussion

This thesis is a first genetic investigation into New Zealand *P. georgianus*. Investigating the evolutionary relationships and population genetics of this key fisheries species has provided important biological and fisheries management insights regarding the taxonomy of the *Pseudocaranx* genus, the population structure and demographic history of New Zealand *P. georgianus*. Additionally, it has provided a valuable basis from which genetic or genomic tools can be applied to this or other fishery species in the future.

This thesis has contributed to the growing species “bar-coding” database (Ward et al. (2009)) by producing the first COI sequences published for New Zealand *P. georgianus*. This contribution to a straightforward and standardised database allows for easy comparisons to the results of older genetic studies. The first control region sequence data has been produced for New Zealand *P. georgianus*, providing genetic data that can be reanalysed or extended in future research. Furthermore, the utility of two mitochondrial gene regions, the COI gene and the control region, for investigating the phylogenetics and population genetics of New Zealand *P. georgianus* has been demonstrated.

6.1 Key findings

6.1.1 *P. georgianus* mitogenome

The first *P. georgianus* mitogenome is assembled and described in this thesis. The *P. georgianus* mitogenome was found to be typical of ray-finned and cartilaginous fish in terms of its overall structure, gene region lengths and stop and start codons.

Using the assembly of the *P. georgianus* mitogenome from whole genome data in silico as an example, I demonstrated the influence a few key methodological choices have on the quality of mitochondrial data as well as some subtleties in how the assembly algorithm handles real life data. The most important determinant of mitogenome quality was the choice of reference mitogenome used in the assembly process, impacting the annotation of the final mitogenome and the resolution of uncertain regions.

I highlighted that mitogenomes of a taxonomically closely related species does not guarantee a robust mitogenome assembly. Instead, the reference mitogenome that is genetically similar is required. Furthermore, the level of genetic similarity of the reference genome required for a robust mitogenome will vary by DNA region due the differing mutational rates along mitogenomes.

The number of mapping iterations used in the assembly process also had an important impact on the quality of the assembly. Using a sufficient number of mapping iterations improved the coverage and identity of more difficult to resolve regions. However, a large number of mapping iterations was unable to mitigate the effects of choosing a poor reference mitogenome and could increase the incorporation of sequencing errors.

Overall, this investigation into a few key methodological choices during the assembly of mitogenomes from whole genome data in silico provides a reminder to keep a genetic perspective when dealing with genetic data, investigate the quality of the assembly and report confidence in mitogenome data produced using these methods.

6.1.2 Evolutionary relationships of *Pseudocaranx* species and population genetics of *P. georgianus*

6.1.2.1 *Pseudocaranx* species occurring in New Zealand

The COI gene supported the current taxonomy of the *Pseudocaranx* genus pending taxonomic verification that the naming of two specimens sampled in Australia as *P. dentex* have not been updated (to *P. georgianus*) since the taxonomic revision of the *Pseudocaranx* genus by Smith-Vaniz and Jelks (2006). In both maximum likelihood and Bayesian phylogenies, *P. georgianus* from New Zealand and Western Australia are monophyletic, indicating that they are the same species.

In contrast, the more variable control region identified a small, but significant genetic difference between *P. georgianus* in New Zealand and *P. georgianus* in Western Australia (F_{ST} : 0.03517, p -value <0.001). Unfortunately, genetic data is unable to differentiate between common ancestry and ongoing gene flow. Hence, it remains unknown if this genetic difference indicates that these fish are distinct species or sub-populations.

The literature suggests that a cryptic *Pseudocaranx* species occurs off the North Cape of New Zealand (see Section 4.1.2). No evidence was found on the COI gene or control region of fish sampled in this region to suggest a species distinct from *P. georgianus* is occurring close to the coast of New Zealand's North Cape. However, all five fish that were sampled in the Three Kings Islands and the Kermadec Islands were highly genetically divergent from *P. georgianus* sampled from the rest of New Zealand as well as Western Australia and these fish could represent the cryptic species described in the literature.

6.1.3 Population structuring of *P. georgianus* in New Zealand

There was no clear evidence to suggest *P. georgianus* in New Zealand are structured as distinct populations nor isolated by distance. This possibly reflects their batch spawning behaviour and capability for long distance migrations (see Section 1.1). However, due to the large geographic range that *P. georgianus* occupies and the genetic difference identified between New Zealand and Western Australian populations, there could be isolation by distance on a broader scale throughout Australasia.

This thesis provided additional support for putative sub-stocks of *P. georgianus* in New Zealand (see Section 5.1.1). No evidence was found to refute the claim that *P. georgianus* in the Bay of Plenty are the same biological stock as fish from TRE2. There appears to be some genetic structuring of *P. georgianus* within TRE7 (on the West Coast of New Zealand's North Island), however, the exact structure of these populations was not able to be resolved.

6.2 Combining molecular genetics with morphological data

The taxonomy of the *Pseudocaranx* genus has been difficult to resolve, partly as a result of a historical reliance on morphology to inform the taxonomy of this genus. For contentious taxa such as *Pseudocaranx*, phylogenetic analyses provide further data to compliment morphological data. For the *Pseudocaranx* genus, the genetic data supported the taxonomic groupings produced from morphological analysis by Smith-Vaniz and Jelks (2006).

Because the taxonomic groupings of the *Pseudocaranx* genus was supported by the genetic data produced in this thesis, it is likely that morphology would be informative for future taxonomic studies of other *Pseudocaranx* species. It is also likely that such studies would result in the description of new species given the success of describing new *Pseudocaranx* species when research is undertaken (see Smith-Vaniz and Jelks (2006); Yamaoka et al. (1992)). However, there remains some among-species overlap in morphological characters (see James and Stephenson (1974); Masuda et al. (1995); Smith-Vaniz and Jelks (2006); Yamaoka et al. (1992)) and the possibility that environmental factors affect

Pseudocaranx morphology (see Masuda et al. (1995)). Therefore, morphological analyses would benefit from being complimented with phylogenetic analyses to provide a robust taxonomic description of the *Pseudocaranx* complex.

The fact that the genetic data supported prior taxonomic studies based on morphology highlights the value of such morphological analyses. However, there remains a place for genetics to validate taxonomy based on morphological analyses and identify evolutionary events that may obscure the results, such as convergent evolution. It appears that the best course of action is to integrate the two disciplines.

6.3 Population connectivity of *P. georgianus*

There is some information on the movements of other *Pseudocaranx* species or *P. georgianus* in Australia (see Afonso et al. (2009); Fairclough et al. (2011); Fowler et al. (2018)), however relatively little is known regarding the movements of New Zealand *P. georgianus*. Investigating the population genetics of New Zealand *P. georgianus* has, as a by-product, provided some information on the movements of these fish.

Genetically, *P. georgianus* from very broad geographic regions in New Zealand were highly genetically similar. This occurs over geographic regions broader than the level of movement described in the literature (see Section 1.1.2). It appears that the level of movement of New Zealand *P. georgianus* might be underestimated by a tag-recapture study undertaken on New Zealand *P. georgianus* (James, 1980). The geographic remoteness of *p. georgianus* has already been overestimated for this species as outlined by James and Stephenson (1974). Little is known regarding the capacity for *P. georgianus* to undertake long-distance movements. However, the level of genetic connectivity of *P. georgianus* over a vast geographic range suggests that *P. georgianus* may have a strong long distance dispersal capacity. This highlights the importance to not assume that there is no gene flow between geographically remote populations. This is particularly important for managing fisheries where a species geographic range covers different jurisdictions, such as *P. georgianus* fisheries in New Zealand and Australia.

6.4 Putting research into practice

6.4.1 Inputs for fisheries stock assessment models

Fish sampled from the Three Kings and the Kermadec Islands likely represent a species distinct from *P. georgianus*, possibly *P. wrighti*. Measures should be put in place to ensure this population/species is not

included in the total allowable catch (TAC) of TRE1 and is accounted for in TRE1 stock assessments. The specimens held in the Te Papa fish collection should be taxonomically identified and genetically sequenced. The results of this thesis suggest that *P. georgianus* from TRE1 and TRE2 are the same biological stock and the genetic data supports the decision to combine these fisheries management boundaries in future stock assessments (see Fisheries Science Group (2018)). The population structure of TRE7 is less obvious, given the low sample size of *P. georgianus* from South Taranaki Bight. However, the results support the decision to further investigate the stock structure of *P. georgianus* from the west coast of the North Island (TRE7) (see Fisheries Science Group (2018)). Importantly, no data is available for the stock structure of *P. georgianus* from New Zealand's South Island, comprising the TRE3 QMA. The commercial fishing intensity of this region is low, however the level of gene flow between *P. georgianus* in the South Island and North Island of New Zealand has important implications on the overall genetic stability of New Zealand *P. georgianus* fisheries.

If future research shows that *P. georgianus* in Australia and New Zealand are indeed the same species or the same population, the New Zealand and Australian *P. georgianus* fisheries may be best managed as a single fishery to ensure the sustainability of these genetically interacting fisheries. Unfortunately, this study does not provide information on the genetic connectivity of *P. georgianus* from Southern and Eastern Australia encompassing the Commonwealth Trawl and Scalefish Hook Sector (Australian Government Department of Fisheries, 2019a) and the Norfolk Island Fisheries (Australian Government Department of Fisheries, 2019b) with *P. georgianus* from New Zealand. Therefore, further research would be needed to determine if *P. georgianus* from these fisheries are the same or different species or populations as the New Zealand *P. georgianus* fisheries.

6.4.2 PCR primers for metabarcoding *Pseudocaranx* species

New Zealand manages *Pseudocaranx* fisheries based on the assumption that a single species and single population is included in each QMA (Fisheries Science Group, 2018). However, *Pseudocaranx* species are not easily distinguished morphologically (4.1.3). *Pseudocaranx* morphology changes with age (Smith-Vaniz and Jelks (2006)) which has already caused taxonomic confusion in the literature (James and Stephenson, 1974). The only morphological feature that is clearly distinct with no overlap between *Pseudocaranx* species is vertebrae counts (Smith-Vaniz and Jelks (2006)) which is not necessarily easy to quantify, especially if the fish needs to remain alive (eg. aquaculture broodstock). On top of this, there remains some question whether this feature is clearly distinct between *Pseudocaranx* species or could be influenced by environmental variables (Masuda et al. (1995)).

This could present challenges when identifying cryptic species that are included in *P. georgianus*

fisheries. The *P. georgianus* mitogenome as well the primers and associated PCR protocols developed in this thesis provide an opportunity to develop PCR primers to metabarcode *Pseudocaranx* species like what has been done by Miya et al. (2015). The control region has already been shown to be capable of delimiting *P. georgianus* and *P. wrighti* (Bearham, 2004) and could address the practical challenge of delimiting *Pseudocaranx* species in research, fisheries management and aquaculture.

6.5 Inferences based on two mitochondrial markers

Mitochondrial DNA studies have been a fundamental component of population genetic research for several decades (Ballard and Whitlock, 2004). However, in more recent years, some limitations of mitochondrial DNA markers have come to light. This included the presence of mitochondrial recombination in some species (eg. Burzynski et al. (2003); Davila et al. (2011); Leducq et al. (2017)), the limitations of testing for isolation by distance (Teske et al., 2018), introgression and direct or indirect selection on mitochondrial DNA markers (Ballard and Whitlock, 2004) and its inability to inform male migration or gene flow (Waples et al., 2008). This has raised questions on the appropriateness of a heavy reliance on mitochondrial DNA markers in phylogenetic and population genetic studies.

There are also limitations associated with using a limited number of DNA markers in phylogenetic and population genetic analyses. Phylogenetic incongruence can occur as a result of different signals held on different DNA markers (Shen et al., 2017) and different markers are appropriate for different purposes (Hellberg, 2009).

A thorough review of the literature outlining the limitations of using a few mitochondrial DNA markers is beyond the scope of this thesis. However, this emerging awareness in the literature of some of these limitations highlights the benefit of using additional genetic markers or undertaking genomics in studies extending the work of this thesis. This will provide more power to tackle complex questions and further investigate ambiguous results regarding the phylogenetics and population genetics of *P. georgianus*.

In recent years, there has been a growing interest using genetic tools to inform the biology and in turn the management of New Zealand fisheries (Bernatchez et al., 2017). There is an ongoing transition from low-resolution genetic studies towards genomics. More data produced using phylogenomics will not be a silver bullet for addressing issues that have come to light in mitochondrial and single gene studies (see Philippe et al., 2011; Philippe and Roure, 2011; Springer and Gatesy, 2018), but it does have the power to address a more comprehensive spectrum of fisheries needs (Bernatchez et al., 2017). Genomic tools have been underutilised in fisheries despite its ability to inform fisheries management, aquaculture, biosecurity (Bernatchez et al., 2017) and systematics (Pyrone, 2015). Typically, a few

genetic markers will provide enough resolution for population genetic studies (Bernatchez et al., 2017) since phylogenetic signals are typically strong over sampled loci (Pyrón, 2015). However, genomics has the advantage of identifying markers that are informative for the research question at hand (Bernatchez et al., 2017).

6.6 Future research directions

This thesis is based on two mitochondrial DNA markers. Although such studies have their limitations, it provides a first look into the phylogenetics and population genetics of this key fishery species for which no genetic work has been undertaken. The results of this thesis highlight possibly fruitful areas of study that can be targeted with additional genetic markers or genomics. By investigating the within-species variation along the length of the *P. georgianus* mitogenome, several species-specific gene-region targets have been described, providing a basis from which biologically informative mitochondrial DNA regions can be identified and targeted with PCR.

For *P. georgianus*, genomics could be used in the description of the cryptic species occurring in the Three Kings and Kermadec Islands. It could also provide a higher-level of resolution to investigate the ambiguous population structure of *P. georgianus* within TRE7 on the west coast of New Zealand's North Island. Moreover, genomics could help distinguish whether *P. georgianus* in New Zealand and Australia are the same species. As is often the case in ecology, scale is important. There was a higher than expected level of genetic connectivity, or panmixia, of *P. georgianus* within and between Australia and New Zealand. This suggests that future studies could benefit from capturing a broader geographic range than was sampled in this study.

The production of the *P. georgianus* mitogenome also provides the opportunity for it to be used as a reference mitogenome to assemble the mitogenomes of other *Pseudocaranx* species. This provides the opportunity to undertake a first genetic investigation into other fisheries species, like what has been done in this thesis. For example, *P. wrighti* is another commercially important fisheries species in Australia (Australian Fisheries Management Authority, 2010; Australian Government Department of Fisheries, 2019a) for which little genetic work has been undertaken. Overall, this thesis has contributed to the wider effort to apply genetic tools to fishery species.

The large majority of the tissues sequenced in this thesis, as well as further tissues that were not sequenced (see Section 3) are held in long-term storage. This provides the opportunity to undertake a genomic study of *P. georgianus*, allow further genetic sequencing of other DNA markers or undertake additional sequencing of the COI gene and control region of *P. georgianus*. These tissues could also

act as a basis from which to undertake serial sampling of *P. georgianus* which would enable researchers to account for seasonal and annual movements of *P. georgianus*. Such research will further extend our knowledge of the population structure, connectivity, movement and reproductive behaviour of *P. georgianus*, all of which contribute to New Zealand's goal of sustainably harvesting *P. georgianus*.

Appendix A

High salt DNA extraction protocol

Ingredients:

- 20mg/ml proteinase K
- DNA extraction buffer (0.4M NaCl, 10mM Tris-HCl pH 8.0, and 2mM EDTA pH 8.0)
- TE buffer (10 mM Tris, 1mM EDTA, pH 8.0)
- 20% sodium dodecyl sulfate (warm to re-dissolve)
- 5M sodium chloride (NaCl saturated dH₂O: Autoclave)
- 100% isopropanol
- 70% ethanol

Sample preparation

1. Select the samples to be used from the freezer and hold on ice during sample prep.
2. Remove the tissue and cut a portion onto a clean kimwipe in a petri dish.
3. Squash out the ethanol or scrape excess salts, weigh ~20-40mg of tissue into a 1.5ml microcentrifuge tube and return tissue to storage. Cut up the tissue. to improve digestion and prevent the tissue from sticking to side of tube.
4. Clean forceps and scissors between each sub-sample by wiping with a clean paper kimwipe, dipping in 98% ethanol and sterilising by passing through a flame

Cell lysis

1. Add 400 μ l of DNA extraction buffer and 80 μ l of SDS (If SDS has precipitated, dissolve it on a heatblock (>40°C).

2. Add 10 μ l of proteinase K.
3. Place in orbital mixer for 2hrs to overnight set to 300rpm and 50°C throughout digestion.

Protein precipitation and removal

1. Centrifuge max speed – 5 min.
2. Transfer supernatant to a new tube.
3. Add 320 μ l of 5M sodium chloride to each tube and mix by inverting the tubes 60 times.
4. Centrifuge max speed – 10 min.
5. Transfer supernatant to a new tube.

DNA precipitation

1. Add 525 μ l of chilled (-20°C) 100% isopropanol to the supernatant.
2. Inverted 60 times.
3. Centrifuge tubes for 20 min at 13K rpm (15.7k xg) at 4°C.
4. Carefully remove supernatant without disturbing pellet.
5. Add 1ml of chilled 70% ethanol to tubes and invert 60 times.
6. Centrifuge for 10 min at 13K rpm (15.7k xg) at 4°C.
7. Remove the supernatant and air dry the pellet for approximately 5 min at 37°C on heat block, leave tubes open and cover with kimwipe - do not over-dry.

DNA rehydration

1. Add 30 μ l of TE buffer to the dried pellet (dilute further if needed) and gently agitate tube to resuspended DNA. Leave to rehydrate for ~2hrs to overnight.
2. Store DNA in fridge until checking quality (no longer than 48 hrs) or store in freezer for long term storage.

DNA Quantification

1. Quantify DNA on the Nanodrop. Record ng/ μ l, 260/280 ratio (>1.8 pure DNA and <1.7 indicates protein contamination) and 260/230 ratio (<1.5 indicate salt contamination, ideally >1.8).
2. Wipe pedestal clean with water.
3. Add 1 μ l of TE buffer and run blank.

4. Wipe pedestal with kimwipe then add $1\mu\text{l}$ of the first sample and click analyse.
5. Repeat this step for each sample, wiping the pedestal clean with a kimwipe between samples, re-blank with TE buffer after every 5 samples.
6. If DNA is too concentrated, dilute then store samples in fridge for at least 2 hours before re-measuring. Gently agitate tube to resuspend DNA before re-measuring. Aiming for around $100\text{-}200\text{ng}/\mu\text{l}$ of DNA.
7. Analyse $1\mu\text{l}$ of DNA on a 1% agarose gel. Compare against $1\mu\text{l}$ of a high molecular weight ladder such as lambda Hind 11. Run gel for 30 minutes at 90V.
8. Stain gel in Ethidium bromide for 20mins, wash in H_2O for 1min and visualize in the UV imaging box.

Figure A.1: Final *P. georgianus* mitochondrial genome sequence

Figure A.2: Final COI alignment.

Figure A.3: Figure continued.

Figure A.5: Final control region alignment (83 of 304 sequences).

[illegible]

Figure A.6: Figure continued.

[illegible]

Figure A.7: Figure continued.

Figure A.8: Final control region alignment of five Trevally sampled from Three Kings and Kermadec Islands.

Figure A.9: Posterior probability distributions for all four Migrate model parameters.

Bibliography

- Afonso, P., Fontes, J., Holland, K., and Santos, R. (2009). Multi-scale patterns of habitat use in a highly mobile reef fish, the white trevally *Pseudocaranx dentex*, and their implications for marine reserve design. *Marine Ecology Progress Series*, 381:273–286.
- Afonso, P., Fontes, J., Morato, T., Holland, K., and Santos, R. (2008). Reproduction and spawning habitat of white trevally, *Pseudocaranx dentex*, in the Azores, central north Atlantic. *Scientia Marina*, 72(2):373–381.
- Allaire, J., Gandrud, C., Russell, K., and Yetman, C. (2017). *networkD3: D3 JavaScript network graphs from R*. R package version 0.4.
- Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., Wickham, H., Cheng, J., Chang, W., and Iannone, R. (2018). *rmarkdown: dynamic documents for R*. R package version 1.11.
- Allendorf, F. and Ryman, N. (2002). *The role of genetics in population viability analysis*, pages 50–85. University of Chicago Press, Chicago.
- Appelhans, T., Detsch, F., Reudenbach, C., and Woellauer, S. (2019). *mapview: interactive viewing of spatial data in R*. R package version 2.7.0.
- Australian Fisheries Management Authority (2010). Norfolk island inshore fishery data summary 2006-2009. Report, Australian Fisheries Management Authority. <https://www.afma.gov.au/sites/default/files/uploads/2014/02/2006-2009-Data-Summary-Norfolk-Island-Inshore-Fishery.pdf>.
- Australian Government Department of Fisheries (2019a). Commonwealth trawl and scalefish hook sectors. <http://www.agriculture.gov.au/abares/research-topics/fisheries/fishery-status/trawl-scalefish-hook-sectors>.

- Australian Government Department of Fisheries (2019b). Norfolk island fishery. <http://www.agriculture.gov.au/abares/research-topics/fisheries/fishery-status/norfolk-island>.
- Avise, J., Arnold, J., Ball, R., Bermingham, E., Lamb, T., Neigel, J., Reeb, C., and Saunders, N. (1987). Intraspecific phylogeography - the mitochondrial dna bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, 18:489–522.
- Ballard, J. and Whitlock, M. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, 13(4):729–744.
- Bandelt, H., Forster, P., and Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16(1):37–48.
- Bearez, P. and Villaroel, C. (2018). First record of *Pseudocaranx chilensis* (Carangidae) from the continental coast of north-central Chile. *Cybium*, 42(3):303–305.
- Bearham, D. (2004). *A mtDNA study of the population structure of silver trevally Pseudocaranx dentex and the relationship between silver and sand trevally Pseudocaranx wrighti in Western Australian waters*. Thesis, Murdoch University.
- Beerli, P. (1998). Estimation of migration rates and population sizes in geographically structured populations. *Advances in Molecular Ecology*, 306:39–53.
- Beerli, P. and Felsenstein, J. (1999). Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*, 152(2):763–773.
- Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D., Barth, J., Beacham, T., Maes, G., Martinsohn, J., Miller, K., Naish, K., Ovenden, J., Primmer, C., Suk, H., Therkildsen, N., and Withler, R. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends in Ecology and Evolution*, 32(9):665–680.
- Bhaska, R. and Barret, S. (2018). *leaflet.extras: extra functionality for 'leaflet' package*. R package version 1.0.0.
- Biomatters (2018). Geneious 11.1. Report, Geneious.
- Brooks, A. (2019). *Using ecological niche modelling to predict climate change responses of ten key fishery species in Aotearoa New Zealand*. Thesis, Victoria University of Wellington.
- Burzynski, A., Zbawicka, M., Skibinski, D., and Wenne, R. (2003). Evidence for recombination of mtDNA in the marine mussel *Mytilus trossulus* from the Baltic. *Molecular Biology and Evolution*, 20(3):388–392.

- Campos, P., de Castro, M., and Bonecker, A. (2010). Occurrence and distribution of Carangidae larvae (Teleostei, Perciformes) from the Southwest Atlantic Ocean, Brazil (12-23 degrees S). *Journal of Applied Ichthyology*, 26(6):920–924.
- Casillas, S. and Barbadilla, A. (2017). Molecular population genetics. *Genetics*, 205(3):1003–1035.
- Castro, J., Santiago, J., and Hernandez-Garcia, V. (1999). Fish associated with fish aggregation devices off the Canary Islands (Central-East Atlantic). *Scientia Marina*, 63(3-4):191–198.
- Charlop-Powers, Z. (2017). *phyloanvas: interactive phylogenetic trees using the 'Phyloanvas' JavaScript library*. R package version 0.1.3.
- Cheng, J., Karambelkar, B., and Xie, Y. (2018). *leaflet: create interactive web maps with the JavaScript 'Leaflet' library*. R package version 2.0.2.
- Cryer, M., Mace, P., and Sullivan, K. (2016). New Zealand's ecosystem approach to fisheries management. *Fisheries Oceanography*, 25:57–70.
- Damerau, M., Freese, M., and Hanel, R. (2018). Multi-gene phylogeny of jacks and pompanos (Carangidae), including placement of monotypic vadigo *Campogramma glaycos*. *Journal of Fish Biology*, 92:190–202.
- Darling, A., Mau, B., and Perna, N. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *Plos One*, 5(6).
- Darriba, D., Taboada, G., Doallo, R., and Posada, D. (2012). jmodeltest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9(8):772–772.
- Davila, J., Arrieta-Montiel, M., Wamboldt, Y., Cao, J., Hagmann, J., Shedje, V., Xu, Y., Weigel, D., Mackenzie, S., and A., S. (2011). Double-strand break repair processes drive evolution of the mitochondrial genome in arabidopsis. *BMC Biology*, 9(64):1–14.
- Department of Agriculture and Water Resources (2018). Fishery status reports 2018. Report, Australian Government. <http://www.agriculture.gov.au/SiteCollectionDocuments/abares/publications/fsr2018.pdf>.
- Dudley, B., Tolimieri, N., and Montgomery, J. (2000). Swimming ability of the larvae of some reef fishes from New Zealand waters. *Marine and Freshwater Research*, 51(8):783–787.
- Duffy, C. and Ah Yong, S. (2015). Annotated checklist of the marine flora and fauna of the Kermadec Islands Marine Reserve and northern Kermadec Ridge. Report, Auckland

- War Memorial Museum. <https://www.aucklandmuseum.com/getmedia/b616a69e-f490-41fb-88a4-65592900054a/bulletin-vol20-kermadec-biodiscovery-expedition-02-duffy-ahyong>.
- Dyer, B. and Westneat, M. (2010). Taxonomy and biogeography of the coastal fishes of Juan Fernandez Archipelago and Desventuradas Islands, Chile. *Revista De Biología Marina Y Oceanografía*, 45:589–617.
- Excoffier, L. and Lischer, H. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3):564–567.
- Excoffier, L., Smouse, P., and Quattro, J. (1992). Anaysis of molecular variance inferred from metric distances among DNA haplotypes - application to human mitochondrial DNA restriction data. *Genetics*, 131(2):479–491.
- Fairclough, D., Potter, I., Lek, E., Bivoltsis, A., and Babcock, R. (2011). The fish communities and main fish populations of the Jurien Bay Marine Park. Report, Centre for Fish and Fisheries Research. <https://www.fisheries.govt.nz/dmsdocument/18337-fisheries-assessment-plenary-may-2017-volume-3-pipi-to-yellow-eyed-mullet>.
- Farmer, B., French, D., Potter, I., Hesp, S., and Hall, N. (2005). Determination of biological parameters for managing the fisheries for mullet and silver trevally in Western Australia. Report, Murdoch University. <https://www.frdc.com.au/Archived-Reports/FRDC>
- Farris, J. (1970). Methods for computing Wagner trees. *Systematic Zoology*, 19(1):83–92.
- Fisheries Science Group (2018). Fisheries assessment plenary, may 2018: stock assessments and stock status. Report, Fisheries New Zealand. <https://www.fisheries.govt.nz/dmsdocument/18337-fisheries-assessment-plenary-may-2017-volume-3-pipi-to-yellow-eyed-mullet>.
- Fowler, A., Chick, R., and Stewart, J. (2018). Patterns and drivers of movement for a coastal benthopelagic fish, *Pseudocaranx georgianus*, on Australia’s southeast coast. *Scientific reports*, 8(1):16738–16738.
- Francis, M., Morrison, M., Leathwick, J., Walsh, C., and Middleton, C. (2005). Predictive models of small fish presence and abundance in northern New Zealand harbours. *Estuarine Coastal and Shelf Science*, 64(2-3):419–435.
- Froese, R. and Pauly, D. (2019). Fishbase. <https://www.fishbase.de/>.
- Fu, Y. (1996). New statistical tests of neutrality for DNA samples from a population. *Genetics*, 143(1):557–570.

- Gebbies, C. (2014). *Population genetic structure of New Zealand blue cod (*Parapercis colias*) based on mitochondrial and microsatellite DNA markers*. Thesis, Victoria University of Wellington.
- Guindon, S., Dufayard, J., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate Maximum-Likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3):307–321.
- Guindon, S. and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52(5):696–704.
- Hahn, C., Bachmann, L., and Chevreux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research*, 41(13).
- Hellberg, M. (2009). Gene flow and isolation among populations of marine animals. *Annual Review of Ecology Evolution and Systematics*, 40:291–310.
- Hodgson, B. (2011). *Kahawai phylogeny and phylogenetics: a genetic investigation into commercial and recreational fisheries management and practice*. Thesis, Victoria University of Wellington.
- Holland, B., Huber, K., Moulton, V., and Lockhart, P. (2004). Using consensus networks to visualize contradictory evidence for species phylogeny. *Molecular Biology and Evolution*, 21(7).
- Horn, P. (1986). Distribution and growth of snapper *Chrysophrys auratus* in the North Taranaki Bight, and management implications of these data. *New Zealand Journal of Marine and Freshwater Research*, 20(3):419–430.
- Hudson, R. (1990). Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology*, pages 1–44.
- Huelsenbeck, J. and Rannala, B. (2004). Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic Biology*, 53:904–913.
- Huelsenbeck, J. and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8):754–755.
- Jaafar, T., Taylor, M., Nor, S., de Bruyn, M., and Carvalho, G. (2012). DNA barcoding reveals cryptic diversity within commercially exploited indo-malay Carangidae (Teleostei: Perciformes). *Plos One*, 7(11).
- James, G. (1976). Eggs and larvae of the trevally *Caranx georgianus* (Teleostei: Carangidae). *New Zealand Journal of Marine and Freshwater Research*, 10(2):301–310.

- James, G. (1978). Trevally and koheru - biology and fisheries. *Fisheries Research Division Occasional Publication (New Zealand)*, 15:50–54.
- James, G. (1980). Tagging experiments on trawl-caught trevally, *Caranx georgianus*, off north-east New Zealand, 1973-79. *New Zealand Journal of Marine and Freshwater Research*, 14(3):249–254.
- James, G. and Stephenson, A. (1974). *Caranx georgianus* Cuvier, 1833 (Pisces: Carangidae) in temperate Australasian waters. *Journal of the Royal Society of New Zealand*, 4(4):401–410.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12):1647–1649.
- Kingman, J. (1982a). The coalescent. *Stochastic Processes and their Applications*, 14:235–248.
- Kingman, J. (1982b). On the genealogy of large populations. *Journal of Applied Probability*, 19:22–43.
- Kingman, J. (2000). Origins of the coalescent: 1974-1982. *Genetics*, 156(4):1461–1463.
- Kress, W., Garcia-Robledo, C., Uriarte, M., and Erickson, D. (2015). DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology and Evolution*, 30(1):25–35.
- Kruskal, J. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, 7:48–50.
- Laikre, L., Palm, S., and Ryman, N. (2005). Genetic population structure of fishes: implications for coastal zone management. *Ambio*, 34(2):111–119.
- Langley, A. (2004). Length and age composition of Trevally (*Pseudocaranx dentex*) in commercial landings from TRE1 purse-seine fishery, 2002-03. *New Zealand Fisheries Assessment Research Document series*.
- Langley, A., Kendrick, T., and Bentley, N. (2015). Stock assessment of trevally in TRE7. Report.
- Leducq, J., Henault, M., Charron, G., Nielly-Thibault, L., Terrat, Y., Fiumera, H., Shapiro, B., and Landry, C. (2017). Mitochondrial recombination and introgression during speciation by hybridization. *Molecular Biology and Evolution*, 34(8):1947–1959.
- Leigh, J. and Bryant, D. (2015). POPART: full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6(9):1110–1116.

- Masuda, R., Kamaishi, T., Kobayashi, T., Tsukamoto, K., and Numachi, K. (1995). Mitochondrial DNA differentiation between two sympatric morphs of striped jack near Japan. *Journal of Fish Biology*, 46(6):1003–1010.
- Meirmans, S., Meirmans, P., and Kirkendall, L. (2012). The costs of sex: facing real-world complexities. *Quarterly Review of Biology*, 87(1):19–40.
- Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J. Y., Sato, K., Minamoto, T., Yamamoto, S., Yamanaka, H., Araki, H., Kondoh, M., and Iwasaki, W. (2015). Mifish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *Royal Society Open Science*, 2(7).
- Nath, H. and Griffiths, R. (1993). The coalescent in 2 colonies with symmetrical migration. *Journal of Mathematical Biology*, 31(8):841–851.
- National Institute of Water and Atmospheric Research (2019). Feature layers by mpi geospatial management. <https://mpi.maps.arcgis.com/home/search.html?q=owner>
- Neira, F., Perry, R., Burrige, C., Lyle, J., and Keane, J. (2015). Molecular discrimination of shelf-spawned eggs of two co-occurring *Trachurus* spp. (Carangidae) in southeastern Australia: a key step to future egg-based biomass estimates. *Ices Journal of Marine Science*, 72(2):614–624.
- Nogueira, N., Ferreira, M., Cordeiro, N., and Canada, P. (1990). The coalescent and the genealogical process in geographically structured population. *Journal of Mathematical Biology*, 29(1):59–75.
- Ojala, D., Montoya, J., and Attardi, G. (1981). Transfer-RNA punctuation model of RNA processing in human mitochondria. *Nature*, 290(5806):470–474.
- Paradis, E. (2018). Analysis of haplotype networks: the randomized minimum spanning tree method. *Methods in Ecology and Evolution*, 9(5):1308–1317.
- Pebesma, E. (2018). *sf: simple features for R*. R package version 0.7-2.
- Pebesma, E. and Bivand, R. (2018). *sp: classes and methods for spatial data*. R package version 1.3-1.
- Philippe, H., Brinkmann, H., Lavrov, D., Littlewood, D. T. J., Manuel, M., Woerheide, G., and Baurain, D. (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. *Plos Biology*, 9(3).
- Philippe, H. and Roure, B. (2011). Difficult phylogenetic questions: more data, maybe; better methods, certainly. *Bmc Biology*, 9.

- Pyron, A. (2015). Post-molecular systematics and the future of phylogenetics. *Trends in Ecology and Evolution*, 30(7):384–389.
- Ray, N., Currat, M., and Excoffier, L. (2003). Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution*, 20(1):76–86.
- Rogers, A., Fraley, A., Bamshad, M., Watkins, S., and Jorde, L. (1996). Mitochondrial mismatch analysis is insensitive to the mutational process. *Molecular Biology and Evolution*, 13(7):895–902.
- Rogers, A. and Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, 9(3):552–569.
- Rosenberg, N. and Nordborg, M. (2002). Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics*, 3(5):380–390.
- Rowling, K. and Raines, L. (2000). Description of the biology and an assessment of the fishery for silver trevally *Pseudocaranx dentex* off New South Wales. Report, NSW Fisheries. <http://www.agriculture.gov.au/SiteCollectionDocuments/abares/publications/fsr2018.pdf>.
- Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J., Guirao-Rico, S., Librado, P., Ramos-Onsins, S., and Sanchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, 34(12):3299–3302.
- Santini, F. and Carnevale, G. (2015). First multilocus and densely sampled timetree of trevallies, pompanos and allies (Carangoidei, Percomorpha) suggests a Cretaceous origin and Eocene radiation of a major clade of piscivores. *Molecular Phylogenetics and Evolution*, 83:33–39.
- Satoh, T., Miya, M., Mabuchi, K., and Nishida, M. (2016). Structure and variation of the mitochondrial genome of fishes. *BMC Genomics*, 17.
- Scott, S. (2010). *Phylogeography of the common New Zealand wrasse species, Notolabrus celidotus, and the phylogenetics of the Pseudolabrine tribe*. Thesis, Victoria University of Wellington.
- Shen, X., Hittinger, C., and Rokas, A. (2017). Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nature Ecology and Evolution*, 1(5).
- Sievert, C., Parmer, C., Hocking, T. Chamberlain, S., Ram, K., Corvellec, M., and Despouy, P. (2018). *plotly: create interactive web graphics via 'plotly.js'*. R package version 4.8.0.
- Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*, 236(4803):787–792.

- Smith, H. (2012). *Characterisation of the mitochondrial genome and the phylogeographic structure of blue cod (*Parapercis colias*)*. Thesis, Victoria University of Wellington.
- Smith-Vaniz, W. and Jelks, H. (2006). Australian trevallies of the genus *Pseudocaranx* (Teleostei: Carangidae), with description of a new species from Western Australia. *Memoirs of Museum Victoria*, 63(1):97–106.
- Springer, M. and Gatesy, J. (2018). On the importance of homology in the age of phylogenomics. *Systematics and Biodiversity*, 16(3):210–228.
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3):585–595.
- Teske, P., Golla, T., Sandoval-Castillo, J., Emami-Khoyi, A., van der Lingen, C., von der Heyden, S., Chiazzari, B., van Vuuren, B., and Beheregaray, L. (2018). Mitochondrial DNA is unsuitable to test for isolation by distance. *Scientific Reports*, 8.
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B., Remm, M., and Rozen, S. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*, 40(15).
- Waples, R., Punt, A., and Cope, J. (2008). Integrating genetic data into management of marine resources: how can we do it better? *Fish and Fisheries*, 9(4):423–449.
- Ward, R., Hanner, R., and Hebert, P. (2009). The campaign to DNA barcode all fishes. *Journal of Fish Biology*, 74(2):329–356.
- Wickham, H., Chang, W., Henry, L., Pedersen, T., Takahashi, K., Wilke, C., and Woo, K. (2018). *ggplot2: create elegant data visualisations using the grammar of graphics*. R package version 3.1.0.
- Wickham, H., François, R., Henry, L., and Müller, K. (2019). *dplyr: a grammar of data manipulation*. R package version 0.8.0.1.
- Wilcox, R. (2015). *A population genetic analysis of the New Zealand spotty (*Notolabrus celidotus*) using mitochondrial DNA and microsatellite DNA markers*. Thesis, Victoria University of Wellington.
- Wright, S. (1943). Isolation by distance. *Genetics*, 28(2):114–138.
- Xie, Y. (2018a). *bookdown: authoring books and technical documents with R Markdown*. R package version 0.9.
- Xie, Y. (2018b). *knitr: a general-purpose package for dynamic report generation in R*. R package version 1.21.

- Xie, Y. (2019). *tinytex: helper functions to install and maintain 'TeX Live', and compile 'LaTeX' documents*. R package version 0.13.
- Yamaoka, K., Han, H., and Taniguchi, N. (1992). Genetic dimorphism in *Pseudocaranx dentex* from Tosa Bay, Japan. *Bulletin of the Japanese Society of Scientific Fisheries*, 58(1):39–44.
- Yang, Z. and Rannala, B. (2012). Molecular phylogenetics: principles and practice. *Nature Reviews Genetics*, 13(5):303–314.
- Ye, F., Samuels, D., Clark, T., and Guo, Y. (2014). High-throughput sequencing in mitochondrial DNA research. *Mitochondrion*, 17:157–163.
- Yu, Y., Ouyang, Y., and Yao, W. (2018). shinycircos: an R/Shiny application for interactive creation of Circos plot. *Bioinformatics*, 34(7):1229–1231.
- Zhu, H. (2019). *kableExtra: construct complex table with 'kable' and pipe syntax*. R package version 1.1.0.
- Zou, K. and Li, M. (2016). Characterization of the mitochondrial genome of the whitefin trevally *Carangoides equula* (Perciformes: Carangidae): a novel initiation codon for ATP6 gene. *Mitochondrial DNA Part A*, 27(3):1779–1780.