Population genetic structure and connectivity patterns of the Kānae (*Mugil cephalus*) in New Zealand inland and coastal waters

By

Angel Balam Jimenez Brito

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

the degree of Doctor of Philosophy

Victoria University of Wellington

Te Whare Wānanga o te Ūpoko o te Ika a Māui



This thesis was conducted under the supervision of

Dr Peter Ritchie (Primary Supervisor)

School of Biological Sciences, Victoria University of Wellington, Wellington

New Zealand.

&

Dr Mark Morrison (Secondary Supervisor)

National Institute of Water and Atmosphere (NIWA), Auckland, New Zealand.

Para Hinemoana:

Tu eres mi más grande bendición, has sido mi fuerza y mi esperanza. No importa lo difícil, tu pequeña sonrisa lo vale todo. Mi amor en su totalidad es para ti. Llenas mi corazón de felicidad y cada logro es tuyo mi amor. Gracias.

To Kimberley:

Te Aroha Nui Kimo. I would have never made it without you. You gave me a place in your life and I am thankful you are part of mine. I love you and thank you.

Para mi Mamá.

Tu sentaste las bases de todo lo que soy y he logrado mamá. Todo esto es consecuencia de tu amor y tu guía. Gracias mamá.

Kānae whakapapa



Te Ahukaramū Charles Royal, 'Tangaroa – the sea - The importance of the sea', Te Ara - the Encyclopedia of New Zealand, http://www.TeAra.govt.nz/en/document/7389/fish-genealogy

Abstract

Mugil cephalus is a cosmopolitan fish species found in most coastal waters from tropical to temperate zones. It is a species common in the near-shore marine environment, and known to reside in estuarine and freshwater systems. Adult *M. cephalus* move out to sea to spawn in aggregations. Their larvae can drift on surface ocean currents for over a month before recruitment to nursery grounds. *Mugil cephalus* is a species that is closely associated with the coastal environment, but it is capable of interoceanic migrations. Population genetic studies have reported high levels of genetic differentiation among populations in the Mediterranean, Atlantic and western Pacific. However, there is no evidence to suggest reproductive incompatibility has arisen among populations. In New Zealand *M. cephalus* supports important recreational, commercial and customary fisheries, but very little is known about the distribution and connectivity among populations.

The aim of this study was to use nuclear microsatellite DNA (msatDNA) and mitochondrial DNA (mtDNA) markers to describe the population genetic structure, connectivity patterns and to determine the phylogeographic history of New Zealand *M. cephalus* populations. Total of 850 samples were collected (576 adults and 274 juveniles) during the summers of 2010 and 2014-2015 from 15 locations around coastal and inland waters of the North Island, and one location in Marlborough Sounds. In addition, 245 mtDNA sequences were added from previously published studies and used to outgroup the New Zealand population and place it into the context of the other Pacific populations.

Seven msatDNA loci were isolated and used to determine the population genetic structure and connectivity patterns of *M. cephalus* in New Zealand. Admixture of four genetically distinct groups or populations was identified and a chaotic spatial distribution of allele frequencies. Within each population there was significant gene flow among locations, no pattern of genetic isolation-by-distance was identified and there was a high proportion of nonmigrant individuals. There was evidence of bottlenecks and seasonal reproductive variation of adults, which could explain the significant shifts in the effective population size among locations.

To test whether the pattern of genetic variation in *M. cephalus* populations was the result of seasonal variability in the reproductive success of adults, DNA from adult and juvenile samples were used to test for differences in the levels of genetic variation between generations (cohorts). Juveniles were grouped by age classes and compared to the adults. The levels of genetic diversity within the groups of juveniles were compared to the adult population and significant genetic bottlenecks between juveniles and adults were detected. This pattern was consistent with the Sweepstake-Reproductive-Success hypothesis. Two spawning groups in the adults were identified, an early spawning group and a late spawning group.

The analysis of DNA sequence data from the mtDNA Cytochrome Oxidase subunit 1 (COX1) gene and D-loop region showed two sympatric haplogroups of *M. cephalus*. New Zealand was most likely colonised by *M. cephalus* migrants from different population sources from the Pacific first ~50,000 and a second wave of migrants from Australia between ~20,000 and ~16,000 years ago. High levels of gene flow were detected, but there has not been enough time for genetic drift to completely sort the lineages.

The findings of this thesis research will help with the understanding of aspects of *M*. *cephalus* dispersal and the genetic structure of populations. The patterns of connectivity can be used to better align the natural boundaries of wild populations to the fishery management stock structure. Understanding the reproductive units, levels of genetic diversity and the patterns of reproduction of *M. cephalus* will assist management efforts to focus on the key habitats threats, risks and the long-term sustainability of the species.

Iwi:

Te Whānau-ā-Apanui; Waikato Tainui; Rangitāne o Wairarapa; Whakatōhea.

People:

Kimberley Maxwell, Hinemoana Jimenez, Columba Brito, Matua Weno, Maren Preuss, Cong Zeng, Shalen Kumar, Catherine Davis, Nicholas Manukau, Rangi Mahuta, Christian Boedeker, Don Nelson, John van der Sman, Neville Higgison, Dorothee Durpoix, Peter Ritchie, Mark Morrison, Tom Trnski, Scott Tindale, Ravini Sachintha, Mark Fenwick, Jim O'Brian, my Family, my adoptive family and friends. Thank you for all your support. Your love was food to my soul and allowed me to keep going even in the darkest hours. I love you all.

Funds and Projects:

MPI GMU200901; MBIE Coastal Conservation Management (CCM) CO1X0907; Victoria Doctoral Scholarship; Victoria Submission Scholarship; Allan Wilson Centre Summer Legacy Projects; VUW-Science Faculty Grants.

Table of Contents

Abstract	ix
Iwi:	xi
People:	xi
Funds and Projects:	xi
Table of Contents	. xiii
List of Figures	.xvii
List of Tables	xxi
List of Equations	.xxv
General Introduction – Chapter I	1
Genetically structured populations and gene flow	4
Grey mullet (Mugil cephalus) biology, distribution and ecology	6
Grey mullet in New Zealand, fisheries and local management	9
Aim of this thesis research and study structure	12
Specific Aims	13
Insight into the contemporary genetic population structure and connectivity patterns of Ka (<i>Mugil cephalus</i>) in New Zealand coastal and inland waters – Chapter II	inae 17
Abstract	18
Keywords: chaotic genetic structure, bottlenecks, heterozygote excess, sweepstakes reproductive success, philopatry	18
Introduction	19
Materials and Methods	22
Data collection	22
Amplification and genotyping	22
Loci assessment and genetic diversity	23
Population structure	27
Bayesian population clustering and spatial distribution	28
Mutation-drift equilibrium	29
Migration and contemporary connectivity patterns	30
Results	31
Loci assessment and genetic diversity	31
Population structure	44
Bayesian population clustering and spatial distribution	51
Mutation-drift equilibrium	52
Migration and contemporary connectivity patterns	52

Discussion	54
Linkage disequilibrium	54
Chaotic population genetic structure	56
Mugil cephalus population genetic structure in the world	59
Conclusion	63
Appendix	65
<i>Mugil cephalus</i> reproductive success and temporal fractioned genetic diversity between adults and juveniles in New Zealand – Chapter III	73
Abstract	74
Key words: chaotic genetic structure, temporal genetic variability, single-sample and temporal effective population size methods, effective population size, <i>Mugil cephalus</i> .	74
Introduction	75
Methods	77
Data collection	77
Amplification and genotyping	78
Loci assessment and genetic diversity	78
Juvenile and adult genetic structure and differentiation levels	79
Mutation-drift equilibrium	79
Effective population size estimations	80
Relatedness	81
Proportion of non-migrant juveniles	81
Results	82
Loci assessment and genetic diversity	82
Juvenile and adult genetic structure and differentiation levels	83
Mutation-drift equilibrium	92
Effective population size estimation	93
Relatedness	96
Proportion of non-migrant juveniles	97
Discussion	98
Sweepstake Reproductive Success	99
Allele frequency differences and heterozygote excess	99
Co-ancestry	101
Linkage Disequilibrium	103
Mugil cephalus migratory and reproductive behaviours	104
Conclusion	107

Appendix	
Population genetic structure and lineage divergence of the Kānae (<i>Mugil cepha</i> Zealand based on two mitochondrial DNA markers – Chapter IV	<i>ulus</i>) in New 111
Abstract	112
Key words: Phylogeography, population genetic structure, secondary contact high-density blocking	t, colonisation, 112
Introduction	113
Materials and Methods	116
Data collection	116
Amplification and alignment	118
Lineage divergence	119
Divergence time and cryptic lineages discovery	119
Neighbour Joining clustering	122
Genetic structure and diversity	122
Mitochondrial DNA and microsatellite DNA structure comparison	123
Demographic history	123
Migration and connectivity patterns	124
Results	125
New Zealand	125
Lineage divergence	126
Genetic structure and diversity	126
Mitochondrial DNA and microsatellite DNA structure comparison	130
Demographic history	131
Migration and connectivity patterns	132
Pacific Rim	138
Discussion	143
GMYC model more than species delimitation	143
Population connectivity and gene flow	144
Lineage divergence and New Zealand colonisation	147
Mugil cephalus in New Zealand compared to the Pacific	151
Conclusions	152
Appendix	153
General Discussion – Chapter V	175
Overview on Mugil cephalus in New Zealand	176
Population genetic structure of <i>M. cephalus</i> in New Zealand	177

177
179
184
187
188
190
193

List of Figures

<i>Figure I.1 <u>Mugil cephalus</u>, first described in 1758 by Linnaeus and it is known in New Zealand as kānae and grey mullet</i>
Figure I.2 Major harbours/estuaries of the North Island and upper South Island where samples were collected for this research. The areas in blue are <u>M. cephalus</u> fishing areas in New Zealand (GMU-grey mullet units). The map on the right represents the six putative populations or stocks identified by McKenzie and Vaughn (2008) based on CPUE (Catch per unit effort)
Figure II.1 Locations where <u>M. cephalus</u> samples were collected between November 2014 and April 2015 23
Figure II.2 Allele frequencies by locus and sampling location
Figure II.3 Rarefaction curves of allele discovery by locus and sampling location
Figure II.4 Principal Component analysis of the genetic relationship between sampling locations
Figure II.5 Factorial Component Analysis (FCA) of the individual genetic distances by sampling location as implemented in Genetix. The three axes explain 42% of the individual variation of the total genetic differences. The red dots represent the individuals sampled in each location
Figure II.6 Bayesian cluster analysis showing the proportion of ancestry in each individual from four genetically

Figure III.6 Proportion of non-migrants and connectivity patterns by sampling location of <u>M. cephalus</u> juveniles in New Zealand. In green are East-coast locations and in red are West-coast locations. Only the flows that are representing a migration volume of more than 5% are shown. The outer circle segments below the locality name represent the total number of individuals in each location. The inner bars show the total volume of migration encoded in the circle segments. To define directionality, in each location emigrants are represented with the same colour of the location and immigrants are coloured as the locality of origin. The width of the curved line is the volume of movements where the exact volume of movements can be visualized only at the beginning and end points. The width of the curved line depends on number of migrants relative to the total volume of migrants across all regions indicated by the outside circle segments.

Figure III.7 Variation of r^2 compared to the expected r^2 of the Ne of juveniles and adults at three different critical values (p = 0.05; 0.02; 0.01). r^2 is shown in solid lines; expected r2 is shown in dotted lines. r^2 variation at 0.05 critical value = yellow; r^2 variation at 0.02 critical value = blue; r^2 variation at 0.01 critical value = red. 110

Figure IV.4 Map of the distribution of haplotype frequencies in New Zealand. a) Frequency distributions of haplotypes from COX1. b) Frequency distribution of D-loop haplotypes clustered in haplogroups L1 and L2.135

Figure IV.6 Contemporary mtDNA connectivity patterns of New Zealand <u>M. cephalus</u>. Only the flows that are representing a migration volume of more than 51% are shown. The colour pattern follows regional distribution: Blue = South Island and southern North Island; orange/yellow = west coast North Island; green = east coast North Island. The outer circle segments below the locality name represent the total number of migrants moving to and from the locality. The inner bars show the total volume of migration encoded in the circle segments. To define directionality, in each location emigrants are represented with the same colour of the location and immigrants are coloured as the locality of origin. The width of the curved line is the volume of movements where the exact volume of movements can be visualized only at the beginning and end points. The width of the curved

Figure IV.8 Posterior probability branch support analysis for the GMYC model that shows the entities that are well defined and which variation could be explained by Yule's speciation model. Only three entities have a significant posterior probability branch support (p > 0.95) and those are the candidate nodes where the time threshold could be confidently located. The red blocks are the nine clusters predicted by the GMYC model. .. 170

Figure IV.10 Neighbour-Joining (NJ) tree of the D-loop sequences from New Zealand and the Pacific Rim. The NJ was based on HKY + G+ I nucleotide substitution model and pair-wise comparisons between sequences. From right to left, haplogroup L2 and L1 were comprised of haplotypes only found in New Zealand. Haplogroup NWP2 are haplotypes distributed from China to Philippines. Haplogroup Pacific IV is distributed from China to Japan. The bar charts on the right represent the number of haplotypes in the NJ tree that come from the Pacific location in the X-axis.

List of Tables

Table II.5 AMOVA analysis of the R_{ST} and F_{ST} variation partitioned in three hierarchical categories. Significantprobabilities are highlighted in red.47

Table III.3 Levels of genetic differentiation among sampling locations grouped by development stage and levels of genetic differentiation among age classes. Below the midline are the R_{ST} genetic differentiation values. Above the midline are the probability of each R_{ST} value in which significant values are highlighted in red (p < 0.05). 85

Table III.4 Analysis of Molecular Variance of differentiation levels within individuals (R_{IS}), among individuals (R_{ST}), among geographic regions (R_{CT} ; within individuals respective to the total variation), and total variation (R_{IT}); grouped by development stage and age classes. Significant probabilities (p < 0.05) are highlighted in red.

Table III.8 Single sample method full likelihood procedure of the population size differences assuming randommating and non-random mating such as implemented in COLONY. Ne = effective population size; CI = confidenceintervals (95%).96

Table III.11 Temporal comparisons of the variation of F-statistics between adults (generation 0) and juveniles (generation 2) in each sampling location at three critical values. Method Pk = Pollak (Pollak 1983); method NT = Nei-Tajima (Nei and Tajima 1981); method JR = Jorde-Ryman (Jorde and Ryman 2007); p = critical value (0.05, 0.02 and 0.01); H = harmonic mean; Ai = number of independent alleles; F = F-statistics at generation 0; F' = F-statistics at generation 2; Ne = effective population size; CI = Confidence intervals (non-parametric Jacknife). In red are the values that showed a significant reduction in the F-statistics as result of reduction in the Ne at each critical value.

Table IV.1 Basic genetic statistics of the sequences from COX1 partial gene and D-loop region sequences in New Zealand. n = number of samples; S = segregating sites; Nh = Number of haplotypes; h = haplotype diversity; $\pi =$

 Table IV.10 Estimation of migration ratio among geographical regions based on D-loop sequences for the 12

 New Zealand locations.

 169

List of Equations

General Introduction

Overharvesting has been a serious threat to coastal fisheries, but in recent years additional threats have emerged, which include the degradation of estuaries where juveniles often spend their early lives. Runoff from of intensive agricultural land use and urban development along the coastal margins has increased pollution levels and nutrient loadings causing shifts in biological activities and oxygen-depleted regions (Elliott and Whitfield 2011; Blaber 2013).

Estuaries are especially sensitive to anthropogenic activities such as overfishing, invasive species, intensive land use in coastal environments and high rates of sedimentation (Hewitt et al. 2004; Morrison 2005). Estuaries are high productivity-eutrophic systems with a range of biogenic habitats such as seagrass, meadows, mangroves and mussel beds (McDowall 1976; Elliott et al. 2007; Francis et al. 2011). Estuaries are already stressed environments with variable salinity levels depending on the freshwater outflow and the tidal inflow of marine water, different turbid areas due to local currents and a structural complexity due to the variety of biogenic habitats it offers (Elliott et al. 2007).

The combination of habitat characteristics allows estuaries to support a variety of species not only providing space and food, but also providing an environment for larvae and juveniles to develop before maturation which is sheltered from predatory assemblages of fish and invertebrates (Elliott et al. 2007; Elliott and Whitfield 2011; Blaber 2013). Human-based activities also have indirect effects on the migration routes, reduction of spawning grounds, reduction of nursery areas and refugia for marine, freshwater and estuarine species. The extensive land-use drained into estuaries results in habitat fragmentation and reduction that directly affects population density and distribution of marine species (McDowall 1976; Hewitt et al. 2004; Morrison 2005; Elliott et al. 2007).

Estuaries in New Zealand, are used as nursery areas by juveniles of commercially important marine fish species such as the snapper *Pagrus auratus*, the kahawai *Arripis trutta*

and *M. cephalus* known as kānae or grey mullet (Francis et al. 2005, 2011; Morrison 2005). Little is known on the impact of human activities in the genetic diversity of those species, their population genetic structure, connectivity patterns of the stocks and the genetic variation among juveniles and adult populations. The full extent of damage in estuarine areas due to human-based activities has not been comprehensively studied in New Zealand; however, in other parts of the world it has had strong repercussions into the levels of genetic diversity of estuarine and marine species that use the estuaries as nursery areas, driving the stock structure of those species into genetic bottlenecks and significant loss of genetic diversity such as in Chesapeake Bay (Richards and Rago 1999; Kemp et al. 2005; Buchheister et al. 2013).

Back in the 1960's, industry, farming and industrialized exploitation of finfish resulted in increased levels of eutrophication in Chesapeake Bay in United States. Several studies have been carried out in this area to assess the impact of anthropogenic activities on fish species. The numbers of fish species declined drastically in a ten-year period, diversity levels dropped and their distribution range were reduced due to the loss of suitable habitats for juveniles to settle (see Richards and Rago 1999; Kemp et al. 2005; Buchheister et al. 2013 and references within).

When population size declines as habitats shrink and become more fragmented, the effect of genetic drift becomes stronger (Frankel 1974; Fahrig and Merriam 1985). The consequence is a reduction in the potential of populations to respond to the fast changes in habitat and environmental conditions (Hoffman et al. 2004). Demographic decline in populations and reduction in gene flow results directly in a reduction of effective population size, there is a loss of low frequency genes and heterozygosity, and higher chances of inbreeding (Palumbi 1994). Inbreeding reduces reproductive success and the lack of genetic variation limits populations to respond to rapid local changes, to adapt if the stress persists and colonise new habitats (Hastings and Harrison 1994). Inbreeding contributes to the reduction of

the phenotypic plasticity that facilitates species to adapt to environmental changes (e.g. Muschick et al. 2011; Oufiero and Whitlow 2016; Wellband and Heath 2017).

Genetically structured populations and gene flow

The effects of genetic drift and mutation are buffered through gene flow between populations that also prevents differentiation due to selection (Palumbi 1994; Morjan and Rieseberg 2004). Higher dispersal capabilities and migratory potential of adults could translate in higher connectivity between populations (Cowen 2006, 2007). Unless dispersers and migrants are able to settle in the new areas and reproduce with the resident population (to establish), there is only demographic connectivity but no gene flow between populations (Buckley et al. 2013). The level of genetic flow among populations is a product of migration and successful reproduction at the new site (Hedgecock and Pudovkin 2011). The factors that affect the reproductive success of new migrants at a site are whether they can find a mate, genetic compatibility, the accessibility of habitat and whether it is suitable for recruitment (Parsons 1996; Grant and Bowen 1998; Ball et al. 2000).

Species with highly structured populations are common in marine environments (e.g. Rossi et al. 1998b; Huang et al. 2001; Bargelloni et al. 2003; Bernal-Ramirez et al. 2003; Sumpton et al. 2008; Hickey et al. 2009; Liu et al. 2009a; Betancur-R et al. 2010; Shen et al. 2011; Krück et al. 2013; Hughes et al. 2014). However, other marine species inhabiting the same locations may show panmictic patterns and high levels of gene flow (e.g. Neethling et al. 2008; Palm et al. 2009; White et al. 2011; Van de Putte et al. 2012; Horne et al. 2013; Varela et al. 2013). The answer to this conundrum lies first in the different dispersal capabilities of each species; and second, it lies in the factors driving the genetic subdivision in each species, where physiological differences, heritable traits and behaviour also need to be accounted as factors driving genetic subdivision (Palumbi 1994; Cowen 2007; Hedgecock and Pudovkin 2011).

Very little attention has been given to how genetic variation is maintained in coastal and estuarine populations that undergo seasonal bottlenecks. Estuaries as a sheltered environment for fish larvae and early juveniles could nurse different generations before they reach maturity and are recruited into the adult population (McDowall 1976; Elliott et al. 2007; Francis et al. 2011; Blaber 2013). The level of reproductive success in marine and estuarine fish species may change seasonally due to the fluctuation of environmental conditions that could affect spawning rates, fertilization, larval growth and survival, and that also could produce habitat fragmentation (Hedgecock et al. 2007a; Hedgecock and Pudovkin 2011). The result is a small portion of the adults successfully reproducing each season. This is the basis of the Sweepstake Reproductive Success theory or SRS (Hedgecock 1994). The SRS theory predicts that the level of genetic diversity is lower in the offspring from one season when compared to the level in the adult population. The risk for species with a sweepstake pattern of reproduction is that losing a cohort could result in significant fluctuations of gene frequencies and possibly the loss of genetic diversity within a generation.

Estuarine nursery areas that are experiencing severe environmental degradation might affect recruitment success and the maintenance of genetic diversity in the adult population (Cushman et al. 2013). Recently, estuaries in northern New Zealand that are nursery areas for *M. cephalus* have suffered environmental declines, due mainly to land-based human activities producing high sediment loads (Morrison et al. 2009, 2014b,c). Adverse effects on juvenile growth have been detected and effects on juvenile survivorship are suspected (Morrison and Mohd-Zain, unpubl. data). This increasing pressure on juvenile populations suggests a need to assess levels of genetic diversity within and among populations, test for the sweepstake effect, determine the amount of gene flow, and identify the possible dispersal processes that maintain the genetic connectivity for this species.

Grey mullet (Mugil cephalus) biology, distribution and ecology

Mugil cephalus (Linnaeus 1758; Figure I.1) commonly known in New Zealand as kānae or grey mullet, is classified in the family Mugilidae (Order: Mugiliformes). It can be differentiated from other *Mugil* species due to a thick, soft and transparent well developed adipose eyelid that covers most of the eye, with a vertical opening (Whitehead et al. 1986; Harrison 1995). The reviews of Thomson (1997) and Eschmeyer and Fricke (2011) on mullets from the family Mugilidae, have synonymised several *Mugil* species with *M. cephalus* based on the wide range of morphometric characteristics of this species.



Figure I.1<u>Mugil cephalus</u>, first described in 1758 by Linnaeus and it is known in New Zealand as kānae and grey mullet.

Mugil cephalus is considered to be a cosmopolitan species with a discontinuous distribution in both hemispheres occurring mainly in tropical, sub-tropical, temperate coastal waters, rivers and lakes between 42° N and 42° S (Briggs 1960; Thomson 1966; Whitfield et al. 2012). It is a euryhaline species and can tolerate fresh and hyperhaline waters. It can survive at a wide range of dissolved oxygen levels and has a preference for muddy or sandy habitats (Young and Potter 2002).

Mugil cephalus spawns in the sea or nearshore marine environments. Eggs and early larval stages drift in ocean currents for over a month and when they reach the stage of post-flexion larvae or early juveniles they migrate to nearer the shore and into estuaries (Wasserman

and Strydom 2011; Whitfield et al. 2012). Temperature is the main factor that triggers the spawning behaviour of *M. cephalus* (Whitfield et al. 2012). Sea surface temperatures around 20°C are the best for *M. cephalus* to spawn and the spawning seasons are variable from summer to winter in different parts of the world (Whitfield et al. 2012 and references within). It is mainly considered a gonochoristic species where each individual has one of two distinct sexes (McDonough et al. 2005). However, the presence of oocytes in developing testis has also been observed as consequences of high temperatures, starvation, overcrowding and exogenous hormones suggesting that protandric hermaphroditism (male changing to hermaphrodite) is possible (Sadovy and Shapiro 1987).

The larvae of *M. cephalus* are planktonic feeders (Zismann et al. 1975). Post-flexion larvae and juveniles enter estuaries in schools and they remain there until adulthood. After maturation they can colonise freshwater environments, remain in the estuary or move to the marine environment (Whitfield 2010). In Shark Bay, Australia, the entire life cycle takes place in the marine environment (Smith and Deguara 2003). It is also believed that one migrant population in North Taiwan is strictly marine (see Hwang et al. 1990; Chang et al. 2000; Huang et al. 2001; Chang and Iizuka 2012). Juveniles change their diet after they reach the size of 25mm from plankton to small invertebrates that migrate through the water column. In different parts of the world, juveniles have been seen to switch their feeding preferences to detritivorous feeding on benthic microalgae, foraminiferans, benthic diatoms, meiofauna and small invertebrates (see Thomson 1963; Blaber 1976; Payne 1976; De Silva and Wijeyaratne 1977; Marais 1980; Michaelis 1993; Rueda 2002; Lawson and Jimoh 2010).

Mugil cephalus is an important source of protein for a wide-taxonomic range of top predators such as sharks, crocodiles, eagles, ospreys, pelicans and dolphins (Whitfield and Blaber 1978; Martucci et al. 1993; Clancy 2005; Liordos and Goutner 2009; Fury and Harrison 2011). When *M. cephalus* matures and is recruited into the adult population its feeding

behaviour changes to foraging benthic environments and having an impact on microbenthic fauna composition and abundance (Torras et al. 2000; Whitfield et al. 2012).

The dispersal capability of *M. cephalus* or vagility is believed to be limited across large open oceans and populations associated with different landmasses are thought to be genetically isolated due to *M. cephalus* preference for coastal environments and estuarine dependence (see Rocha-Olivares et al. 2000, 2005; Hung and Shaw 2006; Jamandre et al. 2009; Shen et al. 2011; Durand et al. 2012b). Highly divergent mitochondrial DNA (mtDNA) lineages have been reported at different locations all over the world where in most cases the levels of genetic divergence among lineages exceeds the reported interspecific variation observed among species of the same genus (e.g. Rocha-Olivares et al. 2000; Heras et al. 2007; Ke et al. 2009; Liu et al. 2010; Shen et al. 2011; Sun et al. 2012; Jamandre et al. 2014). Hence, it is believed that *M. cephalus* is a species complex.

Mugil cephalus is considered by the IUCN (International Union for Conservation of Nature) red list as least concern (LC) and is not under the overexploited status neither the depleted cases assessed by the FAO (Food and Agriculture Organization of the United Nations). However, the first warning for this species was reported, where *M. cephalus* stocks in Taiwanese waters collapsed due to overfishing activities (Hung and Shaw 2006). Human populations all over the world have a long history of using *M. cephalus* as a food resource. *Mugil cephalus* bones, otoliths and paintings on ceramic have been found in archaeological sites dating from 12,600 to 500 years BC in south and north America, the Atlantic coast of South Africa, China and Australia (Bailey and Parkington 1988; Gobalet et al. 2005; Proulx 2006; Attenbrow 2010; Lunney et al. 2010; Benson 2012; Menotti and O'Sullivan 2013; Béarez et al. 2015; Russell et al. 2015). It is important to understand *M. cephalus* genetic structure patterns to ensure a sustainable fishery. In New Zealand, the stocks are considered to be under pressure, but stock status is uncertain (MPI 2016). There is evidence from commercial

fishing Catch Per Unit Effort (CPUE) that the main stock in Grey Mullet Unit one (GMU 1) may actually be comprised of six populations, with low to moderate mixing between them (McKenzie et al. 1999; McKenzie and Vaughan 2008).

Grey mullet in New Zealand, fisheries and local management

The commercial importance of *M. cephalus* in New Zealand is primarily as fishing bait for more commercially valuable species. However, it has great cultural importance for Māori Iwi, and as customary fish, especially in the northern North Island where is an important food source (Paulin and Paul 2006). The biology, ecology, geographic distribution and population dynamics of *M. cephalus* are still fundamental gaps in the knowledge of New Zealand *M. cephalus*.

Reports on the status of *M. cephalus* fisheries to the New Zealand Ministry of Primary Industries (MPI) and its predecessors have been made since 1989 and the stocks divided into East and West coast have been monitored through CPUE indexes (Anon 1989; McKenzie et al. 1999; Hartill 2004; Watson et al. 2005; McKenzie and Vaughan 2008). A life history review of grey mullet (and other species) states that it spawns in the sea from November to February (Morrison et al. 2014a). Its larvae move into the harbours/estuaries of the North Island, where they settle into nursery habitats in the upper reaches (Paulin and Paul 2006; Morrison et al. 2014a,c, 2016a). They mature at about 3 years old when males reach a length of ~30 cm and females reach ~35 cm, and have a maximum age of 14 years; although new ageing work shows a maximum age of 20 years (Morrison et al. 2016; MPI 2016). The main component of the fishery involves age classes from four to eight years old (Hartill 2004). Caution needs to be taken when adopting information from studies in other part of the world for New Zealand management fisheries. There is a lot of variation in growth rates, early-life history, habitat use, and spawning cycles (see above), that could lead to an incorrect management of the fishery. *Mugil cephalus* has a long history of supporting local fisheries since pre-European Māori and was one of the first commercial fisheries in northern New Zealand (Paulin and Paul 2006). The *M. cephalus* fishery has an approximate value of \$3 million per year; annual landings rose up from 420 tonnes in 1974 to a peak of 1,142 tonnes in 1984 (MPI 2016); however, total domestic catches declined after 1984 before the implementation of the Quota Management System (QMS; MPI, 2016). In the Kaipara Harbour, *M. cephalus* catches declined from 385 tonnes in 1996 to ~70 tonnes in 2003 (Hartill 2004; Watson et al. 2005), which caused concern and led to the implementation of a community-managed fishery in 2004 (Paulin and Paul 2006).

Mugil cephalus fishing predominately occurs in the GMU 1 area, which encompasses the northern North Island, from south of the Kawhia Harbour on the west coast, up and around Cape Reinga and down to Cape Runaway, East Cape (Figure I.2). The stock is viewed as two sub-stocks (east and west coast), separated for management purposes. The boundary between them is at North Cape (MPI 2016). It is unknown whether fishery exploitation levels are able to support the Maximum Sustainable Yield (MSY; MPI 2016). Some of this uncertainty is due to limited information on stock structures and boundaries, including genetic structure and the patterns of gene flow across the different fishery areas. A clearer picture of the number of reproductive units, and matching these to stock units, is important for stock assessment modelling (Ovenden 2013).


Figure 1.2 Major harbours/estuaries of the North Island and upper South Island where samples were collected for this research. The areas in blue are M. cephalus fishing areas in New Zealand (GMU-grey mullet units). The map on the right represents the six putative populations or stocks identified by McKenzie and Vaughn (2008) based on CPUE (Catch per unit effort).

There is a need to determine the levels of genetic diversity and genetic differentiation within and among New Zealand populations of *M. cephalus*. The knowledge required to underpin effective fisheries management will be greatly improved with more detailed information about the amount of gene flow, and the possible dispersal processes that maintain connectivity. Assessing the genetic structuring of New Zealand's *M. cephalus* fishery resource will enable management to be better focused on the key habitats and areas for the long-term sustainability of the species.

Aim of this thesis research and study structure

The overall aim of this research project is to determine the genetic population structure of *M. cephalus* in New Zealand coastal and inland waters. Identify the historical and contemporary patterns of connectivity among locations, and understand how the genetic structure changes from adults to juveniles. To address this aims, the genetic structure of *M. cephalus* populations in New Zealand was assessed using 1,095 grey mullet individuals. This was comprised of 274 juveniles (collected in 2010), 576 adults (collected in 2014), and 245 previously published mitochondrial DNA sequences retrieved from GenBank. The sequences retrieved form GenBank were from other locations in the Pacific.

The overall hypothesis of this thesis is that there is more than one genetically distinct population with non-significant or limited gene flow among locations or geographic regions (East-West). The null hypothesis is panmixia, in which there are high levels of gene flow and no genetic distinction among location or geographic regions, as considered from the management perspective in New Zealand (MPI 2016).

There are specific hypotheses to be tested in each of the approaches to understand *M*. *cephalus* population genetic structure in New Zealand. Three different approaches were used to assess the levels of genetic differentiation and connectivity patterns. The purpose of Chapter

I is to present an up-to-date summary of the research topics, methods and the species status in New Zealand and the world.

Specific Aims

• Chapter II. Insights into the contemporary genetic population structure and connectivity patterns of Kānae (*Mugil cephalus*) in New Zealand coastal and inland waters.

The hypothesis is that there is more than one genetically distinct population of *M. cephalus* in New Zealand. To test this hypothesis, we will use nuclear microsatellite DNA markers designed from previous studies to describe the population genetic structure of adult *M. cephalus* in New Zealand's coastal and inland waters. Using microsatellite DNA markers from studies in other regions of the world will allow comparisons on genetic diversity, changes in effective population sizes between New Zealand and other regions of the world. Genetic data will be assessed for spatial patterns and connectivity between the sampled locations using the spatial distribution of allele frequencies, genetic diversity, F_{ST} differentiation levels among sampling locations and test for significant changes in effective population sizes to address demographic events. The individual proportion of ancestries will be identified to cluster individuals with similar ancestries and assess the levels of genetic differentiation between populations. The results of this research could help provide better information about levels of genetic diversity, and to align the natural boundaries of wild *M. cephalus* populations with fishery stock structure boundaries. Also, could provide and insight of the evolutionary forces affecting *M. cephalus* populations in New Zealand.

• Chapter III. *Mugil cephalus* reproductive success and temporal fractioned genetic diversity between adults and juveniles in New Zealand.

The use of estuaries as nursery areas could make *M. cephalus* vulnerable to random and un-seasonal environmental changes that could increase mortality rates in

juveniles, resulting in bottlenecks and genetic diversity loss. The hypothesis is that there are significant differences in the genetic diversity between juveniles and adults and strong effects of bottlenecks in juveniles compared to the adults. To evaluate the levels of genetic diversity and differences in juvenile and adult samples of *M. cephalus*, allele frequencies will be compared between juveniles and adults and among different juvenile age classes. The juveniles will be grouped into age classes based on the size of the fish and the average growing rate during summer conditions. Tests for significant changes in the effective population size between juveniles and adults and among juvenile age classes will be performed to assess the strength of bottlenecks over time. The proportion of migrants and levels of local recruitment in each estuary, will be assessed. The outcome of this research could help to understand seasonal effects on *M. cephalus* genetic structure that can be used to better manage fishing pressures on *M. cephalus* stocks.

• Chapter IV. Population genetic structure and lineage divergence of the Kānae (*Mugil cephalus*) in New Zealand based on two mitochondrial DNA markers.

To address the demographic history of *M. cephalus* in New Zealand and the evolutionary processes shaping *M. cephalus* population genetic structure, a phylogeographic approach will be used. The phylogeographic approach also allows for comparisons with lineages found in other regions of the world and test whether or not *M. cephalus* in New Zealand is isolated. The hypothesis is that New Zealand has unique lineages and isolated with non-significant or limited gene flow from other regions of the world. Two mitochondrial DNA sequences from Cytochrome Oxidase subunit 1 (CO1) gene and D-loop region, will be used to identify the phylogeographic structure, demographic history and historical connectivity patterns of New Zealand's *M. cephalus* populations in both a local New Zealand and a broader Pacific Rim context.

Phylogenetic analyses will be performed to identify highly divergent haplotypes, groups of haplotypes and the percentage of differences between haplotype groups; to assess spatial patterns and connectivity, haplotype frequencies, haplotype diversity and Φ_{ST} levels of differentiation will be used. Changes in effective population size will be used to assess for demographic events over time. The results of this research will provide a better understanding of the diversification patterns of *M. cephalus* in New Zealand and the Pacific, and the importance of New Zealand's populations from a global perspective.

• Chapter V. General Discussion.

General discussion to review the evolutionary processes driving *M. cephalus* genetic differentiation levels and examples that support uncommon patterns of gene flow and geographic distribution.

Insight into the contemporary genetic population structure and connectivity patterns of Kānae (*Mugil cephalus*) in New Zealand coastal and inland waters

Abstract

Mugil cephalus in New Zealand is a commercially, customarily and recreationally important fishery. Its distribution is mainly around the North Island and Marlborough Sounds. There is a gap in the knowledge of *M. cephalus* population genetic structure and the factors behind its distribution in New Zealand. To describe the levels of genetic variability and connectivity, DNA was extracted from 576adult samples collected from 15 locations around the North Island and one in the South Island (Marlborough Sounds), which were genotyped using seven microsatellite DNA markers. The markers followed a stepwise mutation model (SMM) of evolution and R_{ST} was used to assess the levels of genetic differentiation. A Bayesian cluster assignment, deviations from mutation-drift equilibrium and migration rates were assessed. Four genetically distinct clusters with recent admixture were identified (K=4), and the distribution of their allele frequencies were not correlated to distinct geographic locations or show a pattern of Isolation-By-Distance ($R^2 = 0.002$; p >0.050). There were significant levels of gene flow and ~51% of the individuals were non-migrants. Over 50% of the total genetic variation could be explained by the four genetic clusters. The four genetic clusters had an overlapping distribution and were more like a spatially chaotic genetic structure. This pattern is predicted by a Sweepstakes-Reproductive-Success (SRS) model and it might have arisen because of seasonal variability in the mortality rates of juveniles, spawning behaviour and philopatry.

Keywords: chaotic genetic structure, bottlenecks, heterozygote excess, sweepstakes reproductive success, philopatry

Introduction

Genetic population structure is determined by the relative complexity of the environment that affect migration and may act as barriers to gene flow. Within the marine environment, cosmopolitan fish species tend to show genetic homogeneity along their distributional range, because it only takes a small number of migrants to homogenise the genetic variation between populations (Hauser and Carvalho 2008; Selkoe et al. 2010; Hedgecock and Pudovkin 2011; Ovenden 2013). The level of gene flow permeability through environmental barriers can vary over time, which means sporadic gene flow might be a relatively common phenomenon but one that is difficult to document without long-term observations (Lehodey et al. 2006; Gaither et al. 2011; Reid et al. 2016; von Biela et al. 2016). The changes produced by variable environmental barriers could lead to genetic heterogeneity within and between populations that do not appear to be correlated with geographic distances and/or contemporary environmental factors (e.g. Hogan et al. 2010; Broquet et al. 2013; Selkoe et al. 2014; Selwyn et al. 2016).

Species that show genetic heterogeneity between neighbouring populations and genetic homogeneity between distant populations, are defined as having spatial genetic patchiness, or a chaotic genetic structure (Hogan et al. 2010; Broquet et al. 2013). Marine species that show this type of genetic structure often have a significant amount of genetic differences between juvenile recruits and adults (Hedgecock 1994; Hogan et al. 2010; Hedgecock and Pudovkin 2011). This is thought to be caused by the seasonal variability of local oceanographic conditions that affect survival rates during early-life history stages (*i.e.* Hedgecock 1994; Huang et al. 2001; Bekkevold 2002; Swearer et al. 2002; Bekkevold et al. 2007; Hedgecock and Pudovkin 2011; Reid et al. 2016). The result is seasonal bottlenecks in juveniles in which only a small fraction of the genetic diversity found in adults is represented in the surviving recruits (Hedgecock 1994; Hedgecock and Pudovkin 2011). The reproductive success of the

adults as a consequence of seasonal differences in juvenile mortality rates forms the basis of the sweepstake-reproductive-success (SRS) theory (Hedgecock 1994; Hedgecock and Pudovkin 2011).

The SRS could also be strengthened by reproductive behaviours leading to non-random mating such as philopatry (Hendry and Day 2005; Maes et al. 2006). While migration is a key factor that determines the amount of genetic homogeneity across large-scale distances, immigrants that arrive successfully to a location may not be able to easily integrate into a randomly mating group of resident individuals. Reproductive differences in mating cues (e.g. temperature, salinity) could produce asynchronous mating times (Hendry and Day 2005; Maes et al. 2006).

Mugil cephalus is a species closely associated with coastal environments and often considered estuarine dependent, yet it has a world-wide distribution in temperate and tropical waters between ~42° N and ~42°S (Whitfield et al. 2012). A variety of studies that used nuclear DNA markers such as allozymes, single nucleotide polymorphisms (SNP's) and microsatellite DNA (msatDNA) to assess *M. cephalus* population genetic structure, reported genetically distinct populations in Florida and along the China, Taiwan and eastern Australian coasts (see Campton and Mahmoudi 1991; Rossi et al. 1998b; Huang et al. 2001; Liu et al. 2009b; Blel et al. 2010; Shen et al. 2011; Krück et al. 2013). Highly structured populations in neighbouring estuaries were reported along the coast of China assessed with microsatellite DNA (msatDNA) and mitochondrial DNA (mtDNA) markers (Huang et al. 2001; Liu et al. 2009b; Shen et al. 2011). However, there was a lack of differentiation across similar spatial scales in the Mediterranean Sea that were assessed with a set of different msatDNA markers than those used along the coast of China (Rossi et al. 1998b; Blel et al. 2010). In the eastern coast of Australia, a north to south break was found with single nucleotide polymorphisms (SNP's), possibly correlated to northern and southern populations spawning at different temperatures (Krück et

al. 2013). A study that used allozymes of 27 gene loci showed low differentiation levels among North Atlantic, Mediterranean Sea, Africa and Eastern Australia; western Australia was clustered with Galapagos Islands (Rossi et al. 1998b). There has not been a study that used the same nuclear markers to assess the genetic structure of *M. cephalus* making the outcome from each study non comparable. The seemly contradictory levels of genetic differentiation in each location and across large spatial scales could be a result of the markers used or could be a consequence of undetected processes. For example, sporadic migration could potentially promote intermittent gene flow over large distances, enough to maintain the cosmopolitan *M. cephalus* group as a single species with large effective population size and genetically structured populations.

In New Zealand, *M. cephalus* is distributed around the North Island and as far South as the Marlborough Sounds at the top of the South Island (Morrison et al. 2016). Local Ecology Knowledge assessments (LEK; unpubl. data), reported an extended southwards distribution where *M. cephalus* has being customarily fished in Dunedin (~45° S; Don Nelson 2016, personal comment). *Mugil cephalus* is known for seasonal migrations into estuaries and fresh water systems. Spawning typically occurs in the sea from October to March. Eggs and early larval stage fish drift in coastal currents and use tides to get into harbours and estuaries, in which they reach maturity at around three years old (Hartill 2004; Paulin and Paul 2006).

Commercial fishing of *M. cephalus* occurs predominantly in the area GMU1 (Grey mullet Fishery Management Area 1; Figure I.2), where most of New Zealand's annual landings are recorded (Hartill 2004; McKenzie and Vaughan 2008; MPI 2016). *Mugil cephalus* in GMU1 is considered a panmictic population comprised of two stocks (East and West), separated by North Cape, for fishery management purposes by the Ministry of Primary Industries in New Zealand (MPI 2016). Six administrative stocks have been recognized through the assessment of catch per unit effort (CPUE) in the major estuaries of the GMU1 (McKenzie

and Vaughan 2008). CPUE analysis for *M. cephalus* are not totally reliable because there are different fishing practices in different areas of New Zealand; seine nets, set nets, tow nets are some of the commercial fishing methods and the eye mesh of the net varies across New Zealand regions (McKenzie and Vaughan 2008). For example, in Kaipara Harbour the legal eye mesh size of the nets used to catch *M. cephalus* commercially is 90mm while in Wellington the legal size is 100mm (MPI 2016).

The aim of this study was to use microsatellite DNA (msatDNA) markers used in *M*. *cephalus* population genetic studies in the west Pacific and the Mediterranean, to determine the genetic structure of New Zealand's *M. cephalus* populations sampled from coastal, estuarine and inland waters. The data was used to describe the levels of genetic variability, connectivity patterns and test for evidence of genetic bottlenecks or selection.

Materials and Methods

Data collection

Tissue samples from 576 adults were collected from 15 locations around the North Island and one the South Island of New Zealand (Figure II.1). Tissue biopsies were taken from the caudal peduncle and DNA extracted using a phenol/chloroform protocol followed by ethanol precipitation. The purified DNA was re-suspended in TE buffer and stored at 4°C.

Amplification and genotyping

Seven msatDNA loci were obtained from the studies of Miggiano et al. (2005) and Shen et al. (2010) (Table II.1). PCRs were conducted in 15 μ l volumes that contained 1.5 or 2 mM MgCl₂, 0.4 μ M of the forward primer labelled with M-13 tail, 0.6 μ M reverse primer, 0.6 μ M of fluorescently-labelled M-13 primer, 0.6 μ M dNTP's, 0.4 μ g/ul of bovine serum albumin, 20 mM Tris HCl (pH 8.4), 50 mM KCl, and 1 unit of Taq DNA-polymerase (Bioline). Between 50-80 ng/ μ l of template DNA was added. The thermal cycling conditions for each locus is specified in Table II.1. The size of the amplified products were determined using a 3730 Applied

Biosystems Genetic Analyzer (Macrogen Inc., Korea). Resulting chromatograms were visualized in Genemarker v2.2 (SoftGenetics, LLC), and the peaks were scored by eye and recorded manually. The alleles were coded as three digit genotypes and the individuals containing missing data at more than two loci were removed from the subsequent analyses.



Figure II.1 Locations where <u>M. cephalus</u> samples were collected between November 2014 and April 2015.

Loci assessment and genetic diversity

Genotyping errors were evaluated using Micro-Checker v2.2.3 (Van Oosterhout et al. 2004) to check for null alleles, large allele drop-out, allele size shifts and scoring errors due to stuttering, following the Hardy-Weinberg equilibrium (*HWE*) expectations of heterozygosity.

Estimates of the number of alleles per locus, observed heterozygosity (*Ho*), expected heterozygosity (*He*), Wright's coefficient of inbreeding (F_{IS}), proportion of different alleles per locus, Shannon's allelic diversity, and the number of alleles that are different and equally frequent in the data set (number of effective alleles) by locus and sampling location, were calculated in GenAlEx v6.5 (Peakall and Smouse 2012), Arlequin v3.5 (Excoffier and Lischer 2010), and FSTAT v2.9.3.2 (Goudet 1995). Allele frequency distribution plots by locus and population were generated in GenAlEx v6.5. Exact tests for deviation from *HWE* by locus and population were calculated in GENEPOP v4.2 (Rousset 2008); and pairwise linkage disequilibrium tests were estimated using Arlequin v3.5.

Significant values of HWE were calculated using the exact test of Guo and Thompson (1992), with a Markov chain of 10,000 dememorization steps. Significant values for linkage disequilibrium (LD) were calculated using a likelihood-ratio test (LRT) of 20,000 permutations (Slatkin and Excoffier 1996), both as implemented in Arlequin v3.5. Sequential Bonferroni corrections were used to adjust p-values for multiple comparisons (Rice 1989). To identify outlier loci that could be under the positive or balancing selection a coalescent approach was performed with LOSITAN (Antao et al. 2008). We used the method described in Beaumont and Nichols (1996), simulating the distribution of loci based on the F_{ST} and He values. Both an Infinite Allele Model (IAM) and the Step-wise Mutation Model (SMM) were used. A total of 205, 000 simulations were run with a neutral and forced F_{ST} to improve the reliability of the mean values of F_{ST} and 99% confidence intervals. To assess if the sample size was sufficient to reflect the polymorphisms in the population, allele-discovery curves were calculated in R (R Development Core Team), using the PopGenKit v1.0 package (Rioux-Paquette 2012). POWSIM 4.1 (Ryman and Palm 2006) was used to assess the statistical power of the sampled loci. The statistical power and alpha values were assessed using the Chi-squared and Fisher's exact tests under a specified level of population divergence taken from the lowest significant

 F_{ST} value between sampling locations, with 10,000 iterative runs and simulation of 40 generations increasing the effective population size under *HWE*.

R conditions
ers and PCH
îication prim
eats, amplij
marker rep
ellite DNA
1 Microsat
Table 0.

Repeat	Repeat	Primers (5' - 3')	Fluorescent Dye	Allele range	Source	PCR conditions
Mco3	E.	GGAGGACTAGGATTTCGTG	EAM	130.748	Shen et al	05° fanin. (01° 30º: 58° 30º: 77° 00°,401° 30º: 53° 15º: 77° 15° 48° 17° 30min
MCCO	(T)23	CTTATCCCTGACCTTCTCAT	INTELL	0+7-001	2010	111110 71 '00'67 71 '00' 10' '00' 40' '00' '00' '00' '00'
Maalkew	ŨĿ	CAGATTGTTCGGGGGGGGGCAGA			Miggiano et	050 5in. (040 30 £30 30 770 00
INTERTOTSOLA	(D)13	GTCATGATGCTGCTATCAGGCAAA		716-077	al 2005	111110 71 '00/16-4 71 '40-4 CC '400 46/ ACV/60 71 '400 CO '40' +6' '11111 C CC
Mca3		AGCCAAAGTTCCTGTAAGTACG	VIC	00-200	Shen et al	05° fanin. (01° 30e. 60° 30e. 77° 00e.v30.00° 30e. 53° 15e. 77° 15e.v8. 77° 30min
MICCA		TCAGATTAGGACCCGACCATA		007-77	2010	HILLOC 71 YON/GCH 71 YOFH CO YONG HAYAON/GOA 71 YONG AN YONG HAY HILLOC CA
Mael6DM	ET.	ACAAATACTAGATGCTTTGCACACCCTTAC	DET	116.208	Miggiano et	05° 5min. (01° 30° 61° 30° 77° 00° 40° 30° 52° 15° 17° 15° 8° 17° 30min
MICSIONIA	(01)30	GGCCTACAGATITCCTTCTTTATGCTAGGC	191	140-200	al 2005	1000 71 '00'07 77 '00' '00' '00' '00' '0
Mcon	GATAGACALAN	ACTGTGCACTTCTGGTTTTCC	EAM	117 170	Shen et al	05° f.min. (04° 3An. 58° 3An. 77° 0An.v3Av04° 3An. 53° 45° 45° 45° 40° 3Amin
		ACATCTTTGAGGTTGCCC	INTE I	6/1-/11	2010	111110 71 100/102 71 100 100 100 100 100 100 100 100 100
HACSOM		GCAGTGAAATAAACAGGCACCTCTCA	EAM	164 374	Miggiano et	05° 5min. (04° 30n. 63° 30n. 77° 00. by 04° 30n. 53° 45r. 77° 45r. 97° 30min
MICS ZF H		TGTGGGGTATGAAGTCCTCTTGCTC	INTELL	104-724	al 2005	111110C 71 'WYGC+ 71 'YC+ CC 'YOC +64'YOCYGA6 71 'YOC CO 'YOC +64' '11111C C6
Mco2A		AAAGTCTGAATGGAGCGCAG	VIC	104-154	Shen et al	05° fanin. (01° 30º: 60° 30º: 77° 00°,420.00° 30º: 53° 15c. 77° 15c.v.8. 77° 30min
WICH -	llaren	TCCTGAATCTCTGGTGACAGC		+01-+01	2010	111110 71 '00'07 71 '07' 10' '00' +0' '00' +0' '00' '00' '00' '

Population structure

The levels of genetic differentiation among sampled locations were evaluated measuring the variation in allele frequencies assuming IAM with the F_{ST} adapted from Weir and Cockerham (1984). This approach accounts for different sample sizes and uses standardized variation within and among sampling locations through an ANOVA approach in GENEPOP v4.1. Also, pairwise R_{ST} values (Slatkin 1995), were calculated in Arlequin v3.5 and GenAlEx v6.5. R_{ST} values assumed SMM where one mutation changes the allele's length by one repeat unit. Jost's *D* values (Jost 2008), and F'_{ST} (maximum possible value of genetic differentiation with pairwise comparisons given the observed amount of within-population variation), were calculated in GenAlEx v6.5. P-values were adjusted with Bonferroni correction for p-values under multiple comparisons.

The genetic relationships among sampling locations were evaluated using a Principal Component Analysis (PCA) based on F_{ST} values in PCA-GEN v1.2.1 (Goudet 1999) and a Principal Coordinates Analysis (PCoA) based on co-dominant genetic distances in GenAlEx v6.5. In the PCA, individuals are grouped into clusters where the variation from each cluster is subtracted from the total variation (Goudet 1999). The PCoA maximises the correlation between genetic distances where the coefficient of correlation is a dimension or coordinate, in order to cluster distances with similar correlation coefficients compared to the first three principal coordinates (Peakall and Smouse 2012). The individual genetic differences for each sampling location were assessed with a three-dimensional Factorial Component Analysis (FCA) using Genetix v4.05 (Belkhir et al. 2004). The linearized F_{ST} distances and the natural logarithm of the geographic distances along the coast line were used to test for IBD. Significance and linear regression of pairwise combinations among both matrices were evaluated in Arlequin v3.5 using the Mantel's test with 10,000 permutations.

Levels of genetic diversity within and among sampling locations, across New Zealand's regions (East-West-South) and the groups identified in the PCA, PCoA and FCA, were

calculated using an Analysis of Molecular Variance (AMOVA) in Arlequin v3.5 and GenAlEx v6.5. AMOVA runs assuming the null hypothesis of no structure and panmixia among sites. The F and R-statistic values and the variance components are partitioned in three levels of hierarchical subdivision: among groups (sampling locations, geographic regions and/or PCA groups), within population (within each of the groups) and total variance from the interaction of among and within populations. The genetic distance matrix for AMOVA was estimated by pairwise differences in Arlequin v3.5 and co-dominant genetic distances in GenAlEx v6.5. The significance levels of the variance components of F and R-statistics values were tested by 20,000 nonparametric permutations and 999 permutations in Arlequin v3.5 and GenAlEx v6.5 respectively.

Bayesian population clustering and spatial distribution

In order to evaluate the extent of the genetic structure based on the allele frequencies across the seven loci and F_{ST} comparisons within and among localities, a Bayesian approach was implemented using STRUCTURE v2.3.4 (Pritchard et al. 2000). STRUCTURE v2.3.4 assigns multi-locus genotypes to an ancestral cluster and calculates the maximum likelihood of the number of clusters that contribute to each individual ancestry and can explain the allele frequency distribution. It estimates the ancestry allele frequencies and expected allele frequencies, minimizing the *HWE* and *LD* in the original data. A Markov chain Monte Carlo (MCMC) simulation was run for 400,000 iterations and 1,000 permutations with 10% for burnin period (Falush et al. 2003). Admixture model was used with correlated frequencies, sampling locations as a prior, K=1-16 assuming one population per sampling location as maximum number of clusters. Admixture models assume that each individual has an ancestry from one or more genetically distinct populations (K). Locality prior (Locprior) models were used to detect genetic structure and they are helpful when low levels of genetic divergence are found or there is a limited number of loci (Hubisz et al. 2009). STRUCTURE HARVESTER v0.6.92 (Earl and Von Holdt 2012), was used to evaluate the results and estimate K (number of ancestral clusters) using the Log probability of data (L(K)) and plotting L(K) and Delta K (Δ K) across multiple runs following simulations (Evanno et al. 2005). Delta K is based on the rate of change in the log probability of data between successive K-values (Evanno et al. 2005). The probabilities of each model are compared for each value of K, in which individuals were assumed to have been correctly assigned to an ancestral cluster or population when their q value (*i.e.* its posterior probability of belonging to original ancestral population) was at least 80% for the population (Pritchard et al. 2000).

Mutation-drift equilibrium

It is expected that individuals coming from the same population reflect similar patterns in genetic diversity and show similar allele frequencies. Therefore, evidence from effective population size changes, changes in allele frequencies and drift-mutation equilibrium was evaluated. The excess of heterozygotes was calculated in Bottleneck v1.2.02 (Piry et al. 1999), where He was estimated from the number of observed alleles at each locus (k) and a simulated distribution of heterozygosity was generated with the coalescence approach under the three msatDNA evolutionary models: IAM, SMM and the Two-Phase Model (TPM). The average expected equilibrium heterozygosity (Heq) for each locus was compared to He to assess whether there was significant heterozygote excess or a deficit of heterozygotes at each locus using three approaches: a Sign test, standardized differences test, and Wilcoxon's test based on 1000 replications. The test for shifts in the allele frequency distribution from an L-shaped (allele frequency distribution of population that are under mutation-drift equilibrium) were performed in Bottleneck v1.2.02. Differences in population size and a population size contraction signature were assessed with the Garza-Williamson analysis as implemented in Arlequin v3.5 and M-RATIO (Garza and Williamson 2001). This approach calculates M, which is the ratio of total number of alleles (k) compared to the range of allele sizes (r). A simulated distribution of M values assumed to be under mutation-drift equilibrium were calculated in M_P_VAL v100.0 (Garza and Williamson 2001), to evaluate the differences between the distribution of simulated and observed M-ratios. Mutation rates and the effective population sizes were calculated in Arlequin v3.5 for 10,000 iterations, from expected homozygosity assuming populations are in mutation-drift equilibrium.

Migration and contemporary connectivity patterns

Migration ratios and the number of effective migrants among sampling locations were assessed in Migrate-n v3.6 (Beerli 2009), using a Metropolis-Hastings algorithm to sample multi-dimensional distributions of Theta (θ) with Maximum likelihood (ML) and with Bayesian inference (BI) approaches (Beerli and Felsenstein 1999, 2001). One hundred seeds, 50, 000,000 MCMC iterations, a thinning of 1,000,000 and a sampling interval of 1,000 were used. Migrate-n uses a coalescence approach to determine the migration rates between locations using the genealogy to the most recent common ancestor (MRCA). The genealogy is built including all loci in which each step to the MRCA is defined as an ancestral generation and it is drawn as a node on a line that merges all the way back to the MRCA (Kingman 2000). Migrate-n calculates the probability that two individuals in a particular generation have a common ancestor (Beerli 2009). The probabilities are calculated integrating all possible relationships in each ancestral generation following the genealogical tree. To estimate the number of immigrants per generation, the total migration rate (M) and effective population size (θ) per generation are scaled to the mutational rate (m) (Beerli 2009). In each node or generation, the effective population size or theta (θ) is calculated. To obtain the value of θm (theta scaled to the mutational rate), theta is multiplied by the mutation rate (m) that is usually inferred from the data set, and multiplied by a constant that depends on the ploidy level and the inheritance of the DNA markers (χ) (Beerli 2009). In bi-parental nuclear markers $\chi = 4$ (two allele copies per parent). The immigration rate in each generation (M) is calculated using θm and multiplying it by the number of migrants and dividing it by the mutation rate (Beerli 2009). The result reflects the number of variants brought to the population each generation that are not a consequence of mutation. Migrate-n integrates all migration rates (M) and effective population sizes per generation (θm) to estimate the number of migrants since the MRCA (Beerli 2009).The results were represented graphically with the R package circlize (Gu et al. 2014). The width of the curved lines in the circular plot represents the number of migrants and is dependent on the total volume of migrants across all regions. An assignment analysis was calculated based on the expected frequency of each individual's genotype and subsequently assigned to the population where the expected frequency was highest. The highest likelihood and p-values as first generation migrants between locations were performed using the software Geneclass2 (Piry et al. 2004), using the Rannala and Mountain Bayesian method (Rannala and Mountain 1997). Likelihood estimations were calculated with L_home. The probability that each individual is not a migrant but a resident was estimated using 1,000,000 simulations (Monte Carlo resampling) under Paetkau's algorithm (Paetkau et al. 2004).

Results

Loci assessment and genetic diversity

The seven loci were successfully amplified for all 576 individuals and there were no missing genotypes in the data set. Based on the *He* and the *HWE* allele frequencies, there was no evidence of null alleles, allele drop-out or shifts in allele size due to mistakes in scoring.

Table II.2 Summary of the seven microsatellite DNA markers. N_A = number of different alleles; A = proportion of different alleles; Af = number of effective alleles; Ar = Shannon's index of allele diversity; H_0 = Observed heterozygosity; He = Expected heterozygosity; F_{IS} = inbreeding coefficient; F_{IT} = average heterozygosity within the population relative to the total expected heterozygosity; F_{ST} = average heterozygosity between populations relative to the total expected heterozygosity; HWE = probability of deviation from Hardy-Weinberg equilibrium after Bonferroni correction.

Locus	NA	A	Af	Ar	Но	He	F _{1S}	F <i>IT</i>	F _{ST}	HWE (p)
Mce3	55	18.813	7.121	2.338	0.837	0.846	0.011	0.049	0.038	0.996
Mcs16EM	24	14.375	9.243	2.369	0.887	0.883	-0.005	0.026	0.031	0.905
Mce2	23	9.375	4.513	1.759	0.816	0.766	-0.064	-0.001	0.060	0.193
Mcs16DM	16	11.438	7.177	2.121	0.883	0.847	-0.043	-0.015	0.027	0.801
Mce27	19	7.250	3.586	1.480	0.765	0.699	-0.094	-0.056	0.035	0.957
Mcs2FH	29	13.125	7.204	2.141	0.959	0.847	-0.133	-0.093	0.035	0.231
Mce24	23	13.563	7.029	2.210	0.933	0.851	-0.096	0.055	0.037	0.737

Most of the loci were highly polymorphic with an average of 20 alleles per locus. Locus Mcs16DM had the lowest number of alleles and Mce3 had the highest number of alleles (16 and 55 respectively; Table II.2). The marker Mce27 had the lowest value of observed heterozygosity (Ho = 0.765) and Mcs2FH had the highest (Ho = 0.959). The lowest proportion of expected heterozygotes was found in the marker Mce27 (He = 0.699) and the highest value was found in the marker Mcs16EM (He = 0.883; Table II.2). There was no evidence of significant deviation from HWE in each locus before and after Bonferroni correction for p-values in multiple tests (Table II.2). There were paired combinations of loci that showed significant LD values (Appendix Table II.8). The linkage of loci was not consistent across sampling locations; for example, the msatDNA marker Mcs16DM was linked with Mcs16EM and McS2FH markers in Waikato, and was significantly linked to Mce3 in Kaipara Harbour. Linkage Disequilibrium is not expected to introduce bias into any of the analyses because is not a measure of the distance between loci (Slatkin 2008). The probability to find Linkage between pairs of loci that have been inherited from a shared recent common ancestor, even when they are not located in the same chromosome (Marth et al. 2004; Lawrence et al. 2005;

Achaz 2009; Huang et al. 2014; Skelly et al. 2016). Although the concept itself refers to the genetic linkage in loci we assume physically proximate, genetic association measurements such as *LD* are not a linear function of distance between loci (Lawrence et al. 2005; Skelly et al. 2016). This concept comes from the algebraic basis of how *LD* is calculated, where if the probability at which the frequency of two alleles occur together is equal to the product of their probabilities, the loci are not statistically independent and they are inherited together or linked (Slatkin 2008).

The allele frequency distributions (Figure II.2), showed similar allele frequencies between Waikato and Bay of Islands, Hokianga and Dargaville, Wairarapa and Hauraki Gulf, and Whangape and Waimihia. The distribution of allele frequencies was consistent with SMM evolution pattern (Figure II.2) Based on both Nei's allele diversity (defined as the probability of two random alleles being different in the sample), and Shannon's diversity function modified for allele frequencies, all sampling locations held similar levels of genetic diversity (Table II.3). Shannon's genetic diversity index has a better level of resolution because it takes into account the sample size and is not biased by the missing alleles in each locality. Waiotahe, Whakatane, Hauraki Gulf, Wairarapa and South Island locations showed genetic diversity levels under two (Ar = 1.907; Ar = 1.721; Ar = 1.877; Ar = 1.894; Ar = 1.576, respectively; Table II.3). Those same locations showed low number of private alleles (zero and one; Table II.3). The South Island location showed the lowest value of genetic diversity and no private alleles probably as a result of low sample size (N = 5; Table II.3). Kaipara Harbour showed high levels of genetic diversity (Ar = 2.117; Table II.3); however, there were not private alleles found in this location. Kaipara Harbour is likely to an area with high proportion of individuals coming from other locations. Napier, Bay of Islands, Dargaville, and Waikato showed the highest numbers of private alleles by sampling location (nine, seven, seven and six respectively; Table II.3).















Figure II.2 Cont.







Figure II.2 Cont.















Table II.3 Basic statistics summaries for each sampling population. N = total number of samples; A = proportion of different alleles; Af = number of effective alleles; Ad = allele diversity; Ar = Shannon's index of allele diversity; P_A = number of private alleles; H₀ = Observed heterozygosity; He = Expected heterozygosity; F_{IS} = inbreeding coefficient.

Dopulation	Latituda	Longitudo	N	4	٨f	Ad	4 =	D .	H	He	Fre
Topulation	Lautuue	Longitude	11	A	AJ	Au	A1	IA	110	ne	175
Napier	-39.483989	176.87965	48	16.000	7.996	0.8635	2.272	9	0.854	0.855	0.002
Waiotahe	-38.038384	177.18284	12	9.000	5.888	0.8447	1.907	1	0.905	0.810	-0.124
Whakatane	-37.948243	176.99674	6	7.000	5.206	0.8354	1.721	0	0.786	0.766	-0.025
Hauraki Gulf	-36.829214	174.83479	48	12.000	4.996	0.7894	1.877	0	0.967	0.781	-0.245
Bay of Islands	-35.237716	174.09942	48	14.429	7.101	0.8357	2.169	7	0.827	0.829	0.001
Waimihia	-35.327874	173.1786	47	15.143	7.892	0.8648	2.285	4	0.935	0.856	-0.095
Hokianga	-35.407473	173.4328	48	15.000	7.480	0.8458	2.213	3	0.821	0.837	0.014
Dargaville	-35.981893	173.75731	48	15.571	6.879	0.8397	2.178	7	0.899	0.831	-0.086
Pouto	-36.301823	174.08338	38	12.714	6.983	0.8506	2.114	2	0.820	0.839	0.027
Kaipara Harbour	-36.341742	174.22599	48	14.000	6.840	0.8335	2.117	0	0.848	0.825	-0.029
Waikato	-37.557986	175.12449	48	14.571	7.849	0.8498	2.221	6	0.860	0.842	-0.019
Whangape	-37.460357	175.06742	48	15.000	7.297	0.8593	2.234	5	0.921	0.850	-0.086
Patea	-39.546669	174.56649	24	12.571	6.941	0.8671	2.161	2	0.905	0.849	-0.067
Porirua	-41.097363	174.90011	12	10.000	6.072	0.8638	2.013	2	0.833	0.828	-0.008
Wairarapa	-41.256098	175.20365	48	12.286	4.786	0.7814	1.894	1	0.917	0.773	-0.181
South Island	-41.243534	174.04032	5	5.714	4.642	0.8317	1.576	0	0.800	0.749	-0.073
То	tal (mean)		576	36.000	6.553	0.8410125	2.060	3.0625	0.869	0.820	-0.062

Most of the msatDNA loci appeared to conform to neutral expectations. Only the locus Mce2 showed a pattern that could be interpreted as positive selection after the simulations under both the IAM and SMM evolutionary models (p < 0.01; Appendix Figure II.8). The loci Mce3 and Mcs2FH did not reach a plateau of allele discovery (Figure II.3), so additional alleles could be discovered if the sample size was increased. Simulations of divergence from a single ancestral population conducted in POWSIM 4.1, showed that the msatDNA loci used in this study had enough statistical power to detect significant pairwise differentiation levels in F_{ST} as low as 0.0025 after 40 generations under *HWE* equilibrium (Table II.11).

Population structure

There were low levels of genetic differentiation among sampling locations with both pairwise F_{ST} and Slatkin's R_{ST} statistics being significant ($F_{ST} < 0.05$; $R_{ST} < 0.01$; Table II.4; Appendix Table II.7). Slatkin's R_{ST} statistic is better suited to assess gene flow and isolation when multiple loci and alleles are used. It takes into account the total variance in allele size across all locations and the variance in allele size within each location (Slatkin 1995). Slatkin's R_{ST} statistics were more suitable for assessing differentiation with msatDNA markers under SMM evolution like the loci in this study (Figure II.2).

The AMOVA analysis showed significant differences within sampling locations and among individuals which is characteristic of highly structured populations overlapping in geographical distribution (Table II.5). Slatkin's R_{ST} showed significant levels of differentiation in all thre hierarchical levels while F_{ST} only showed significant differences among individuals (Table II.5). Four groups were identified in the PCA (Figure II.4), where two axes explained over 50% of the total variation. Wairarapa and Hauraki Gulf were grouped together as well as Waikato and Bay of Islands, and Patea and Porirua. The variation of the PCA1 and PCA2 was not enough to identify other groups in the remaining locations. The PCoA performed in GenAlEx where the coordinates explained 14.57 and 10.16% of the variation showed no clear groups. With the FCA performed in Genetix, the factorials explained 42% of the variation and showed similar clusters as the ones found with PCA (see above). The individual genetic distances are small between populations that have undergone recent admixture (Belkhir et al. 2004). Wairarapa and Hauraki Gulf, and Waikato and Bay of Islands, were clustered together (Figure II.5). Some of the individual genetic distances of Napier and Dargaville overlapped forming a cluster as well as Waimimihia and Whangape Lakes which showed a cluster of overlapping of individuals genetic distances (Figure II.5). Samples from two contiguous locations, Kaipara Harbour and Pouto, were scattered suggesting that the three factorial components could not explain the variation found (Figure II.5).

The Mantel test of correlation coefficient between the linearized R_{ST} genetic distances and the linearized geographical distances, showed low and no significant correlation ($R^2 = 0.002$; p > 0.05). There was no pattern of Isolation-By-Distance.

μ	
wi	
es	
lqτ	
d ta	
ileı	
etai	
D	
d.	
re	
ë ir	
are	
es	
ulu	
t ve	
an	
ific	
gni	
Si	
ul.	
энс	
<i>ag</i> (
di_{li}	
he	
10	
log	
r al	
F_{S1}	
ul;	
энс	
<i>ag</i> (
di	
the	
M	
e l o	
$_T b$	
R_S	
es.	
lex	
ina	
CS	
Sti	
tat	
-S	
Η	
wit	
Sə	
liti	
ca	
lo	
ing	
ldı	
an	
00	
пол	
an	
ш	0
ati	I.I
nti	le
210	r_{al}
liffe	n I
сa	re i
eti	_
-	s a
Jen	ues a
2 Gen	values a
: 0.2 Gen	p-values a
ble 0.2 Gen	act p-values a

	Napier	Waiotahe	Whakatane	Bay of Islands	Hauraki Gulf	Waimimihia	Hokianga	Dargaville	Pouto	Kaipara Harbour	Waikato	Whangape	Patea	Porirua	Wairarapa	South Island
Napier	*	0.016	0.012	0.021	0.027	0.014	0.008	0.004	0.013	0.014	0.023	0.013	0.020	0.023	0.032	0.014
Waiotahe	-0.004	*	-0.006	0.033	0.039	0.011	0.019	0.022	0.005	0.015	0.037	0.013	-0.002	0.005	0.039	-0.002
Whakatane	0.026	0.122	*	0.005	0.035	-0.001	-0.004	0.007	-0.005	-0.004	0.015	-0.007	-0.004	0.009	0.023	-0.040
ay of Islands	060.0	0.209	-0.046	*	0.039	0.010	0.004	0.012	0.015	0.008	0.000	0.012	0.025	0.031	0.036	0.005
Iauraki Gulf	0.065	0.181	-0.00	0.029	*	0.033	0.028	0.022	0.037	0.014	0.049	0.025	0.044	0.059	0.00	0.044
Waimimihia	0.123	0.223	-0.014	0.038	0.015	*	0.002	0.008	0.007	0.007	0.014	0.003	0.009	0.015	0.030	-0.002
Hokianga	0.048	0.128	-0.044	0.002	0.005	0.029	*	0.000	0.008	0.002	0.011	0.007	0.021	0.024	0.023	-0.009
Dargaville	0.013	0.069	-0.035	0.022	0.024	0.072	0.003	*	0.00	0.007	0.017	0.007	0.021	0.023	0.023	0.008
Pouto	0.052	0.145	-0.046	0.003	0.010	0.030	-0.008	0.005	*	0.015	0.018	0.008	0.004	0.017	0.035	-0.001
ipara Harbour	0.025	0.099	-0.029	0.026	0.002	0.046	-0.003	-0.001	0.004	*	0.014	0.008	0.020	0.027	0.011	0.000
Waikato	0.016	0.080	-0.027	0.030	0.036	0.090	0.008	-0.012	0.013	0.005	*	0.018	0.028	0.028	0.047	0.016
Whangape	0.092	0.158	0.002	0.051	0.015	0.017	0.034	0.058	0.042	0.032	0.068	*	0.010	0.022	0.019	-0.001
Patea	0.066	0.139	-0.038	0.029	0.008	0.002	0.016	0.034	0.007	0.019	0.054	0.017	*	0.004	0.048	0.003
Porirua	0.162	0.228	0.009	0.094	0.130	0.065	0.101	0.113	0.078	0.143	0.145	0.128	0.036	*	0.059	0.019
Wairarapa	060.0	0.242	-0.041	0.002	0.007	0.019	-0.001	0.024	-0.003	0.018	0.036	0.033	0.015	0.109	*	0.021
South Island	-0.018	0.043	-0.060	-0.002	-0.020	-0.018	-0.046	-0.041	-0.033	-0.059	-0.031	-0.008	-0.044	0.047	0.001	*
Table II.5 AMOVA analysis of the R_{ST} and F_{ST} variation partitioned in three hierarchical categories. Significant probabilities are highlighted in red.

				F	RST					
Source	df	SS	MS	Est. Var.	%	R-S	tatistics	р	Geographic region	ns (West-East-South)
Among Populations	15	69106.623	4607.108	46.462	4%	R _{IT}	0.179	0.001	$R_{IT} = 0.180$	0.001
Among Individuals	560	734193.599	1311.060	167.022	14%	R _{ST}	0.039	0.001	$R_{ST} = 0.008$	0.014
Within Individuals	576	562761.000	977.016	977.016	82%	R _{IS}	0.146	0.001	$R_{\rm IS}=0.173$	0.001
Total	1151	1366061.222		1190.500	100%					
				F	TST					
Source	df	SS	MS	Est. Var.	%	F-S	tatistics	р	Geographic region	ns (West-East-South)
Among Populations	15	96.980	6.465	0.052	2%	F _{IT}	-0.048	1.000	$F_{TT} = -0.027$	1.000
Among Individuals	560	1565.630	2.796	0.000	0%	F_{ST}	0.017	0.000	$F_{ST} = 0.009$	0.001
Within Individuals	576	1772.000	3.076	3.076	98%	F _{IS}	-0.030	1.000	$F_{IS} = -0.036$	1.000
Total	1151	3434.609		3.128	100%					



PCA2 17.82%

Figure II.4 Principal Component analysis of the genetic relationship between sampling locations







Figure II.5 Cont.



Figure II.5 Cont.

Bayesian population clustering and spatial distribution

The Bayesian approach of genetic structure analysis using an admixture model grouped the individuals into four clusters based in their allele frequencies and F_{ST} pairwise comparisons by individual and locations. (Figure II.6). The mean probability reached a plateau after four clusters, representing four ancestral populations that could explain the observed allele frequency distribution. The validation approach of probability comparisons described by Evanno et al. (2005), K= 4 showed the highest number of differences in the ΔK value (Appendix Figure II.9).



Figure II.6 Bayesian cluster analysis showing the proportion of ancestry in each individual from four genetically distinct populations (K = 4), that can explain the observed distribution of allele frequencies.

Mutation-drift equilibrium

There was no evidence of genetic bottlenecks under the TPM and SMM evolutionary models (Table II.6). However, under the IAM evolutionary model, Waikato, Bay of Islands, Hokianga, Napier, Kaipara Harbour and Pouto samples showed a significant excess of heterozygotes. The distributions of allele frequencies were L-shaped, the expected shape for a population in mutation-drift equilibrium except for the South Island which had a shift mode allele frequency distribution probably due to the low sample size (N = 5; Table II.3). The Garza-Williamson index shows M values that are lower than the critical value of a genetic bottleneck (M < 0.68 (Garza and Williamson 2001); Table II.6), suggesting recent bottlenecks among the sampling locations.

Table II.6 Mutation-drift equilibrium analysis under IAM (Infinite allele model), SMM (Step-wise mutation model), TPM (Two-phase mutation model); mode-shift of allele frequencies and Garza-Williamson M-ratio for each sampling location and locus. Red p values are significant and red M values are below the 0.68 for populations with no effects of recent bottleneck according to Garza-Williamson, 2001.

	W	licoxon's tes	st		Ga	rza-William	son
Sampling location	IAM	TPM	SMM	Allele frequency distribution	θ	N_{ef}	М
Napier	0.011719	0.8125	0.99219	normal L-shaped	3.74293	93.57325	0.29768
Waiotahe	0.234375	0.710938	0.96094	normal L-shaped	3.31191	82.79775	0.28824
Whakatane	0.710938	0.945313	0.97266	normal L-shaped	3.13792	78.448	0.20039
Hauraki Gulf	0.34375	0.972656	0.99219	normal L-shaped	2.50833	62.70825	0.28339
Bay of Islands	0.039063	0.65625	1	normal L-shaped	3.14289	78.57225	0.27564
Waimimihia	0.148438	0.34375	0.97266	normal L-shaped	3.77688	94.422	0.31654
Hokianga	0.039063	0.46875	0.98828	normal L-shaped	3.33438	83.3595	0.26871
Dargaville	0.34375	0.765625	0.99609	normal L-shaped	3.21506	80.3765	0.28032
Pouto	0.039063	0.40625	0.94531	normal L-shaped	3.43512	85.878	0.26497
Kaipara Harbour	0.011719	0.59375	0.99609	normal L-shaped	3.10379	77.59475	0.30728
Waikato	0.011719	0.34375	0.85156	normal L-shaped	3.4182	85.455	0.27193
Whangape	0.148438	0.40625	0.99219	normal L-shaped	3.63713	90.92825	0.2955
Patea	0.148438	0.46875	0.97266	normal L-shaped	3.84019	96.00475	0.26959
Porirua	0.34375	0.8125	0.98828	normal L-shaped	3.75179	93.79475	0.24082
Wairarapa	0.53125	0.992188	1	normal L-shaped	2.42738	60.6845	0.27409
South Island	0.1875	0.148438	0.14844	Shifted Mode	3.07284	76.821	0.22744

Migration and contemporary connectivity patterns

The assignment analysis under the Rannala and Mountain (1997) algorithm, showed only 134 individuals out of 576 correctly assigned to the location of origin (23.3%). Of those, fifteen individuals from ten locations could be considered first generation migrants (F_0) with a probability below 0.01 (Appendix Table II.9). The Bayesian analysis of migration rates in Migrate-n showed Wairarapa and Hauraki Gulf as the locations with more immigrants (Appendix Table II.10; Figure II.7). The rest sampling locations showed a low proportion of migrants (Appendix Table II.10; Figure II.7).



Figure II.7 Connectivity patterns of <u>M. cephalus</u> in New Zealand based on the proportion of migrants and the proportion of non-migrants by sampling location. Only the flows that are representing a migration volume of more than 15% are shown. The colour pattern follows regional distribution: Blue = South Island and southern North Island; red/orange = west coast North Island. The outer circle segments below the locality name represent the total number of migrants moving to and from the locality. The inner bars show the total volume of migration encoded in the circle segments. To define directionality, in each location emigrants are represented with the same colour of the location and immigrants are coloured as the locality of origin. The width of the curved line is the volume of movements where the exact volume of movements can be visualized only at the beginning and end points. The width of the curved line depends on number of migrants relative to the total volume of migrants across all regions indicated by the outside circle segments.

Discussion

The msatDNA markers used in this study were highly polymorphic and showed enough statistical power to assess the genetic differences found among populations. The msatDNA analysis showed that the distribution of allele frequencies of *M. cephalus* throughout the North Island, New Zealand could be explained by the recent admixture of four genetically distinct ancestral populations that overlap geographically. There was no evidence of spatial correlation with the levels of genetic differentiation among the four clusters, neither Isolation-By-Distance. High levels of gene flow among locations were identified and the four clusters showed a spatial chaotic genetic structure and *He* excess characteristic of recent bottlenecks.

Linkage disequilibrium

There was linkage disequilibrium (*LD*) between some of the msatDNA loci but it was not consistent among all population comparisons. This pattern of *LD* could be the result of low effective population sizes and genetic structuring of *M. cephalus* populations. Alternatively, it could be a consequence of secondary contact and recent admixture of genetically distinct populations (Lawrence et al. 2005; Sedghifar et al. 2015; Skelly et al. 2016).

There are three main reasons why *LD* could be seen in nuclear DNA markers or genes that are not physically proximate. First, the interaction of genes, or epistasis, where one locus is altered or masked by the effects in another locus (Cordell 2002). This pattern will show a consistent linkage between two markers situated in different chromosomes such as the case of the human chromosomes from Mexican-Americans in USA, where the interaction of genes in chromosome 2 and 15 increases their susceptibility to acquire type 2 diabetes (Cox et al. 1999; Cordell 2002). For *M. cephalus* in New Zealand, there was not a consistent linkage pattern across the individuals assessed.

A second possibility is that two loci show *LD* due to the effects of recent bottlenecks (Lawrence et al. 2005; Slatkin 2008). In random mating populations under population expansion and with a large population size, all possible allelic configurations are assumed to

occur at random (Lawrence et al. 2005; Skelly et al. 2016). In those populations, the probability of finding allele configurations that show linked loci when they are not physically proximate, is typically low (Slatkin and Excoffier 1996). The probability of finding linked alleles that are not physically proximate increases when the recent bottlenecks lead to a temporary excess of He in comparison to the expected levels of He (e.g. Marth et al. 2004; Lawrence et al. 2005; Achaz 2009; Huang et al. 2014; Skelly et al. 2016). Species that show an excess of heterozygotes in their populations will show linkage between pairs of loci in one population that may not be present in other populations (Lawrence et al. 2005; Slatkin 2008; Skelly et al. 2016). In New Zealand, the effects of recent bottlenecks were found in all sampled locations and significant He excess was detected in only six of them (Waikato, Bay of Islands, Hokianga, Napier, Kaipara Harbour and Pouto). The reduction in effective population size and the subsequent effect of genetic drift may account for the presence of random associations between pairs of loci. Mugil cephalus showed ancestry of four ancestral and genetically distinct populations with overlapping geographic distribution, which could be a result of sweepstakereproductive-success, in which bottlenecks in juveniles due to seasonal variability in their mortality rates, causes different reproductive success of the adult and a patchy spatial distribution of the allele frequencies (see below) (Hedgecock 1994; Hedgecock and Pudovkin 2011).

The bottleneck effects found in this study could also be an artefact of low number of loci used. Low number of loci and skewed sample sizes could fail to detect changes in the effective population size (bottleneck) and differentiation between populations (Laikre et al. 2005; Ryman and Palm 2006; Jorde and Ryman 2007). The seven loci used in this study showed significant levels of differentiation and they were powerful enough to detect significant differentiation (levels as low as significant F_{ST} = 0.025), even after 40 generations under HW equilibrium. The sample size used for each location allowed to detect high levels of

polymorphisms in each locus. Most of the loci used in this study will require >10 samples to discover a new allele in the population which increases exponentially the sampling effort in order to add little more variation to be assessed in the population analysis.

The third possibility for significant linkage of non-proximate alleles is that allopatric populations came into contact. In secondary contacts, the differentiated genes from the founder populations will be affected at each generation by segregation and recombination of chromosomes, homogenizing the allele frequencies after a few generations (Sedghifar et al. 2015). During admixture, alleles of the same ancestry will be co-inherited leading to a non-random association of alleles showing significant linkage between pairs of loci even when they are not proximate or in the same chromosome (Sedghifar et al. 2015; Skelly et al. 2016).

The analysis of mtDNA markers could help to understand the demographic processes that could be involved in significant *LD* found. The effect of recent bottlenecks or admixture between allopatric populations could result in the non-random association of alleles in pairs of loci that are not physically proximate for *M. cephalus* in New Zealand.

Chaotic population genetic structure

High levels of allele diversity were identified at the loci sampled in this study. Although there was gene flow among the sampled locations and admixture was identified, ancestry of four genetically distinct populations was identified where one population was predominantly found in freshwater locations. The analysis performed in the software STRUCTURE that assigns individuals into ancestral clusters could underestimate the number of contributing populations when there are unbalanced sample sizes (Wang 2017). Specifically, it merges populations with small sample size (Wang 2017). It is possible that the population distributed mainly in freshwater systems, the South Island and Porirua is an artefact of the uneven sample sizes and STRUCTURE could not accurately identify the real number of genetically distinct clusters within those locations. Genetic distances showed no correlation with geographic distance. The hypothesis of IBD was rejected ($R^2 = 0.02$ p>0.05). The general spatial pattern of the allele frequencies did not correspond to a geographic cline or position. The four genetically distinct populations were consistently presented across the analyses used.

Species such as *M. cephalus* in New Zealand that showed genetic heterogeneity but no correlation with geographic distances are considered to have a chaotic genetic structure (Hogan et al. 2010; Broquet et al. 2013; Selwyn et al. 2016). One of the factors that is argued to cause this type of structure in marine species is when there is a significant variance in reproductive success that occurs during larval and juvenile recruitment stages (Selwyn et al. 2016). The allele frequency differences between generations of offspring is determined by differential mortality of larvae and juveniles that occurs between seasons leading to seasonal bottlenecks with different effects of drift (Hedgecock and Pudovkin 2011). In this situation only the offspring from a portion of the reproducing adults are successful in any given season. This phenomenon is called Sweepstake-Reproductive-Success or SRS (Hedgecock 1994; Hedgecock et al. 2007a; Christie et al. 2010; Hedgecock and Pudovkin 2011).

The factors affecting the mortality of larvae and juveniles could be different in each locality and different each season, resulting in neighbouring locations that are genetically different, and also some genetically distinct populations with overlapping geographic distributions (Hedgecock 1994; Hogan et al. 2010; Selkoe et al. 2010; Broquet et al. 2013). The genetic diversity of the whole population is maintained because of the combined contribution all the cohorts make to the adult population, *i.e.* the adults are comprised of overlapping generations (Hedgecock 1994; Hogan et al. 2010; Selkoe et al. 2010; Broquet et al. 2010; Broquet et al. 2013).

Mugil cephalus in New Zealand showed the effect of recent bottlenecks and significant *He* excess was detected in Waikato, Bay of Islands, Hokianga, Napier, Kaipara Harbour and

Pouto. The differences in effective population sizes among New Zealand locations may be the result low sample sizes of the sympatric populations; however, this fails to explain the levels of gene flow and the geographical patchiness of the genetic structure of *M. cephalus*. A most likely explanation is that *M. cephalus* is chaotic genetic structure is an outcome of dissimilar seasonal contribution of genetic diversity from recruited juveniles as predicted by the Sweepstakes-Reproductive-Success hypothesis (SRS) (Hedgecock 1994; Hogan et al. 2010; Selkoe et al. 2010; Broquet et al. 2013). In the SRS hypothesis the fluctuations in rates of survival of larvae and juveniles could lead to seasonal bottlenecks and shifted allele frequencies in which the overall genetic diversity could be maintained by the seasonal contribution of overlapping generation (Hedgecock 1994). The SRS could be tested including juveniles assessing the levels of genetic diversity compared to the adults and testing for effects of bottlenecks in juveniles from different spawning seasons.

One of the consequences of rapid reductions in allele diversity is a *He* excess when compared to the expected heterozygosity (*Ho*) levels (Maruyama and Fuerstt 1985; Cornuet and Luikart 1996; Piry et al. 1999; Galtier et al. 2000). Loci under SMM could look like more suited to a IAM model when strong drift eliminates alleles and the allele frequency distribution appears to be fragmented (Piry et al. 1999). Unfortunately, despite *M. cephalus* populations showing the effects of SRS model in this study, the factors behind the shifts in *Ne* and *He* excess are difficult to identify with any certainty. It could be the result of kinship aggregations of larvae and juveniles that settled in the same location as the adults (Broquet et al. 2013; Iacchei et al. 2016). It could also be the result of reproductive behaviours such as philopatry or a strong correlation with environmental variables (Bentzen and Bradbury 2016; Bonanomi et al. 2016; Selwyn et al. 2016).

Previous fisheries management studies of *M. cephalus* in New Zealand have identified six putative sub-stocks based on catch per unit effort and landings in the GMU1 fishery area

(McKenzie and Vaughan 2008; Figure I.2). The sub-stocks identified were the Lower Waikato River, Manukau Harbour, Kaipara Harbour, North West Coast, East Northland (including Bay of Islands) and Hauraki Gulf (McKenzie and Vaughan 2008). These are some of the locations analysed in this study with strong effects of bottlenecks. Future studies could investigate how the environment and human activities contribute to genetic diversity patterns.

The Kaipara Harbour and the other areas mentioned above are used as nursery areas by *M. cephalus* along with other major estuaries around the North Island (Francis et al. 2005; Morrison et al. 2016). Fish species with nursery areas in environments such as estuaries seem to be sensitive to the effects of SRS (Christie et al. 2010; Hedgecock and Pudovkin 2011; Broquet et al. 2013). Estuaries act as sink environments for larvae and juveniles, and typically hold populations with overlapping generations (Christie et al. 2010; Broquet et al. 2013). The collective contributions from multiple cohorts made to the adult population ensure that the overall level of diversity is maintained (Hedgecock 1994; Hedgecock and Pudovkin 2011). The risk is that if one season's cohort is lost, there could be a gap for the transmission of alleles to the adult population. The more often these gaps occur, the levels of genetic diversity are more likely to decline over time (Palumbi 1994; Hedgecock and Pudovkin 2011; Broquet et al. 2013). Some of the nursery estuarine environments are surrounded by areas of intensive landuse, which increase sediment and nitrogen loads that modifies or degrades the suitable habitat (Larned et al. 2015). There is growing need to assess the impacts of human-based activities and the threat they might pose to species with sensitive nursery grounds and which rely on sustained cohort success to maintain genetic diversity.

Mugil cephalus population genetic structure in the world

Significant levels of genetic differentiation with msatDNA markers have been found in other *M. cephalus* populations in Tunisia and the western Pacific (Liu et al. 2009b; Blel et al. 2010; Shen et al. 2011; Krück et al. 2013). *Mugil cephalus* populations in New Zealand showed similar levels of heterozygosity to their counterparts in Tunisia and higher levels than those

from the China Sea (New Zealand, He = 0.749 - 0.856; Table II.3; Tunisia, He = 0.8 - 0.93; Blel *et al.* 2010; China Sea, NWP1: He = 0.6; NWP2: He = 0.57; NWP3: He = 0.33; Shen *et al.* 2011). The main differences among studies were the factors identified driving the genetic structure in each geographic region.

In Tunisia, the significant level of differentiation was caused mainly by a habitat preference resulting in what appeared to be selection at the locus *Prl*, which encodes the prolactin protein (Blel et al. 2010). The prolactin protein plays a role in the ability to tolerate movements from a saline to freshwater environments and is the most important hormone that controls the acclimation to different salinity levels in fish (Pickford and Phillips 1959; Manzon 2002; Sakamoto and McCormick 2006). In Tunisia, a panmictic population was identified with extensive gene flow among locations, only the allele frequencies of the *Prl* locus in individuals from each location were positively correlated with different salinity levels (Blel et al. 2010).

In the China Sea and west Pacific, the genetic differentiation levels of *M. cephalus* populations determined using mtDNA and msatDNA markers, were correlated with spatial factors (Shen et al. 2011). The study suggested that the finding of genetically distinct populations could be explained as a consequence of Pleistocene glaciation events. Each glaciation during the Plio-Pleistocene changed the sea level and modified the major current patterns in the China Sea. The results of these major climatic changes were periods of isolation among the three *M. cephalus* populations followed by secondary contact as the climate warmed and the sea level rose (Shen et al. 2011). The pattern of genetic structure was one of differentiated populations that overlapped in their geographic distribution, but without any significant gene flow among them.

Mugil cephalus in New Zealand might possible follow similar correlation with environmental variables as the populations found in Tunisia (Blel et al. 2010). Individuals clustered in a population with a predominant ancestry of one of the four populations identified in this study were distributed mainly in freshwater locations including the Patea River in the south west, Whangape Lake off the Waikato River, Waimimihia Lake in the northwest coast, Whakatane and Waiotahe Rivers in Bay of Plenty on the east coast. The variation in the allele frequencies of the locus Prl was not assessed in this study; however, the variation in the allele frequencies in New Zealand freshwater locations and Porirua Harbour seemed to be influenced by the concentration of Escherichia coli which is mainly related to human-based activities (Appendix Figure II.10; measured with distance-based redundancy analysis (dbRDA), generated to visualise the results of distance-based linear models (DistLM), correlating environmental variation matrices with similarity distances of allele frequencies; Jimenez 2017 unpubl. data). High levels of pollution can influence the survival of the organisms leading to a selection scenario with the fixation of alleles. Human activities may also be responsible for the strong effects of bottlenecks in the Waikato River, the Kaipara Harbour, Bay of Islands and the Hauraki Gulf which were some locations identified as heavily impacted by the fishery (McKenzie et al. 1999; Watson et al. 2005; McKenzie and Vaughan 2008; Morrison et al. 2014a,c). A more thorough analysis is required to assess the human impact on genetic diversity and the genetic structure of *M. cephalus* in New Zealand.

Genetically differentiated populations of *M. cephalus* as a consequence of philopatry have been reported in Taiwan (Huang et al. 2001). Two sub-populations share spawning grounds in Taiwan's coastal waters; a nomadic population that travels to the Chinese sea after the spawning season and a resident non-migratory population. These appear to have become two differentiated populations because of the fixation of one allele of the glucose 6-phosphateisomerase-A (GPI-A) locus in the nomadic population. The protein produced by the GPI-A gene is involved in glycolysis, gluconeogenesis and as a neurotrophic factor. In comparison, our study showed gene flow among locations in the east, west and south coasts of New Zealand but there were also high rates of non-migratory individuals, which is a finding other *M*. *cephalus* studies have not reported. It is possible that a combination of different spawning grounds, key reproductive behaviours such as philopatry (natal homing) and season-based differential reproductive success, are isolating *M. cephalus* individuals into groups with similar characteristics.

Marine fish species such as the Atlantic cod Gadus morhua in Iceland and Greenland have shown a chaotic genetic structure as a result of philopatry (Bonanomi et al. 2016). This species has genetically structured populations in neighbouring locations in Iceland, but high levels of gene flow between Iceland and Greenland populations. By analysing msatDNA from stored voucher specimens collected in both locations since the 1960's and the historical records of their migration movements, Greenland populations returning to Iceland to spawn each season were identified (Bonanomi et al. 2016). Sampling of juvenile M. cephalus in New Zealand has shown that juveniles are estuary-dependent (Francis et al. 2011; Morrison et al. 2016). The precise location of spawning grounds in New Zealand are unknown although it is most likely to be in the marine environment which is consistent with observations of M. cephalus in other parts of the world. While juveniles are known to be dependent of low-salinity conditions for successful growth, spawning is strictly marine-based (Whitfield et al. 2012; Górski et al. 2015; Morrison et al. 2016). Otolith-chemistry analysis of the same adults used in this genetics study, and other adults collected in the same locations, showed that there were different migratory and non-migratory groups (Górski et al. 2015; Morrison et al. 2016; Gillanders et al. unpubl. data). Some individuals appear to often move between marine and estuarine/freshwater systems, while others migrate from marine to freshwater environments and did not seem to return to the sea; other M. cephalus individuals remained in the marine environment for all of their lives (Morrison et al. 2016). Some adults spawn at least once in their life and others spawn intermittently without any particular periodicity, which would be typical features of species with SRS (Hedgecock and Pudovkin 2011; Morrison et al. 2016). It is possible that there are multiple spawning grounds in New Zealand and that each population identified in this study has a different spawning ground. The spawning behaviour of *M*. *cephalus* in New Zealand could be the cause for the SRS and the chaotic genetic structure.

Conclusion

Mugil cephalus in New Zealand showed ancestry and recent admixture of four genetically differentiated populations, which overlap in their geographic distribution. Gene flow among locations and a high rate of non-migratory individuals in each location were identified. To date there has been no reported evidence of morphological characters that could affect mating or niche specialization that could have caused the population genetic differentiation levels. The chaotic spatial pattern observed with msatDNA markers in the genetic structure of New Zealand *M. cephalus* may be a consequence of variability in the reproduction success of adults due to their spawning behaviour. This would be typical of a species with a SRS pattern. The chaotic distribution of genetic diversity is also reflected in significant shifts in *Ne* among locations. Comparing levels of genetic variation between adults and larvae or juvenile samples, can help to assess the applicability of the SRS model. Philopatry could be another factor driving the chaotic genetic structure of *M. cephalus* in New Zealand that would explain the patchy geographical distribution of the four populations.

The proportion of genetic variation that is caused by environmental variability and human-based activities still needs to be evaluated. Human activities could also accentuate the bottlenecks and affect the genetic diversity and structure of *M. cephalus* in New Zealand. The six locations in the GMU1 fishery area showed the strongest effects of recent bottlenecks; which matched those areas with high fishery pressure and were identified as putative fishery stocks based on CPUE (McKenzie and Vaughan 2008). Whether the genetic bottlenecks are a consequence of the small effective population sizes of the four genetically distinct populations, the consequence of fishery pressure and human-based activities in estuarine nursery areas for the management GMU1 area, or a combination of both, remains uncertain and further work is needed. This is the first study reporting a chaotic population genetic structure in *M. cephalus* species; however, further studies are required to disentangle the causal factors behind it.

Appendix

Table II.7 Basic statistics summary by sampling location and microsatellite DNA marker. N = number of samples; A = number of different alleles; Af = number of effective alleles; Ar = Shannon's index of allele diversity; Ho = Observed heterozygosity; He = Expected Heterozygosity; $F_{1S} =$ inbreeding coefficient.

Population/Locus		Mce3	Mcs16EM	Mce2	Mcs16DM	Mce27	Mcs2FH	Mce24
	Ν	48	48	48	48	48	48	48
	Α	29	17	9	13	14	15	15
	Af	12.872	12.625	4.122	7.529	5.284	7.211	6.330
Napier	Ar	2.920	2.649	1.734	2.225	1.938	2.222	2.214
-	H_O	0.875	1.000	0.771	0.917	0.646	0.917	0.854
	He	0.922	0.921	0.757	0.867	0.811	0.861	0.842
	F _{IS}	0.051	-0.086	-0.018	-0.057	0.203	-0.064	-0.014
	N	12	12	12	12	12	12	12
	Α	16	8	9	8	7	6	9
	Af	10.667	6.545	5.236	5.878	3.200	4.966	4.721
Waiotahe	Ar	2.590	1.955	1.885	1.910	1.485	1.676	1.846
	H_O	1.000	0.667	1.000	1.000	0.917	0.917	0.833
	He	0.906	0.847	0.809	0.830	0.688	0.799	0.788
	F _{IS}	-0.103	0.213	-0.236	-0.205	-0.333	-0.148	-0.057
	Ν	6	6	6	6	6	6	6
	Α	7	10	8	6	3	7	8
	Af	5.143	8.000	6.545	3.273	1.946	6.000	5.538
Whakatane	Ar	1.792	2.210	1.979	1.474	0.824	1.864	1.907
	H_O	0.833	0.833	0.833	0.667	0.500	1.000	0.833
	He	0.806	0.875	0.847	0.694	0.486	0.833	0.819
	F _{IS}	-0.034	0.048	0.016	0.040	-0.029	-0.200	-0.017
	N	48	48	48	48	48	48	48
	Α	19	15	8	12	6	10	14
	Af	6.312	5.499	3.752	7.303	3.678	3.026	5.402
Hauraki Gulf	Ar	2.277	2.149	1.576	2.169	1.456	1.411	2.100
	H_O	1.000	1.000	0.958	1.000	0.813	1.000	1.000
	He	0.842	0.818	0.734	0.863	0.728	0.669	0.815
	F _{IS}	-0.188	-0.222	-0.307	-0.159	-0.116	-0.494	-0.227
Bay of Islands	Ν	48	48	48	48	48	48	48
	Α	19	17	11	14	9	15	16
	Af	5.519	10.642	3.303	8.565	3.526	8.727	9.423
	Ar	2.286	2.540	1.616	2.339	1.550	2.369	2.482
	H_O	0.667	0.875	0.667	0.896	0.813	0.938	0.938
	He	0.819	0.906	0.697	0.883	0.716	0.885	0.894
	F _{IS}	0.186	0.034	0.044	-0.014	-0.134	-0.059	-0.049
	Ν	48	48	48	48	48	48	48
	Α	23	15	11	14	9	16	18
	Af	6.630	10.691	5.633	9.846	3.666	9.179	9.600
Waimimihia	Ar	2.485	2.492	1.970	2.428	1.620	2.441	2.561
	H_O	0.875	0.938	0.958	1.000	0.875	1.000	0.896
	He	0.849	0.906	0.822	0.898	0.727	0.891	0.896
	F _{IS}	-0.030	-0.034	-0.165	-0.113	-0.203	-0.122	0.000
	N	48	48	48	48	48	48	48
	Α	23	19	10	13	7	19	14
Hokianga	Af	7.629	12.800	3.749	8.214	3.368	9.216	7.385
	Ar	2.543	2.694	1.717	2.298	1.491	2.489	2.259
	H_O	0.750	0.896	0.771	0.750	0.771	0.896	0.917

Population/Locus		Mce3	Mcs16EM	Mce2	Mcs16DM	Mce27	Mcs2FH	Mce24
	He	0.869	0.922	0.733	0.878	0.703	0.891	0.865
	F _{IS}	0.137	0.028	-0.051	0.146	-0.096	-0.005	-0.060
	N	48	48	48	48	48	48	48
	Α	29	18	12	12	8	16	14
	Af	8.199	11.434	3.303	7.178	4.295	7.986	5.760
Dargaville	Ar	2.720	2.603	1.615	2.181	1.673	2.296	2.159
	H_O	0.792	0.938	0.813	0.979	0.854	0.958	0.958
	He	0.878	0.913	0.697	0.861	0.767	0.875	0.826
	F _{IS}	0.098	-0.027	-0.165	-0.138	-0.113	-0.096	-0.160
	Ν	38	38	38	38	38	38	38
	Α	16	15	10	12	8	14	14
	Af	5.213	11.194	5.776	8.997	3.698	7.640	6.361
Pouto	Ar	2.090	2.534	1.930	2.306	1.529	2.264	2.143
	H_O	0.632	0.921	0.789	0.974	0.684	0.895	0.842
	He	0.808	0.911	0.827	0.889	0.730	0.869	0.843
	F _{IS}	0.219	-0.011	0.045	-0.095	0.062	-0.029	0.001
	Ν	48	48	48	48	48	48	48
	Α	25	16	8	13	7	15	14
	Af	9.501	8.170	3.932	7.291	2.801	7.257	8.930
Kaipara Harbour	Ar	2.666	2.369	1.663	2.207	1.302	2.246	2.366
	H_O	0.792	0.958	0.896	0.854	0.604	0.938	0.896
	He	0.895	0.878	0.746	0.863	0.643	0.862	0.888
	F _{IS}	0.115	-0.092	-0.201	0.010	0.060	-0.087	-0.009
	Ν	48	48	48	48	48	48	48
	Α	21	18	9	14	6	19	15
	Af	7.100	12.420	3.418	9.462	3.640	10.195	8.711
Waikato	Ar	2.469	2.663	1.560	2.417	1.491	2.559	2.391
Waikato	H_O	0.771	0.958	0.750	0.938	0.667	0.958	0.979
	He	0.859	0.919	0.707	0.894	0.725	0.902	0.885
	F _{IS}	0.103	-0.042	-0.060	-0.048	0.081	-0.063	-0.106
	N	47	47	47	47	47	47	47
	Α	25	18	11	13	9	14	15
	Af	7.670	10.909	4.942	8.513	4.277	8.004	6.766
Whangape	Ar	2.572	2.600	1.922	2.309	1.720	2.255	2.262
Whangape	H_O	0.915	0.957	0.809	0.872	0.979	0.936	0.979
	He	0.870	0.908	0.798	0.883	0.766	0.875	0.852
	F _{IS}	-0.052	-0.054	-0.014	0.012	-0.277	-0.070	-0.148
	Ν	24	24	24	24	24	24	24
	Α	18	12	9	11	7	15	16
	Af	6.583	7.024	5.703	7.385	4.535	9.071	8.288
Patea	Ar	2.370	2.152	1.940	2.170	1.688	2.396	2.412
	H_O	0.875	0.750	0.875	0.875	0.958	1.000	1.000
	He	0.848	0.858	0.825	0.865	0.780	0.890	0.879
	F _{IS}	-0.032	0.126	-0.061	-0.012	-0.229	-0.124	-0.137
	Ν	12	12	12	12	12	12	12
	Α	13	11	9	9	6	10	12
	Af	6.261	7.200	5.434	4.364	4.645	7.024	7.579
Porirua	Ar	2.210	2.167	1.921	1.807	1.632	2.109	2.246
	H_O	0.833	0.500	0.750	0.833	0.917	1.000	1.000
	He	0.840	0.861	0.816	0.771	0.785	0.858	0.868
	F _{IS}	0.008	0.419	0.081	-0.081	-0.168	-0.166	-0.152
Wairarapa	N	48	48	48	48	48	48	48

Population/Locus		Mce3	Mcs16EM	Mce2	Mcs16DM	Mce27	Mcs2FH	Mce24
	Α	13	13	11	14	6	13	16
	Af	4.785	5.585	3.193	6.867	2.894	4.760	5.415
	Ar	1.949	2.101	1.606	2.188	1.336	1.959	2.123
	H_O	0.979	1.000	0.813	0.979	0.646	1.000	1.000
	Не	0.791	0.821	0.687	0.854	0.655	0.790	0.815
	F _{IS}	-0.238	-0.218	-0.183	-0.146	0.013	-0.266	-0.227
	Ν	5	5	5	5	5	5	5
	Α	5	8	5	5	4	6	7
	Af	3.846	7.143	4.167	4.167	1.923	5.000	6.250
South Island	Ar	1.471	2.025	1.505	1.505	0.940	1.696	1.887
	H_O	0.800	1.000	0.600	0.600	0.600	1.000	1.000
	He	0.740	0.860	0.760	0.760	0.480	0.800	0.840
	F _{IS}	-0.081	-0.163	0.211	0.211	-0.250	-0.250	-0.190

Table II.8 Significant linked pairs of loci by sampling location after Bonferroni correction for uneven sample sizes. The sampling locations that are not shown did not show any linked pair of loci.

Population	Linked loci	LD(p)
Napier	Mce27/Mcs2FH	0.00505
Whatstone	Mcs16DM/Mce27	0.02188
wnakatane	Mce2/Mce24	0.00129
	Mce3/Mcs2FH	0.00485
	Mcs16EM/Mcs2FH	0.00752
Hauraki Gulf	Mce2/Mcs2FH	0.00356
	Mce3/Mce24	0.00050
	Mcs2FH/Mce24	0.00010
Bay of Islands	Mce3/Mcs2FH	0.02287
Waimimihia	Mcs16EM/Mce24	0.04436
	Mce2/Mcs 16 DM	0.02208
Hokianga	Mcs16EM/Mcs2FH	0.02188
	Mce27/Mce24	0.02337
Dorgovillo	Mce27/Mce24	0.01158
Dargavine	Mcs2FH/Mce24	0.00356
Kaipara Harbour	Mce3/Mcs16DM	0.00574
Waikato	Mce3/Mcs16DM	0.03614
Whangape	Mce27/Mce24	0.04030
Dataa	Mcs16DM/Mcs2FH	0.02297
1 atca	Mce27/Mcs2FH	0.01030
Porirua	Mce3/Mcs16EM	0.04119
	Mce3/Mcs16EM	0.01050
	Mcs16DM/Mce27	0.00921
Wairarapa	Mce3/Mcs2FH	0.01327
	Mce2/Mcs2FH	0.00832
	Mcs2FH/Mce24	0.04723

Table II.9 Significant probabilities of the first generation migrants. The assignment analysis was calculated based on the expected frequency of each individual's genotype and subsequently assigned to the population where the expected frequency was highest. The highest likelihood and p-values as first generation migrants between locations were performed using the Rannala and Mountain Bayesian method. Likelihood estimations were calculated with L_home. The probability that each individual is not a migrant but a resident was estimated using 1,000,000 simulations (Monte Carlo resampling) under Paetkau's algorithm (Paetkau et al. 2004).

1st generation immigrants	=-Log(L_home/L_max)	р	Receiving
			populations
WAIK38	2.382	0.009	Bay of Islands
BOI42	3.546	0.003	Waiotahe
BOI43	3.275	0.005	Dargaville
BOI48	2.999	0.009	Dargaville
HOKI06	3.839	0.001	Whakatane
NAPI34	3.243	0.004	South Island
NAPI36	3.593	0.002	Pouto
NAPI37	3.095	0.006	Pouto
KAI44	3.156	0.005	South Island
WAIR48	2.858	0.005	South Island
WHAK2	6.979	0.007	Bay of Islands
PORI05	4.044	0.007	Bay of Islands
WHAN26	3.238	0.006	Waikato
WAIMI04	3.213	0.006	South Island
WAIMI10	3.155	0.007	Porirua

	Waikato	Bay of Islands	Hokianga	Napier	Kaipara Harbour	Dargaville	Wairarapa	Pouto	South Island	Whakatane	Patea	Porirua	Waiotahi	Whangape	Waimihia	Hauraki Gulf
Waikato	0.6754	0.2229	0.0053	0.006	0.0053	0.009	0.0188	0.0054	0.0053	0.0052	0.0052	0.0052	0.0052	0.0083	0.0097	0.0076
Bay of Islands	0.0313	0.768	0.0053	0.0063	0.0056	0.0091	0.086	0.0057	0.0052	0.0052	0.0052	0.0052	0.0052	0.008	0.0077	0.0407
Hokianga	0.0068	0.0448	0.6719	0.0068	0.0054	0.0059	0.1855	0.0062	0.0052	0.0052	0.0052	0.0052	0.0052	0.0054	0.0059	0.0293
Napier	0.0104	0.0071	0.0054	0.6742	0.0053	0.0102	0.2168	0.0096	0.0052	0.0052	0.0053	0.0052	0.0052	0.007	0.0066	0.0212
Kaipara Harbour	0.0079	0.0218	0.0054	0.0056	0.6719	0.007	0.1355	0.0054	0.0052	0.0052	0.0052	0.0052	0.0052	0.0056	0.0063	0.1016
Dargaville	0.0082	0.0127	0.0053	0.0103	0.0053	0.6727	0.1736	0.0069	0.0052	0.0052	0.0052	0.0052	0.0052	0.0067	0.0062	0.066
Wairarapa	0.0089	0.007	0.0053	0.0057	0.0053	0.0054	0.6775	0.0054	0.0052	0.0052	0.0052	0.0052	0.0052	0.0077	0.0059	0.2372
Pouto	0.008	0.0109	0.0063	0.0063	0.0063	0.0066	0.2191	0.6734	0.0062	0.0062	0.0062	0.0062	0.0062	0.009	0.0073	0.0159
South Island	0.0192	0.0256	0.0159	0.0161	0.0159	0.0164	0.0778	0.0161	0.6825	0.0159	0.0159	0.0159	0.0159	0.016	0.0167	0.0183
Whakatane	0.0156	0.029	0.0152	0.0152	0.0152	0.0153	0.0689	0.0158	0.0152	0.6818	0.0152	0.0151	0.0152	0.017	0.029	0.0215
Patea	0.009	0.0168	0.0083	1600.0	0.0084	0.0085	0.15	0.0091	0.0083	0.0083	0.675	0.0083	0.0083	0.0088	0.0541	0.0096
Porirua	0.0173	0.0132	0.012	0.0122	0.012	0.0146	0.0562	0.0123	0.0119	0.0119	0.0119	0.6786	0.0119	0.0127	0.0917	0.0195
Waiotahi	0.0128	0.0138	0.0119	0.0119	0.012	0.012	0.1159	0.0124	0.0119	0.0119	0.0119	0.0119	0.6786	0.0121	0.0409	0.018
Whangape	0.007	0.0351	0.0054	0.0061	0.0054	0.006	0.1733	0.0056	0.0053	0.0053	0.0053	0.0053	0.0053	0.6723	0.0241	0.0333
Waimihia	0.0115	0.0455	0.0053	0.0063	0.0053	0.0065	0.1701	0.0055	0.0052	0.0052	0.0052	0.0052	0.0052	0.0138	0.6768	0.0274
Hauraki Gulf	0.0084	0.0136	0.0053	0.0061	0.0053	0.0058	0.0737	0.0054	0.0052	0.0052	0.0052	0.0052	0.0052	0.0064	0.0077	0.8362

Table 0.3 Migration rates based on Bayesian inference approach using a Metropolis-Hastings algorithm to sample multi-dimensional distributions of Theta (θ) with as default in Migrate-n (Beerli and Felsenstein 1999, 2001).

Table II.11 Results of POWSIM 4.1 (Ryman and Palm 2006) to show the statistical power of the genetic markers used in this study. The simulation was conducted for 40 generations adrift for 10, 000 runs or replicates.

Proportion of significance (p < 0.05) for

summed/combined test statistics

Loci	Alleles	runs	# Chi2	Chi2	# Fisher	Fisher
7	184	10000	10000	1	9999	0.999
Total	184	10000	10000	1	9999	0.999

Average F'_{*ST*} over al 10, 000 runs = 0.0025

Expected $F_{ST} = 0.0025$

F'sr calculated from average Hs and HT = 0.0025



Figure II.8 Distribution of loci that could be under positive or balancing selection based on the F_{ST} and He values. Simulation were ran under IAM and SMM evolutionary models with a p<0.01.



Figure II.9 Number of genetic clusters (K) of best fit for the seven microsatellite DNA loci under ΔK (Evanno et al. 2005); and the log probability data approaches. a) Delta K = mean (|L''(K)|) / sd (L(K)); b) rate of change of the likelihood distribution (mean); c) absolute value of the second order rate change of the likelihood distribution (mean); d) L(K) (mean \pm SD).



Figure II.10 Distance-based redundancy analysis (dbRDA), generated to visualise the results of distance-based linear model (DistLM), correlating environmental variation matrices with similarity distances of allele frequencies. The allele frequencies used were from individuals clustered in the population distributed in freshwater systems and Porirua. Similarity matrices were built using Bray-Curtis resemblance matrices of allele frequencies. The p-values for individual predictor variables were obtained using 9,999 permutations.

Mugil cephalus reproductive success and temporal fractioned genetic diversity between adults and juveniles in New Zealand

Abstract

Four genetically distinct populations of Mugil cephalus have been identified in New Zealand, which showed a chaotic genetic structure. The aim of this study was to test whether the genetic variation in those four populations is a result of differences in the reproductive success of adults following the SRS hypothesis. Seven microsatellite DNA markers were amplified from 210 adults collected in 2014 and 274 juveniles collected in 2010, sampled in five estuaries in the North Island, New Zealand. The data collected from the juveniles was grouped in four age-classes based on their size and average growth rate during summer. The levels of genetic diversity within the groups of juveniles were compared to the adult population by calculating the differences in allele frequencies, effective population size (Ne), number of breeders (Nb) and He-excess. Significant differences between the Ne of juveniles and adults were detected, which was consistent with the SRS hypothesis. Different spawning groups were identified, an early spawning group and a late group of spawners. It is also possible that the late spawning group might undergo multiple spawning events during summer. Approximately 40% of the genetic variation between adults and juveniles could be explained by the SRS. The observed *He* excess could not be explained by the seasonal variation in *Ne* between juveniles and adults in which environmental variables and human-based activities could play an important role.

Key words: chaotic genetic structure, temporal genetic variability, single-sample and temporal effective population size methods, effective population size, *Mugil cephalus*

Introduction

Traditionally, marine populations have been viewed as demographically open with individuals moving freely between locations. In the locations that are connected by the migratory movements of some individuals gene flow occurs when there is random mating between migrants and residents at a site, if not, the population would remain genetically heterogeneous (Palumbi 1994; Cowen 2006; Hedgecock and Pudovkin 2011). The maintenance of genetic heterogeneity across the geographic distribution of species could be caused by a variety of processes, including the availability of suitable habitat, the direction and routes of migration, whether there are mating opportunities for migrants and the fitness of offspring (Flowers et al. 2002; Waters et al. 2013b). The natural variability of environmental or climatic conditions can cause changes in the reproductive success of adults and in the offspring survival rates during dispersal and settlement (see Hedgecock 1994; Parsons 1996; Grant and Bowen 1998; Ball et al. 2000). One of the consequences of the variability in reproductive success is the development of a population genetic structure that is spatially patchy, or seemingly chaotic (Hogan et al. 2010; Broquet et al. 2013; Selwyn et al. 2016).

The unseasonal fluctuations in environmental conditions can disrupt breeding success and juvenile recruitment. These can have significant impacts on the levels of genetic variability seen in juveniles compared to that in adults (Hedgecock 1994). If the offspring from one season have limited ability to tolerate or adjust to environmental changes, the increased mortality could cause a strong genetic bottlenecks. Larvae and juveniles that survive the environmental pressures will only be a subset of the total genetic variation found in the adult population (Hedgecock 1994; Flowers et al. 2002; Hedgecock et al. 2007a; Hedgecock and Pudovkin 2011). The genetic variation in the adults will be maintained by the cumulative contribution from subsets of offspring over successive generations (Hedgecock 1994). The difference in levels of genetic diversity between adults and offspring cohorts, is strongly influenced by the dispersal and settlement success of juveniles and what proportion of the genetic diversity each cohort represents (Hedgecock and Pudovkin 2011).

The observation of lower levels of genetic diversity within a cohort of offspring compared to the adult population, forms the basis of the Sweepstakes-Reproductive-Success (SRS) hypothesis (Hedgecock 1994). The hypothesis suggests that allele frequency differences observed among juvenile cohorts are caused by extreme variation in the success of offspring produced by parents. The difference often occurs by chance. The high levels of relatedness among the success offspring that form a cohort only amounts to a partial representation of the variation in the adult population (Flowers et al. 2002).

Siblings experience similar fates because they share the same time and location at conception. Time and location at conception could be two crucial factors that determine the likelihood of juvenile survival (Christie et al. 2010; Hedgecock and Pudovkin 2011; Broquet et al. 2013). If a cohort of offspring completely fails (does not survive) it could impact the ability of a population to maintain its genetic variation (Hedgecock 1994). For example, species that have nursery grounds such as estuaries, might be particularly vulnerable to the loss of a year-class of recruits. High nutrient loading, pollution and increased sedimentation from intensified land use often have a detrimental impact on estuarine environments (Hurst et al. 2009; Elliott and Whitfield 2011; Francis et al. 2011). These types of stressors on estuaries might increase the levels of offspring mortality or reduce the success rates of adult reproduction. This has been documented in areas such as Chesapeake Bay, USA (see Richards and Rago 1999; Kemp et al. 2005; Buchheister et al. 2013).

Estuaries in New Zealand play an important role in the reproductive cycle of fish species such as the kānae/grey mullet *Mugil cephalus*, the snapper *Pagrus auratus* and the kahawai *Arripis trutta*, by providing food sources, refuge from predators, and protection from high-energy coastal waters (Francis et al. 2005). Populations of *M. cephalus* in New Zealand

76

were found to have a spatial chaotic genetic structure in their distribution that could be dependent of seasonal reproductive success of the adults (Chapter II). The nursery grounds for *M. cephalus* are estuaries with extensive, shallow and open tidal mud flats in northern New Zealand (Morrison et al. 2016). Otolith chemistry analysis of *M. cephalus* fish from harbours of the GMU1 (Figure I.2) showed that it is likely *M. cephalus* has a protracted spawning season (Morrison et al. 2016).

There was evidence of bottlenecks in the adults from the main estuaries used by *M. cephalus* in GMU1 (Chapter II). Waikato, Kaipara Harbour, Hokianga, Bay of Islands, Hauraki Gulf and Bay of Plenty showed significant bottlenecks and have been previously thought as possible distinct stocks (Figure I.2) based on CPUE assessments (McKenzie and Vaughan 2008). The bottlenecks of *M. cephalus* in the GMU1 harbours and the use of estuaries as nursey areas, both suggest that *M. cephalus* might be a species with a seasonal sweepstakes-like effect. The aim of this study was to test for the predicted SRS of genetic diversity in *M. cephalus* by comparing the levels of diversity between juvenile cohorts and the adult population. Understanding the spatial and temporal patterns of genetic variation of *M. cephalus* is important for the long-term management of their genetic variation and the evolutionary potential of its populations.

Methods

Data collection

Seven microsatellite DNA (msatDNA) loci were amplified from 274 juvenile samples collected in five major estuaries in the North Island, Kaipara Harbour, Hokianga, Bay of Islands, Hauraki Gulf and Bay of Plenty, during summer 2010; and, 210 adult samples collected in the same locations during summer 2014. The DNA was extracted from tissue biopsy from the caudal peduncle following the phenol/chloroform extraction protocol and DNA was stored in TE buffer at -4°C. In order to make any comparisons between juveniles and adults, the

juveniles were grouped in age classes. The age of the juveniles was estimated with an average hatch length of 2.6 mm, an average monthly growth rate of 17 mm (for summer conditions) and the length of each individual as shown in equation 1 (Whitfield et al. 2012).

$$A = \left\{ \left(\frac{L - 2.6mm}{17mm}\right) 30 days \right\} - 30$$

Equation III.1 Age class delimitation of <u>M. cephalus</u> juveniles. A = Age class; L = length of the fish when caught; 2.6mm = average hatching length for <u>M. cephalus</u>; 17mm = average growth rate of <u>M. cephalus</u> during summer conditions; 30 days refers to the total growth in a 30-day period; -30 refers to the average drifting time of larvae before settlement in nursery areas.

Where A is the age class, L is the length of the juvenile fish when caught, 2.6 is the hatching length (mm) and 17 (mm) the average monthly growth rate (Whitfield et al. 2012). The time larvae expend drifting in coastal currents was subtracted from the equation. The juveniles were grouped into four age classes: 30-60 days, 90-120 days, 150-180 days, 210-330 days and the adults.

Amplification and genotyping

The amplification of the seven msatDNA loci, PCR primers and PCR conditions were provided in Chapter II (Table II.1). Amplified fragments were sent to Macrogen Inc. for fragment analysis. Resulting chromatograms were visualized in Genemarker v2.2 Demo version (Softgenetics), and the peaks were scored by eye and recorded manually. The alleles were coded as three digit genotypes and the individuals containing missing data at more than two amplified products were removed from the analysis.

Loci assessment and genetic diversity

Genotyping errors were evaluated in Micro-Checker v2.2.3 (Van Oosterhout et al. 2004) separately on juveniles and adults in order to identify the presence of null alleles, large allele drop-out, allele size shifts and scoring errors due to stuttering under the Hardy-Weinberg equilibrium (*HWE*) and expected heterozygosity (*He*). Estimates of number of alleles per locus, observed heterozygosity (*Ho*), *He*, Wright's coefficient of inbreeding (F_{IS}), allelic richness, Shannon's allelic diversity, number of private alleles by locus, sampling location and age class

were calculated in GenAlEx v6.5 (Peakall and Smouse 2012), Arlequin v3.5 (Excoffier and Lischer 2010), and FSTAT v2.9.3.2 (Goudet 1995). Allele frequency distribution plots by locus, sampling location and age class were generated in GenAlEx v6.5. Exact tests for deviation from *HWE* and allele counts were calculated in GENEPOP v4.2 (Rousset 2008); Significant values of *HWE* were calculated using the exact test of Guo and Thompson (1992), with a Markov chain of 10,000 dememorization steps. Bonferroni corrections were used to adjust p-values for multiple comparisons (Rice 1989).

Juvenile and adult genetic structure and differentiation levels

The differences between the allele frequencies of each age class was tested with nonparametric multiple comparisons of non-independent samples with Friedman's ANOVA and Kendall's coefficient of concordance in Dell Statistica. A factorial correspondence analysis was performed in Genetix v4.05 (Belkhir et al. 2004) in order to visualize the genetic distances between samples. Pairwise R_{ST} values (Slatkin 1995), and the levels of genetic differentiation within and among age classes and adults were calculated using an analysis of molecular variance (AMOVA) in Arlequin v3.5 and GenAlEx v6.5. The genetic distance matrix for AMOVA was estimated by pairwise differences in Arlequin v3.5 and co-dominant genetic distances in GenAlEx v6.5. The significance levels of the variance components of R-statistics values were tested by 20,000 nonparametric permutations and 999 permutations in Arlequin v3.5 and GenAlEx v6.5 respectively.

Mutation-drift equilibrium

The excess of heterozygotes for juveniles age classes and adults was calculated in Bottleneck v1.2.02 (Piry et al. 1999), where He was estimated from the number of observed alleles at each locus (k), and a simulated distribution of heterozygosity was generated with the coalescence approach under the three msatDNA evolutionary models: IAM, SMM and the two-Phase Model (TPM). The average expected equilibrium heterozygosity (Heq) for each locus was compared to He and used to assess whether there was a significant heterozygote excess or deficit using a Wilcoxon's test based on 1,000 replications. The test for shifts in the allele frequency distribution from an L-shaped (allele frequency distribution of population that are under mutation-drift equilibrium) was performed in Bottleneck v1.2.02. The effective population sizes calculated from expected heterozygosity and assuming populations that are in mutation-drift equilibrium and the mutation rate were calculated in Arlequin v3.5 for 10,000 iterations. Differences in population size and a population size contraction signature were assessed with the Garza-Williamson analysis as implemented in Arlequin v3.5 and M-RATIO (Garza and Williamson 2001).

Effective population size estimations

An isolation by Time scenario was assessed with a partial Mantel test correlation between the Shannon's allelic diversity indexes, R_{ST} and age class differences in GENEPOP v4.1. Temporal changes in the effective population size (*Ne*) between samples were estimated using a temporal approach to assess the differences in age classes and a single-sample approach for punctual comparisons among juvenile and adults. The temporal variance of F_{ST} was estimated with pairwise comparisons between age classes and generations (adults = generation 0; juveniles = generation 2) following the Jorde and Ryman (2007) method such as implemented in NeEstimator v2 (Do et al. 2014). Three single-sample comparisons were used to assess differences in *Ne* and the number of breeders (*Nb*) in NeEstimator v2 (see below). A method assuming molecular co-ancestry to estimate the possible *Nb* in the adult sample that could explain the variation in each age class based in Nomura algorithm (Nomura 2008). Two methods were used to evaluate the variation in the *Ne*; one based in *He* excess and the second based in linkage disequilibrium (*LD*) as the non-random association of alleles across the seven loci among and within samples (Waples and Do 2008). A kinship reconstruction of *Ne* was performed in COLONY (Jones and Wang 2010).

Relatedness

Mugil cephalus juveniles swim in schools of siblings into estuaries a month after spawning (Whitfield et al. 2012). We tested whether juvenile samples showed a higher level of relatedness to each other compared to a similar number of samples taken from the adult population. Relatedness pairwise comparisons were calculated in GenAlEx v 6.5 under the indices developed by Queller and Goodnight (1989), Ritland (1996) and Lynch and Ritland (1999).

Proportion of non-migrant juveniles

Recruitment ratios and the number of effective migrants between sampling locations were tested in Migrate-n v3.6 (Beerli 2009), using a Metropolis-Hastings algorithm to sample multi-dimensional distributions of Theta (θ) with Maximum likelihood (ML) and with Bayesian inference (BI) approaches (Beerli and Felsenstein 1999, 2001). One hundred seeds, 50, 000,000 MCMC iterations, a thinning of 1,000,000 and a sampling interval of 1,000 were used. Migrate-n calculates the probability that two individuals in a particular generation have a common ancestor (Beerli 2009). The probabilities are calculated integrating all possible relationships in each ancestral generation following the genealogical tree. To estimate the number of immigrants per generation, the total migration rate (M) and effective population size (θ) per generation are scaled to the mutational rate (m) (Beerli 2009). In each node or generation, the effective population size or theta (θ) is calculated. To obtain the value of θm (theta scaled to the mutational rate), theta is multiplied by the mutation rate (m) that is usually inferred from the data set, and multiplied by a constant that depends on the ploidy level and the inheritance of the DNA markers (χ) (Beerli 2009). In bi-parental nuclear markers $\chi = 4$ (two allele copies per parent). The immigration rate in each generation (M) is calculated using θm and multiplying it by the number of migrants and dividing it by the mutation rate (Beerli 2009). The result reflects the number of variants brought to the population each generation that are not a consequence of mutation. Migrate-n integrates all migration rates (M) and effective

population sizes per generation (θm) to estimate the number of migrants since the MRCA (Beerli 2009). The results were represented graphically with the R package circlize (Gu et al. 2014). The width of the curved lines in the circular plot represents the number of migrants and is dependent on the total volume of migrants across all regions.

Results

Loci assessment and genetic diversity

Based on the *He* and the *HWE* allele frequencies, null alleles were identified in all markers at 95% confidence interval in the juveniles and none in the adults (Appendix Table III.9). The frequency of null alleles in each locus was low (average frequency = 0.05 ± 0.03 ; Appendix Table III.9). The null alleles are often the result of scoring mistakes, stuttering and amplification failure, resulting in missing alleles and low levels of heterozygosity (Van Oosterhout et al. 2004). Heterozygote excess was identified in Mcs16EM, Mce2 and Mce24 loci (Table III.1), and the genotyping failure rate was 0.000 for all loci (Appendix Table III.9). Two loci showed significant deviation from *HWE* expectations (loci Mce3 and Mce27; Table III.1). The msatDNA loci were highly polymorphic and showed similar levels of allele diversity among age classes and compared to the adults (Table III.2). Kaipara Harbour showed significant inbreeding in the adults and with the exception of Bay of Plenty, all juvenile sample locations and all juvenile age classes showed significant inbreeding (Table III.2).

Table III.1 Basic Statistics of seven msatDNA loci limited to the adult samples and juveniles in Bay of Islands, Bay of Plenty, Hauraki Gulf, Hokianga and Kaipara Harbour. N_A = number of different alleles; A = proportion of different alleles; Af = number of effective alleles; Ar = Shannon's index of allele diversity; H_0 = Observed heterozygosity; He = Expected heterozygosity; F_{IS} = inbreeding coefficient; F_{IT} = average heterozygosity within the population relative to the total expected heterozygosity; F_{ST} = average heterozygosity between populations relative to the total expected heterozygosity; HWE(p) = probability of deviation from Hardy-Weinberg equilibrium. In red are the significant values at $p \le 0.05$.

Locus	N_A	A	A_{f}	$A_{ m r}$	H_{O}	H_E	F _{IS}	F _{IT}	F _{ST}	HWE (p)
Mce3	91	28	11.602	2.726	0.784	0.904	0.133	0.158	0.030	0.000
Mcs16EM	65	14.4	9.013	2.359	0.914	0.886	-0.032	0.004	0.034	0.082
Mce2	80	10.4	3.968	1.642	0.755	0.739	-0.021	-0.007	0.014	0.633
Mcs16DM	72	13.6	8.139	2.257	0.811	0.872	0.070	0.088	0.019	0.575
Mce27	80	16	7.613	2.203	0.681	0.833	0.183	0.222	0.049	0.000
Mcs2FH	91	18.2	7.230	2.311	0.787	0.860	0.085	0.098	0.014	0.069
Mce24	72	16	8.519	2.354	0.887	0.873	-0.015	0.029	0.043	0.925
-------	----	----	-------	-------	-------	-------	--------	-------	-------	-------
-------	----	----	-------	-------	-------	-------	--------	-------	-------	-------

Table III.2 Basic statistics of sampling locations, development stage and juvenile age classes. N= sample size; A = proportion of different alleles; Af = number of effective alleles; Ar = Shannon's index of allele diversity; H_0 = Observed heterozygosity; He = Expected heterozygosity; F_{IS} = inbreeding coefficient; p = probability of the inbreeding coefficient (F_{IS}). In red are the significant values at $p \le 0.05$.

Development stage	Population	N	A	A_{f}	$A_{ m r}$	H_0	H_E	F _{1S}	Р
	Bay of Plenty	18	14	6.840	2.117	0.848	0.824	0.26328	0.051
	Hauraki Gulf	48	11.14	6.242	2.023	0.865	0.815	-0.14288	0.955
Adults	Bay of Islands	48	12	4.995	1.877	0.967	0.781	0.09472	0.095
	Hokianga	48	14.42	7.100	2.169	0.827	0.828	0.06736	0.188
	Kaipara Harbour	48	15	7.480	2.213	0.821	0.837	0.29945	0.000198
	Bay of Plenty	58	18.14	9.602	2.462	0.800	0.879	0.03706	0.374
	Hauraki Gulf	31	13.14	8.537	2.258	0.799	0.862	0.57414	0
Juveniles	Bay of Islands	51	16.28	8.773	2.36	0.808	0.866	0.33409	0.000099
	Hokianga	47	14.71	6.744	2.187	0.797	0.839	0.48696	0
	Kaipara Harbour	60	15.28	8.183	2.306	0.777	0.866	0.38018	0.000198
	Class 30-60 days	115	19.42	8.529	2.380	0.777	0.867	0.50079	0
Juvenile Age	Class 90-120 days	81	19.71	10.622	2.536	0.801	0.888	0.16464	0.013685
Classes	Class 150-180 days	26	13.57	7.557	2.223	0.810	0.850	0.67926	0
	Class 210-330 days	25	9.28	6.248	1.953	0.759	0.822	0.27693	0.02737
Adults pooled		210	21.28	7.105	2.234	0.865	0.836	0.16231	0.000978
То	457	16.65	8.012	2.265	0.803	0.852	0.05739	0.0001	

Juvenile and adult genetic structure and differentiation levels

The combined distribution of allele frequencies of juvenile age classes and adults in each locus seemed to complement each other covering the whole allelic range found in each location (Figure III.1; Figure III.2). Each juvenile age class contribute with a different subset of alleles (Figure III.1; Figure III.2). The differences in the allele frequencies of each age class and adults were statistically significant after multiple comparisons assuming non-independent samples and different sample sizes (Figure III.1 and Figure III.2; Friedman ANOVA Chi Sqr. (N = 195, df = 9), F-value = 19.25808 p = 0.02309; Kendall's Coefficient Concordance = 0.01097 Average rank r = 0.00588).

The factorial correspondence analysis of the distances of individual allele frequencies showed subdivision among juvenile age classes and between juveniles and adults (Figure III.3).

There was a subdivision between the 210-330 days *M. cephalus* age class and the other juvenile age classes with Axis 1 explaining 47.51% of the total variation (Figure III.3a). There was an overlap in the genetic distances between 30-60 and 90-120 days age classes suggesting they were no significant differences between them (Figure III.3a). The age class 150-180 days showed no overlap in the genetic distances same as the 210-330 days age classes, suggesting distinctive groups. The adults and juveniles were highly differentiated where the Axe 1 explained 48.79% of the total variation (Figure III.3b). Three different groups were identified with the PCA analysis where the 30-60 days and 90-120 days age classes were clustered together (Figure III.4). The 210-330 days age class and the adults were the other two distinct groups (Figure III.4).

The levels of genetic differentiation among locations were different between adults and juveniles (Table III.3). Bay of Islands showed significant levels of genetic differentiation with all sampling locations with the exception of Kaipara Harbour (Table III.3). In the juveniles, Hokianga showed levels of genetic differentiation among sampling locations with the exception of Bay of Islands (Table III.3). In the age class comparisons, significant levels of genetic differentiation were found in the age class of 210-330 days and the adults. The youngest age class (30-60 days), was not significantly different from the adults. The rest of the age classes did not show significant levels of genetic differentiation among them (Table III.3).

		R ST A	Adults		
	Bay of Plenty	Hauraki Gulf	Bay of Islands	Hokianga	Kaipara Harbour
Bay of Plenty	*	0.00644	0.00089	0.02109	0.17553
Hauraki Gulf	0.06735	*	0.00198	0.17325	0.29779
Bay of Islands	0.10069	0.02896	*	0.3267	0.0196
Hokianga	0.04608	0.00543	0.00231	*	0.53005
Kaipara Harbour	0.02049	0.00306	0.02804	-0.001	*
		R _{ST} Ju	uveniles		
	Bay of Plenty	Hauraki Gulf	Bay of Islands	Hokianga	Kaipara Harbour
Bay of Plenty	*	0.53698	0.59925	0.2179	0.25106
Hauraki Gulf	-0.00209	*	0.4457	0.02317	0.73161
Bay of Islands	-0.00473	0.00224	*	0.00772	0.09484
Hokianga	0.01216	0.044	0.05406	*	0.01445
Kaipara Harbour	0.00918	-0.00517	0.02006	0.0435	*
		R _{ST} (Classes		
	Class 30-60 days	Class 90-120 days	Class 150-180 days	Class 210-330 days	Adults
Class 30-60 days	*	0.12356	0.53079	0.00752	0.36198
Class 90-120 days	0.01698	*	0.2296	0.01554	0.0004
Class 150-180 days	-0.00127	0.01053	*	0.00337	0.0297
Class 210-330 days	0.09628	0.04839	0.15312	*	0.00257
Adults	0.00243	0.03584	0.03651	0.07696	*

Table III.3 Levels of genetic differentiation among sampling locations grouped by development stage and levels of genetic differentiation among age classes. Below the midline are the R_{ST} genetic differentiation values. Above the midline are the probability of each R_{ST} value in which significant values are highlighted in red ($p \le 0.05$).

The hierarchical assessment of genetic differences showed that most of the variation is held among individuals, between age classes and within sampling locations. The total genetic differentiation among sampling locations was non-significant (Table III.4).

Table III.4 Analysis of Molecular Variance of differentiation levels within individuals (R_{IS}), among individuals (R_{ST}), among geographic regions (R_{CT} ; within individuals respective to the total variation), and total variation (R_{IT}); grouped by development stage and age classes. Significant probabilities ($p \le 0.05$) are highlighted in red.

	AMO	VA Adults	р	AMO	VA Juveniles	р	AMO	VA Age Classes	р
Within locations	R _{IS}	0.11092	0.0005	R _{IS}	0.38324	0	R _{IS}	0.24134	0
Among locations	R _{ST}	0.01035	0.00396	R _{ST}	0.01832	0	R <i>st</i>	0.03013	0
Within Individuals respective to the total variation	R _{CT}	0.01331	0.04554	R _{CT}	-0.00047	0.05505	R _{CT}	0.00364	0.0048
Total variation	R _{IT}	0.13183	0.2001	R _{IT}	0.39425	0.48149	R _{IT}	0.26688	0.4125







Figure III.1 Cont.



Figure III.1 Cont.



Figure III.1 Cont.







Figure III.3. Factorial correspondence plot of genetic variation between age classes. a) Juvenile age classes comparison; the red dots are the highlighted individuals of that particular age class b) Juvenile age classes compared with the adult genetic variation; adults are in red.



Figure III.4 Principal component analysis between juveniles and adults. Both axes explain 83.34% of the total genetic variation.

Mutation-drift equilibrium

There was evidence of genetic effects of a contraction of population size due to significant changes in mutation-drift equilibrium. Under IAM evolutionary model, the adults showed an excess of heterozygotes in Bay of Islands, Hokianga and Kaipara Harbour. The juveniles showed heterozygote excess in all sampled locations except Hokianga, and in the age classes 90-120 days and 210-330 days. The excess of heterozygotes is a characteristic signal of populations that have gone through recent genetic bottlenecks. In recent bottlenecks, the allele diversity is reduced faster than the heterozygosity where the observed heterozygosity (*Ho*) is larger than the expected heterozygosity (*He*) of a population in mutation-drift equilibrium (Piry et al. 1999). While the reduction in allele diversity could be seen after the bottleneck, the loss of heterozygosity occurs due to the effect of genetic drift each generation after the bottleneck (Maruyama and Fuerstt 1985; Piry et al. 1999).

Under the TPM and SMM evolutionary models there was no evidence of significant changes in mutation-drift equilibrium (Table III.5). The exception was the age class 210-330 days in the age class comparison that showed a significant effect of genetic bottlenecks in the TPM evolutionary model (Table III.5). The distribution of allele frequencies was normal L-shaped without significant shifts in the mode. Populations that are in mutation-drift equilibrium with large *Ne* are expected to have high proportion of low frequency alleles (Piry et al. 1999). Plotting the allele groups against their frequencies the plot resembles an "L" in which the mode shifts if there is allele diversity loss (Piry et al. 1999).

The Garza-Williamson index showed M values that are lower than the critical value (M < 0.68; Table III.5), suggesting a reduction of the effective population size in the sampling location in adults and juveniles, and each age class. The effects of recent genetic bottlenecks were stronger in the age class 210-330 days.

Table III.5. Mutation-drift equilibrium analysis of under infinite allele model (IAM),; two-phase model (TPM) and step wise mutation model (SMM); mode shift of allele frequencies and Garza-Williamson index for allele diversity loss (M). The comparisons were made by sampling location, development stage and age classes. In red are the values that were significant with a $p \leq 0.05$ for the evolutionary model comparisons. The M values that were significant were below the expected 0.68 for populations with in mutations-drift equilibrium.

<u> </u>		W	/ilcoxon's te	est	Allele frequency	Ga	Garza-Williamson			
Stage	Sampling location	IAM	TPM	SMM	usuibuuon	θ	Ne	М		
	Bay of Plenty	0.46875	0.96094	0.96094	normal L-shaped	3.19318	79.8295	0.19334		
	Hauraki Gulf	0.34375	0.97266	0.99219	normal L-shaped	2.50833	62.70825	0.20381		
Adults	Bay of Islands	0.03906	0.65625	1	normal L-shaped	3.14289	78.57225	0.24756		
	Hokianga	0.01172	0.59375	0.99609	normal L-shaped	3.33438	83.3595	0.2367		
	Kaipara Harbour	0.03906	0.53125	0.98828	normal L-shaped	3.10379	77.59475	0.22783		
Adults pooled		0.03906	0.85156	1	normal L-shaped	3.18399	79.59975	0.21693		
	Bay of Plenty	0.00391	0.65625	0.99219	normal L-shaped	4.71859	117.96475	0.25857		
	Hauraki Gulf	0.00781	0.1875	0.46875	normal L-shaped	4.43611	110.90275	0.18673		
Juveniles	Bay of Islands	0.00781	0.65625	0.99219	normal L-shaped	4.16305	104.07625	0.24037		
	Hokianga	0.14844	0.94531	0.99609	normal L-shaped	3.42792	85.698	0.224		
	Kaipara Harbour	0.00391	0.34375	0.98828	normal L-shaped	4.18452	104.613	0.22102		
	Class 30-60 days	0.40625	0.98047	0.99609	normal L-shaped	4.50638	112.6595	0.1814		
Juvenile Age	Class 90-120 days	0.00781	0.53125	0.8125	normal L-shaped	4.94167	123.54175	0.19114		
Classes	Class 150-180 days	0.34375	0.53125	0.99219	normal L-shaped	3.97207	99.30175	0.13281		
	Class 210-330 days	0.00391	0.01953	0.53125	normal L-shaped	3.64285	91.07125	0.09654		
Juveniles pooled		0.00391	0.76563	1	normal L-shaped	4.18604	104.651	0.22164		

Effective population size estimation

The IBT assessment correlating Shannon's allele diversity index, R_{ST} genetic distances and age class differences using a Mantel test was non-significant ($R^2 = 0.0475 \text{ p} = 0.409$). The sample sizes used to assess the single-sample estimates of effective population size and number of effective breeders (*Ne/Nb*), ranged from 25 in the 210-330 days age class to 210 in the adults (Table III.6). The assessment of the *Nb* under the co-ancestry method showed an increase from the age class 30-60 days to 90-120 days (*Nb* = 17.6 and *Nb* = 64.3, respectively; Table III.6). The age class to 150-180 days showed a reduction in the *Nb* to 21.7 and there was a significant increment in the *Nb* in the adults reaching *Nb* = 91.5 (Table III.6). The assessment was not useful to describe the *Nb* that could explain the genetic diversity in the 210-30 days age class that showed *Nb* = infinite (Table III.6). The single-sampled methods to assess *Ne* and *Nb* used in this study, might show negative and infinite values as a result of low sample size (Do et al. 2014). Negative and infinite *Ne* values could usually be interpreted as sampling error where there is no evidence for variation in genetic diversity as a result of a finite number of parents (Do et al. 2014). The single-sampled method based on *He* excess showed all *Ne* as infinite for all age classes for all the critical values with the exception of the adults (Table III.6). There was no evidence of *He* excess and is showed by negative values of the *He* excess index (*D*; Table III.6). The biological meaning is that the parents that could explain the variation in the *He* excess between juveniles and adults were not sampled.

The linkage disequilibrium method used to assess *Ne* showed significant results for the *P* critical value of 0.05. There was a reduction in the *Ne* from the age class 30-60 days to 90-120 days (Ne = 282.5 and Ne = 87.5, respectively; Table III.6), followed by an increase in the 150-80 days age class (Ne = 190; Table III.6), suggesting that a different group of parents could explain the genetic variation found in the 150-180 days age class compared to the age classes 30-60 days and 90-120 days . The adults showed the highest value for *Ne* (Ne = 947.7; Table III.6). The age class 210-30 days under the *LD* method showed negative values of *Ne* (Table III.6). The comparisons of *Ne/Nb* between juveniles and adults by sampling locations, showed significant differences in *Nb* between juveniles and adults with different *P* critical values in the *LD* method (Appendix Table III.10). The method of *He* excess showed infinite and negative values the same as in the age-class analysis and is not shown in the table (Appendix Table III.10).

The temporal estimate of *Ne* based on F-statistics differences among generations, showed that there is a significant inbreeding in the juveniles (generation 2) compared to the adults (generation 0; Table III.7). This pattern was shown independently in each method used

to estimate Ne and calculate F in generation 0 and F' in generation 2 at three different critical

values of *P* (Table III.7).

Table III.6 Single-sample methods to assess the differences in the effective population size (Ne) and Number of breeders (Nb) among juvenile age classes and the adults. N = number of samples; H = Harmonic mean size; F1 = single-sample variation of *F*-statistics; Nb = Number of breeders; CI = Confidence intervals for Ne with non-parametric Jacknife loci; p = critical value; Ai = Number of alleles that are independent; D = index for the excess or deficiency of heterozygotes; Ne = effective population size; inf = infinite.

<i>Ne/Nb</i> estimator	Age class	Ν	H	F <i>1</i>	Nb	(CI			
	30-60 days	40	28	0.0284	17.6	7.8	31.3			
	90-120 days	81	49	0.00777	64.3	0.1	322.8			
Co-ancestry	150-180 days	26	20.5	0.02304	21.7	0.5	80.1			
	210-330 days	25	11.2	-0.00099	Inf	Inf	Inf			
	Adults	210	210	0.00546	91.5	0.1	459.3			
			Р	H	Ai	D	Ne	С	Ι	
	30-60		0.05	28	48	-0.13263	Inf	Inf	Inf	
	davs	40	0.02	28	75	-0.12021	Inf	Inf	Inf	
	days		0.01	28	121	-0.06613	Inf	Inf	Inf	
He excess	90-120 days		0.05	49	49	-0.13095	Inf	Inf	Inf	
		81	0.02	49	88	-0.15033	Inf	Inf	Inf	
			0.01	49	113	-0.11895	Inf	Inf	Inf	
	150-180 days		0.05	20.5	46	-0.09226	Inf	Inf	Inf	
		26	0.02	20.5	69	-0.11493	Inf	Inf	Inf	
	uays		0.01	20.5	88	-0.08354	Inf	Inf	Inf	
	210-330		0.05	11.2	44	-0.07422	Inf	47.6	Inf	
	days	25	0.02	11.2	58	-0.04975	Inf	37.3	Inf	
			0.01	11.2	58	-0.04975	Inf	37.3	Inf	
	Adults	lts 210	0.05	210	43	0.01566	32.4	11.7	Inf	
			0.02	210	71	-0.00165	Inf	23.5	Inf	
			0.01	210	83	0.00073	682.1	23.9	Inf	
			р	H	Ai	\mathbf{r}^2	r ² expected	Ne		CI
	30-60		0.05	22.1	811	0.044944	0.04386	282.5	34.9	Inf
	davs	40	0.02	22.2	2043	0.042607	0.043022	-745.4	49	Inf
	uays		0.01	25.7	4269	0.036438	0.038148	-181.8	33.3	Inf
	90-120		0.05	35.8	1134	0.026909	0.02319	87.5	37.4	Inf
	days	81	0.02	40.4	2621	0.022444	0.019782	123.1	35.2	Inf
	uays		0.01	40.1	4151	0.020904	0.019521	238.9	46	Inf
Linkage	150-180		0.05	18.5	760	0.058677	0.057071	190	24.1	Inf
Disequilibrium	davs	26	0.02	17.5	1951	0.060761	0.061977	-254.9	39.9	Inf
	uays		0.01	18.9	2535	0.053614	0.055092	-210.1	32.8	Inf
	210-330		0.05	12	527	0.056046	0.056649	-512.1	20.9	Inf
	davs	25	0.02	15	971	0.053817	0.054828	-306.3	40.5	Inf
	uays		0.01	15	971	0.053817	0.054828	-306.3	40.5	Inf
			0.05	210	785	0.005185	0.004834	947.7	402.5	Inf
	Adults	ults 210	0.02	210	2133	0.005229	0.004834	841.3	499.8	2287.6
			0.01	210	2910	0.005163	0.004834	1011.8	599.6	2836.6

The inbreeding in juveniles was stronger when assessed by location which showed lower values of F' at generation 2 compared to F at generation 0 (*T*able III.7). The estimated *Ne*

based on mating behaviour did not showed strong differences between a random mating and non-random mating based on the kinship relationship among adults and juveniles, and within juveniles. The estimated *Ne* based on random mating and partial kinship between juveniles and adults showed a lower *Ne* in juveniles (Ne = 215) compared to the adults (Ne = 293; Table III.8).

Table III.7 Temporal comparisons of the variation of F-statistics between adults (generation 0) and juveniles (generation 2) at three critical values. Method Pk = Pollak (Pollak 1983); method NT = Nei-Tajima (Nei and Tajima 1981); method JR = Jorde-Ryman (Jorde and Ryman 2007); p = critical value (0.05, 0.02 and 0.01); H = harmonic mean; Ai = number of independent alleles; F = F-statistics at generation 0; F' = F-statistics at generation 2; Ne = harmonic mean of the effective population size in generations 0 through 2); <math>CI = Confidence intervals (non-parametric Jacknife). In red are the values that showed a significant reduction in the F-statistics as result of reduction in the Ne at each critical value.

Generations	Method	Р	Н	Ai	F	F'	Ne	(CI
	Pk NT	0.05	148.1	42	0.06825	0.0615	16.3	7.5	29.7
		0.02	148.8	80	0.05531	0.0486	20.6	12.4	31.8
		0.01	152.6	102	0.04584	0.03929	25.5	17.7	35.4
		0.05	148.6	42	0.0694	0.06268	16	7.2	29.5
Adult (0)- Iuvonilo (2)		0.02	149.1	80	0.05471	0.04801	20.8	12.4	32.5
Juvenne (2)		0.01	152.6	102	0.04555	0.03899	25.6	17.5	36.1
		0.05	148.6	42	0.10387	0.09488	10.5	5.5	123.7
	JR	0.02	149.1	80	0.08338	0.07527	13.3	7.2	82.3
		0.01	152.6	102	0.06002	0.05267	19	12.5	39.7

Table III.8 Single sample method full likelihood procedure of the population size differences assuming random mating and non-random mating such as implemented in COLONY. Ne = effective population size; CI = confidence intervals (95%).

Development Stage	Ne	CI	,
Adults	293	240	357
Juveniles	215	175	269

Relatedness

The mean values of relatedness between juveniles and adults were similar (r \approx -0.02;

Figure III.5), and there were significant differences between juveniles and adults with *N* of 21945 ($p \le 0.0001$) in the three methods used to calculate relatedness. Only one average value for LRM is shown and represents the relatedness in juveniles. The adult relatedness was not applicable under LRM algorithm.



Figure III.5 Relatedness levels comparison between juveniles and adults, assessed as the mean within population pairwise values (r). Three estimators were used to calculate the levels of relatedness: Ritland (1996) estimator (RI), Lynch and Ritland (1999) estimator (LRM), and Queller and Goodnight (1989) estimator (QG). The differences between juveniles and adults were significant ($p \le 0.05$) and were tested with Wilcoxon's pair comparisons between dependent samples with N of 21945. There only one average value for LRM which represent the relatedness in juveniles. The adult relatedness was not applicable under LRM algorithm.

Proportion of non-migrant juveniles

There was a high proportion of non-migrant juveniles in each sampling location (~85-

90%; Figure III.6). Migrants were detected among Hauraki Gulf, Bay of Islands, Hokianga and

Kaipara Harbour. The levels of connectivity identified among locations indicated that some

juveniles recruited to different locations even when they potentially shared parents (Figure

III.6).



Figure III.6 Proportion of non-migrants and connectivity patterns by sampling location of <u>M. cephalus</u> juveniles in New Zealand. In green are East-coast locations and in red are West-coast locations. Only the flows that are representing a migration volume of more than 5% are shown. The outer circle segments below the locality name represent the total number of individuals in each location. The inner bars show the total volume of migration encoded in the circle segments. To define directionality, in each location emigrants are represented with the same colour of the location and immigrants are coloured as the locality of origin. The width of the curved line is the volume of movements where the exact volume of movements can be visualized only at the beginning and end points. The width of the curved line depends on number of migrants relative to the total volume of migrants across all regions indicated by the outside circle segments.

Discussion

Mugil cephalus population genetic structure in New Zealand showed no correlation with geographic distances and showed recent admixture of four genetically distinct ancestral populations with a chaotic spatial distribution (Chapter II). In this study, the Sweepstake-

Reproductive-Success (SRS) hypothesis, which is a factor that leads to a pattern of chaotic spatial distribution of the four populations identified in the previous chapter, was tested. The overall pattern found in this study was a significant allele frequency difference between the juvenile age classes when compared to a sample of the adult population. There were lower genetic diversity levels within a juvenile age class and higher levels of genetic differentiation. There was a significant reduction in the effective population size of the pooled juvenile samples and when each age class was treated separately, when they were compared to the adults.

Sweepstake Reproductive Success

The chaotic genetic structure in marine species is often a consequence of seasonal variation in the reproductive success of adults (see Flowers et al. 2002; Hedgecock and Pudovkin 2011; Johnson et al. 2011; Broquet et al. 2013; Horne et al. 2013; Pusack et al. 2014). The seasonal variation in how reproductive marine animals contribute offspring to a future pool of reproductively mature adults, where different effects of bottlenecks could be identified in juvenile cohorts due to different mortality rates forms the body of the Sweepstake-reproductive-success (SRS) hypothesis (Hedgecock 1994).

Allele frequency differences and heterozygote excess

Mugil cephalus in New Zealand showed significant differences between the allele frequencies of each juvenile age class and the adults. The combined distribution of juvenile and adult allele frequencies seemed to complement each other filling the gaps along the allelic range, which is finding consistent with one of the central predictions of the SRS hypothesis. The complementarily allele frequencies between age classes is possibly the result of overlapping generations found in each estuarine location. Only a fraction of the larvae spawned each generation survive and go on to contribute their allelic diversity to the adult group. Differential survival happens because of natural variability in environmental conditions (Hedgecock 1994; Hedgecock et al. 2007b; Christie et al. 2010; Hedgecock and Pudovkin 2011).

The observed differences in allele frequencies could alternatively be the result of a Wahlund effect (Pusack et al. 2014). The Wahlund effect might be expected when genetically distinct cohorts come into contact and are sampled and unintentionally considered members of the same group (e.g. limpets from the genus Siphonaria in western Australia; Johnson and Black 1982, 1984, 2012). A consequence of the Wahlund effect is a low level of heterozygotes than the expected based on the different allele frequencies in the two populations, which arose due to the independent effect of genetic drift in each group (Pusack et al. 2014). However, this pattern was not observed in the present study and some of the msatDNA loci showed a heterozygote excess. In a population that underwent a recent bottleneck, allelic diversity is reduced faster than heterozygosity when compared to the expectation of the HWE (Maruyama and Fuerstt 1985; Cornuet and Luikart 1996; Piry et al. 1999; Galtier et al. 2000). The effects of recent bottlenecks also could be reflected in significant departures from the Stepwise mutation model (SMM) to the Infinite allele model or IAM (Piry et al. 1999). The msatDNA markers in this study follow a SMM model (Chapter II). Loci under a SMM evolutionary model that are exposed to bottlenecks will show random gaps in the allele size range leading to an apparent IAM model and significant levels of He excess (Putman and Carbone 2014; Sharma et al. 2016).

Differences in the effective population size (*Ne*) between cohorts and adults is one of the key SRS predictions (Turner *et al.* 2002; Fraser *et al.* 2007; Riquet *et al.* 2016). There were significant differences between the *Ne* estimates for *M. cephalus* in New Zealand depending on the method used and the method based *He* excess was not reliable. The method to assess *Ne* based on *He* excess, same as the Co-ancestry and *LD* methods, require single-sample comparisons and assumes ancestral population sizes to account for temporal variation between samples that are made of overlapping generations (Fraser et al. 2007; Anderson and Garza 2009; Riquet et al. 2016). The negative or infinite values are interpreted as no evidence of

changes in the *Ne* as a result of a finite number of parents which means that the parents with the possible combination of alleles that could explain the observed allele frequencies in the juveniles, were either not sampled or other factors are causing an increase in the observed *He* (Do et al. 2014). It is possible that for *M. cephalus* in New Zealand, a bigger sample size of the adults would reliably estimate the *Ne* differences between juveniles and adults based on the *He* excess method; however, equally possible is that non-random and unseasonal factors such as human-based activities, could be causing the genetic bottlenecks and the increase in the observed number of *He*.

The population genetic structure analysis of *M. cephalus* adults in New Zealand assessed with a sample size of 576 adults sampled from 15 locations around the North Island, showed a significantly higher proportion of heterozygotes in six locations: Napier, Bay of Islands, Hokianga, Pouto, Kaipara Harbour and Waikato River mouth (Chapter II). Those six locations are nursery estuarine environments for *M. cephalus* in New Zealand surrounded by areas of intensive land-use, which increase sediment and nitrogen loads that modifies or degrades the suitable habitat (Larned et al. 2015). Also, they are priority areas for *M. cephalus* fishery (McKenzie et al. 1999; Watson et al. 2005; McKenzie and Vaughan 2008; Morrison et al. 2014a,c). Further studies are needed to assess the causes of the observed levels in *He* excess in *M. cephalus* juveniles and adults.

Co-ancestry

There was a higher level in the *Nb* of the juvenile age class of 90-120 days followed by a lower level in the 150-180 days age class. The co-ancestry method is usually sensitive to age structured populations and is expected to follow a similar pattern than the *Ne* (Kamath et al. 2015). The co-ancestry method is based in subtracting the putative non-sibling pairs at each locus from the molecular co-ancestry between samples defining the possible number of breeders that could raise the observed number of distinct sample pairs (Do et al. 2014). Overlapping generations will increase the number of allelic configurations affecting directly the number of distinct sample pairs, thus the variation pattern in the *Nb* may differ from the variation pattern of the *Ne* (Do et al. 2014). To overcome this problem, in this study the juveniles were categorized into age classes in each location and pairwise comparisons were performed with the adults; however, the age structure in adults was unknown and the comparisons were made with the pooled adults from all locations.

Considering that the age classes 30-60 days and 90-120 days were clustered as a single group and the distribution of their genetic distances within the factorial components overlapped, a possible explanation to the increase in Nb in the age class 90-120 days is that spawning occurs not as single one-off event where all individuals gather to spawn; but as a continuous that could last several days or weeks. The low Nb in the age class 30-60 days suggests that are the offspring of the breeders reproducing in the last days of the spawning event. Significant differences were found between 150-180 days and the cluster from 30-60 and 90-120 days age classes. The differences between this groups could be a result of two separate spawning events in the same season. One early spawning event which explains the Nb and genetic distances of the 150-180 days age class. The second spawning event lasts several days and explains the similar genetic distances between 30-60 days and 90-120 days age classes and their difference in Nb. Long spawning seasons for M. cephalus have been identified in different parts of the world in which temperature ($\sim 20^{\circ}$ C) is the environmental trigger to spawn (Whitfield et al. 2012). Otolith chemistry assessments have revealed the possibility of a protracted spawning season for M. cephalus in New Zealand (Morrison et al. 2016). The spawning season of *M. cephalus* in New Zealand was thought to be during summer from November to February; however, isotope composition on the otoliths showed the season could start in October and finish in April (Morrison et al. 2016).

Overlapping adult generations in our data set is a potential bias and we cannot rule out the possibility of adults with shared ancestries which could also explain similar genetic distances and different *Nb* between 30-60 days and 90-120 days age class. Adults from the same generation could belong to the same cohort or share ancestry from at least one parent which will result in less variation in their offspring (Li and Hedgecock 1998; Waples 1998; Hedgecock and Pudovkin 2011; Christie et al. 2012). A correlation of age dating analysis with the adults sampled during spawning could help to identify spawning groups and if they are age related.

Linkage Disequilibrium

The *Ne* in juveniles was significantly lower than in the adults. There was a reduction in *Ne* from 30-60 days to 90-120 days age class and an increase towards 150-180 days age class which could be also reflecting the nature of two spawning events (see above). The variation in the *Ne* among age classes reflect different factors causing genetic bottlenecks with different strength in each generation (Iacchei et al. 2013; Do et al. 2014; Selwyn et al. 2016). Under a SRS scenario, the effect of bottlenecks are independent on each generation and the genetic distances not necessarily follow an Isolation-By-Time pattern (Li and Hedgecock 1998; Waples 1998; Hedgecock and Pudovkin 2011; Christie et al. 2012). The genetic distances among *M. cephalus* individuals in New Zealand of different age classes did not follow a linearized correlation with age in an Isolation-By-Time (IBT) pattern (IBT— $R^2 = 0.0475$ p = 0.409).

Linkage disequilibrium methods to assess *Ne* are usually biased by residual disequilibria that has not been dissipated through recombination (Waples 2005; Kamath et al. 2015). Linkage disequilibrium was identified in the loci used in a previous chapter (Chapter II). Residual disequilibria due to different effects of drift in isolated populations before admixture (Appendix Figure III.7), and significant reduction in the effective population size was

detected in each age class of *M. cephalus* juveniles. The difference between the expected asymptote ($r^2_{expected}$) and the observed correlation of allele frequencies at different loci (r^2) is the proportion of the *Ne* changes due to residual disequilibria (Waples 2005). Residual disequilibria has an effect on the *Ne* estimation, where a reduction in the *Ne* reflects the new *Ne* (Waples 2005; Do et al. 2014). An increase in *Ne* will reflect the *Ne* from previous generations (Waples 2005; Do et al. 2014).

Mugil cephalus migratory and reproductive behaviours

Significant levels of gene flow among locations were identified; however, there was a high proportion of non-migrant juveniles (85-90%). It is highly unlikely that juveniles go out from estuaries to swim into other estuaries after settlement. It is more likely that either there are *M. cephalus* adults in New Zealand that can spawn several times during the spawning season, or there could be different spawning grounds where offspring from related parents could get into different estuaries resulting in juveniles in different estuaries with similar allele frequencies. Contrarily to *M. cephalus* in New Zealand, most of the marine species reported with different reproductive success of adults have shown limited or no gene flow between locations (see Carvalho and Hauser 1994; Moberg and Burton 2000; Hauser et al. 2002; Bernal-Ramirez et al. 2003; Hedgecock and Pudovkin 2011). In the few reported cases of marine species that fit the SRS predictions and where gene flow was identified, the latter was a consequence of the stochastic events allowing intermittent migratory episodes (see Aspinwall 1974; Johnson and Black 1982, 2012; Knowlton et al. 1997; Maes et al. 2006; Schtickzelle and Quinn 2007).

Behaviour and ecological flexibility of a species to occupy different ecological niches could also be factors driving SRS in marine species (Joyce et al. 2005; Barluenga et al. 2006; Azuma et al. 2008). One sector of the population could become more specialized and occupy a specific ecological niche, which means they more often mate only with those that have similar preferences (*i.e.* assortative mating). The result is reproductively isolated eco-morph populations, which would evolve in sympatry such as cichlid fishes (Joyce et al. 2005; Barluenga et al. 2006; Azuma et al. 2008).

Mugil cephalus ecological flexibility allows it to inhabit marine, estuarine and freshwater systems and is a highly migratory species (Whitfield et al. 2012). Isolated populations of marine species can respond to different reproduction cues in the environment. If migrants were adapted to different environmental cues for reproduction, after a secondary contact this could eventually lead to the formation of separate species in sympatry. This particular scenario has been reported for the Atlantic herring *Clupea harengus* that inhabits the Baltic Sea (Bekkevold et al. 2005, 2007; Jørgensen et al. 2005; Laikre et al. 2005; Ruzzante et al. 2006). There are two populations of *C. harengus* in the Baltic Sea that are separated because one spawns during summer and the other during winter. The change in the spawning times in *C. harengus* is believed to be the result of founder events from an extant population from the Atlantic that used to spawn during winter. The migrant offspring have maintained the reproductive time of the ancestors and remain isolated even though their distributions now overlap (Bekkevold et al. 2005, 2007; Jørgensen et al. 2005; Laikre et al. 2005; Ruzzante et al. 2006).

Temperature seems to be the important determinant factor that triggers spawning in M. *cephalus* (Whitfield et al. 2012). It has been reported that temperatures around 20°C, are optimal for M. *cephalus*. More importantly, the spawning seasons of M. *cephalus* are different in other parts of the world (Whitfield et al. 2012 and references within). For example, in the Mediterranean the spawning season is from June to October (summer), in the Atlantic and Indian oceans it takes place from October to March during winter conditions (Whitfield et al. 2012). In eastern Australia it is from March to July (Kesteven 1953). The spawning season of M. *cephalus* populations in New Zealand, is believed to be during summer from November to

February (MPI 2009). However, otolith analysis of *M. cephalus* in New Zealand revealed that they have a protracted spawning season in spring/summer from October to March (McKenzie et al. 1999).

Reproductive strategies might strengthen the effects of SRS in marine fish populations with chaotic genetic structure. Philopatry in fish has been reported as a factor aiding to SRS and chaotic genetic structure in marine species (Bekkevold et al. 2005, 2007; Svedäng et al. 2007). Philopatry in marine fish populations and *M. cephalus* may be difficult to imagine because it requires adult fish returning to their spawning grounds, but they only experienced the area as passive drifters in the water column in the form of eggs or recently hatched larvae (Bentzen and Bradbury 2016). The mechanisms that allow marine fish species to have such navigation abilities are yet to be identified; however, chemical cues, geomagnetic imprinting or a combination of both are the possible candidate mechanisms (Thorrold et al. 2001, 2007; Jones et al. 2005; Putman et al. 2013).

Philopatry has been identified in two *M. cephalus* populations in Taiwan (Hsu et al. 2007). Two different spawning grounds of *M. cephalus* populations in Taiwan were identified based on gonadal development, oocyte sizes and similarities in environmental conditions. They showed that adults return to the same spawning ground each season. Genetically differentiated populations of *M. cephalus* as a consequence of philopatry have also been reported in Taiwan and China Seas (Huang et al. 2001). Two sub-populations share spawning grounds in Taiwan's coastal waters; a nomadic population that travels to the Chinese sea after the spawning season and a resident non-migratory population. These appear to have become two differentiated populations because of the fixation of one allele of the glucose 6-phosphate-isomerase-A (GPI-A) locus in the nomadic population. The protein produced by the GPI-A gene is involved in glycolysis, gluconeogenesis and as a neurotrophic factor.

In New Zealand, direct evidence of philopatry in *M. cephalus* populations has not been reported. Philopatry would explain the high proportion of non-migrant juveniles (85%-90%), the non-significant differentiation levels among locations and the similar proportion of ancestries found in distant locations (Figure II.6). Direct and indirect methods such as tracking, capture-recapture and fine-scale chemical analysis, used in combination with genetic methods, could help identify *M. cephalus* spawning grounds and test for the different patterns of spawning and philopatry.

Conclusion

M. cephalus populations in New Zealand showed evidence of a SRS effect when the allele frequencies, genetic diversity and shifts in *Ne* were compared between juvenile age classes and adults. Two spawning groups were identified, an early spawning group and a late spawning group. Philopatry could explain the non-significant levels of genetic differentiation among locations; further analysis is needed to test philopatry in *M. cephalus* in New Zealand. The SRS did not have enough resolution to explain all the variation found in each sampling location and a correlation among environmental variables, human-based activities and genetic diversity levels is needed.

Appendix

Locus	Null allele Frequency	Lower Boundary 0.025	Upper Boundary 0.975	Genotyping failure rate
Mce3	0.0649	0.0424	0.0931	0.00
Mcs16EM	0.0162	0.0000	0.0548	0.00
Mce2	0.0104	0.0000	0.0445	0.00
Mcs16DM	0.0641	0.0360	0.1018	0.00
Mce27	0.1168	0.0864	0.1533	0.00
Mcs2FH	0.0701	0.0468	0.0995	0.00
Mce24	0.0403	0.0155	0.0757	0.00
Mean	0.0546			
+SD	0.0363			

Table III.9 Null allele frequency and genotyping failure rate in juveniles by each locus.

Table III.10 Effective population size (Ne) and Number of breeders (Nb) differences between adults and juveniles by sampling location at three critical values. D. E. = developmental stage; N = number of samples; H = Harmonic mean size; F1 = single-sample variation of F-statistics; Nb = Effective number of breeders; CI = Confidence intervals for Ne with non-parametric Jacknife loci; p = critical value; Ai = Number of alleles that are independent; Ne = effective population size; inf = infinite.

Ne/Nb estimator	Location	D.E.	N	H	F1	Nb	CI				
	Bay of	Juveniles	57	34.5	0.03886	12.9	5.7	22.8			
	Plenty	Adults	18	18	0.03827	13.1	2.7	31.5			
	Hauraki	Juveniles	23	12.2	-0.00156	Inf	Inf	Inf			
	Gulf	Adults	48	48	-0.03031	Inf	Inf	Inf			
C	Bay of	Juveniles	38	26.9	0.04366	11.5	7	16.9			
Co-ancestry	Islands	Adults	48	48	0.00154	324.2	0.3	1627.3			
	TT-1-1	Juveniles	25	17.3	0.02643	18.9	4.6	43.3			
	Hokianga	Adults	48	48	0.01585	31.5	2.3	98.3			
	Kaipara	Juveniles	29	15.1	-0.01688	Inf	Inf	Inf			
	Harbour	Adults	48	48	-0.01698	Inf	Inf	Inf			
				р	H	Ai	\mathbf{r}^2	r ² _{expected}	Ne	CI	
				0.05	27.9	755	0.03225	0.028511	80.7	22	Inf
		Juveniles	57	0.02	26	2935	0.034093	0.031528	118.4	23.3	Inf
	Bay of			0.01	28.7	3582	0.028068	0.02672	226.8	30.8	Inf
	Plenty			0.05	18	863	0.069362	0.065893	87	33.8	Inf
		Adults	18	0.02	18	2115	0.066122	0.065893	1338.9	71.4	Inf
				0.01	18	2115	0.066122	0.065893	1338.9	71.4	Inf
	Hauraki Gulf			0.05	9.8	814	0.092704	0.093743	-298.1	14.7	Inf
		Juveniles	23	0.02	11.9	1536	0.077275	0.078084	-382.3	4.2	Inf
				0.01	11.9	1536	0.077275	0.078084	-382.3	4.2	Inf
				0.05	48	643	0.021069	0.022218	-292.1	398.6	Inf
		Adults	48	0.02	48	1566	0.021363	0.022218	-392.1	440.4	Inf
				0.01	48	2482	0.02213	0.022218	-3779.4	203.2	Inf
				0.05	21.1	1010	0.0466	0.045726	350.9	25.8	Inf
		Juveniles	38	0.02	22.4	2741	0.044307	0.042944	224.2	29.3	Inf
Linkage	Bay of			0.01	24	3329	0.037344	0.038945	-194.1	37.3	Inf
Disequilibrium	Islands			0.05	48	1022	0.023011	0.022218	418.3	136.5	Inf
		Adults	48	0.02	48	2639	0.022732	0.022218	646	167.3	Inf
				0.01	48	3592	0.022875	0.022218	505.6	149	Inf
				0.05	15	587	0.069211	0.067418	170.1	15.5	Inf
		Juveniles	25	0.02	16	2466	0.060266	0.062327	-151.1	24.4	Inf
	Haltionaa			0.01	16	2466	0.060266	0.062327	-151.1	24.4	Inf
	поклапда			0.05	48	1023	0.022379	0.022218	2072.7	141.1	Inf
		Adults	48	0.02	48	2643	0.021709	0.022218	-656.9	481.1	Inf
				0.01	48	3862	0.022328	0.022218	3030.2	240.3	Inf
				0.05	13.7	884	0.046009	0.046299	-1063.4	45.9	Inf
		Juveniles	29	0.02	14.9	1063	0.044949	0.044918	1043.5	52	Inf
	Kaipara			0.01	17.7	1599	0.042563	0.043005	-697.5	47.1	Inf
	Harbour			0.05	48	935	0.022217	0.022218	-441760.3	260.3	Inf
		Adults	48	0.02	48	2132	0.022524	0.022218	1085.4	177.2	Inf
				0.01	48	3443	0.022393	0.022218	1906.3	230.5	Inf

Table III.11 Temporal comparisons of the variation of F-statistics between adults (generation 0) and juveniles (generation 2) in each sampling location at three critical values. Method Pk = Pollak (Pollak 1983); method NT = Nei-Tajima (Nei and Tajima 1981); method JR = Jorde-Ryman (Jorde and Ryman 2007); p = critical value (0.05, 0.02 and 0.01); H = harmonic mean; Ai = number of independent alleles; F = F-statistics at generation 0; F' = F-statistics at generation 2; Ne = effective population size; CI = Confidence intervals (non-parametric Jacknife). In red are the values that showed a significant reduction in the F-statistics as result of reduction in the Ne at each critical value.

Location	Method	р	H	Ai	F	F'	Ne	(CI
		0.05	23.3	41	0.11069	0.06783	14.7	6.6	34.8
	Pk	0.02	23.7	89	0.08146	0.03921	25.5	14.7	49
		0.01	23.8	116	0.07177	0.0297	33.7	19.8	65.8
Dow of		0.05	23.4	41	0.10653	0.06376	15.7	7.2	37
Day 01 Dionty	NT	0.02	23.7	89	0.08024	0.03798	26.3	15.1	51.2
rienty		0.01	23.8	116	0.07125	0.02917	34.3	20.4	65.8
		0.05	23.4	41	0.13776	0.09349	10.7	5.6	139.3
	JR	0.02	23.7	89	0.11658	0.07302	13.7	7.8	57.8
		0.01	23.8	116	0.11392	0.07044	14.2	8.2	52
	Dir	0.05	20.5	38	0.21279	0.16393	6.1	3.5	10.2
	Pk	0.02	19.5	69	0.14899	0.09759	10.2	6.9	15.4
		0.01	19.7	87	0.1277	0.07695	13	9	19.1
Hauraki Gulf Bay of		0.05	20.3	38	0.20117	0.15193	6.6	3.8	10.9
	NT	0.02	19.5	69	0.14227	0.09088	11	7.3	16.6
		0.01	19.7	87	0.12346	0.07266	13.8	9.4	20.4
		0.05	20.3	38	0.30506	0.23456	4.3	2.9	8.4
	JR	0.02	19.5	69	0.20441	0.14466	6.9	5	11.1
		0.01	19.7	87	0.18842	0.12998	7.7	5.6	12.5
	Pk	0.05	33.2	45	0.07146	0.04131	24.2	7.7	99.1
		0.02	33.8	83	0.06351	0.03392	29.5	16.2	58.2
		0.01	34.3	106	0.05716	0.02799	35.7	22.5	60.7
		0.05	33.3	45	0.07272	0.04272	23.4	7.5	91.3
	NT	0.02	33.8	83	0.06298	0.03343	29.9	16.1	61.1
Isianus		0.01	34.3	106	0.05681	0.02765	36.2	22.2	63.5
	JR	0.05	33.3	45	0.11554	0.08415	11.9	5.6	Inf
		0.02	33.8	83	0.08439	0.05418	18.5	9.9	129.6
		0.01	34.3	106	0.0779	0.04788	20.9	11.6	104
		0.05	24.6	47	0.11953	0.07886	12.7	5.8	28
	Pk	0.02	25	85	0.09598	0.05591	17.9	9.2	37.2
		0.01	25	96	0.08907	0.04906	20.4	10.4	44
		0.05	24.7	47	0.11622	0.07573	13.2	6.1	29.2
Hokianga	NT	0.02	25	85	0.09252	0.0525	19	9.8	40.1
		0.01	25	96	0.08642	0.04645	21.5	10.9	47.4
		0.05	24.7	47	0.15194	0.10819	9.2	5.2	43.8
	JR	0.02	25	85	0.11435	0.07261	13.8	8.1	45.3
		0.01	25	96	0.1098	0.06826	14.7	8.4	58.2
		0.05	25.5	44	0.12423	0.08497	11.8	6.1	22.3
	Pk	0.02	26.6	70	0.11483	0.07728	12.9	8.2	20.5
		0.01	27.5	86	0.09596	0.05953	16.8	10.4	27.6
Kainara		0.05	25.1	44	0.11575	0.07588	13.2	7	25.1
Harbour	NT	0.02	26.2	70	0.10827	0.07016	14.3	9	22.8
		0.01	27.1	86	0.09187	0.0549	18.2	11.2	30.3
		0.05	25.1	44	0.12821	0.08148	12.3	7.1	43.9
	JR	0.02	26.2	70	0.11752	0.07131	14	8.4	42.8
		0.01	27.1	86	0.11324	0.06722	14.9	8.6	54.2



Figure III.7 Variation of r^2 compared to the expected r^2 of the Ne of juveniles and adults at three different critical values (p = 0.05; 0.02; 0.01). r^2 is shown in solid lines; expected r^2 is shown in dotted lines. r^2 variation at 0.05 critical value = yellow; r^2 variation at 0.02 critical value = blue; r^2 variation at 0.01 critical value = red.

Population genetic structure and lineage divergence of the Kānae (*Mugil cephalus*) in New Zealand based on two mitochondrial DNA markers

Abstract

Mugil cephalus has a circumglobal distribution which suggests it is capable of transoceanic migrations even though it is a species commonly associated to coastal areas, estuaries and rivers. Previous population genetic studies based on mtDNA have reported highly structured populations in the north Pacific and north Atlantic. In New Zealand, recent admixture of four genetically distinct populations was identified with nuclear microsatellite DNA markers. In this study, the phylogeographic structure was investigated using DNA sequences from the mitochondrial DNA Cytochrome Oxidase I (COX1) gene and D-loop region. Samples were collected from 12 sites in New Zealand, and compared to samples from Australia, Chile, and previously published sequences from Pacific locations. Analyses of the data showed two sympatric haplogroups of *M. cephalus*, which were both in mutation-drift equilibrium. The wo haplogroups did not match the four populations identified with nuclear markers suggesting a secondary contact between allopatric populations followed by population expansion. There was no evidence of isolation-by-distance, latitudinal-gradients of haplotype diversity or phylogeographic breaks within the New Zealand populations; instead, there appeared to be extensive levels of gene flow among the sampled locations. Based on the available samples and the locations surveyed in New Zealand and the Pacific, New Zealand was most likely colonised by two allopatric *M. cephalus* populations from the Pacific between ~50,000 and ~16,000 years ago; since then, M. cephalus in New Zealand has been in genetic isolation from other Pacific areas.

Key words: Phylogeography, population genetic structure, secondary contact, colonisation, high-density blocking.

Introduction

Identifying the barriers that cause intra-specific genetic breaks in marine species is important for understanding patterns of distribution and evolution processes. Long-term environmental barriers to gene flow are the most common factors considered to genetically structure genetic diversity and to isolate populations (Palumbi 1994; Kotoulas et al. 1995). Less well understood is when environmental conditions have changed through time resulting in short bouts of gene flow between otherwise isolated populations (Reid et al. 2016). For example, phylogeographic analyses have suggested that changes in sea-level due to glacial periods caused shifts in surface and bottom currents, allowing intermittent gene flow among populations of the peacock grouper (Cephalopholis argus) in the Indo-Pacific, the grey mullet (Mugil cephalus) in the China Sea, and the bluefish (Pomatomus saltatrix) in the south-east Atlantic (Gaither et al. 2011; Shen et al. 2011; Reid et al. 2016, respectively). Species with a high dispersal power could overcome barriers colonising new habitats and promoting gene flow between populations; however, the changes in environment conditions might allow species with limited dispersal to colonise distant areas (Kotoulas et al. 1995; Lewallen et al. 2011; Shen et al. 2011; Hyde et al. 2014; Reid et al. 2016). Intermittent gene flow can be difficult to identify because it produces complicated patterns of genetic connectivity instead of a simple relationship between the population genetic structure and geographic distribution (Hedgecock et al. 2007a; Hellberg 2009). Phylogeographic studies have used the finding of genetically distinct lineages and different haplotype frequencies to identify the historical processes that have restricted gene flow and caused population genetic differentiation (Jiggins 2006; Bolnick and Kirkpatrick 2007; Fitzpatrick et al. 2008; Wolinsky 2010).

After a new population become established demographic processes such as highdensity blocking or competitive exclusion can limit ongoing gene flow from the source or other populations, even when environmental conditions are suitable for successful migration (Ostfeld and Canham 1995; Waters et al. 2013b; Borowski et al. 2016). Competitive exclusion takes place when a second wave of migrants arrive to a colonised area and occupy the same ecological niche as the residents (Waters 2011). Competitive exclusion can occur, not only for the secondary dispersers, but also for individuals behind each colonisation front and it can affect other species attempting to colonise a site (Waters 2011). The lack of ongoing gene flow from the source population, the founder effect, and genetic drift can act in concert to cause newly founded population to rapidly become genetically distinct (Hewitt 1999, 2000). If a second migration wave is able to establish successfully from a different source population, the newly foundered population can contain two or more divergent haplogroups that occur in sympatry. In cases where the two groups continue to intermix, divergent mitochondrial DNA haplotypes might persist but nuclear loci will become homogenized (Palumbi 1994; Hewitt 1999; Hendry and Day 2005; Pujolar et al. 2006). If the secondary colonisers are less likely to reproduce with individuals from the first colonisation event the differences between the two group will be maintained in both the mitochondrial and nuclear genomes (Hendry and Day 2005; Maes et al. 2006). Reproductive isolation and the persistence of two genetically distinct populations with overlapping distributions have been reported in the Atlantic herring (Clupea harengus) in the Baltic Sea (Bekkevold et al. 2005, 2007; Jørgensen et al. 2005; Ruzzante et al. 2006). The separate C. harengus groups are maintain by difference in spawning times; one population spawns during the summer and the other during the winter. This is thought to have been the result of several colonisation events from a winter-spawning population that is now extinct.

Mugil cephalus is a cosmopolitan coastal fish species and has larvae that can drift on surface ocean currents for over a month before settlement (Cardona 2000; Salvarina et al. 2010; Whitfield et al. 2012). This species can tolerate a wide range of dissolved oxygen levels and fresh and hyperhaline waters, which enables them to colonise estuarine and freshwater systems (see Whitfield et al. 2012 for a review). Significant levels of genetic differentiation among populations have been reported among the Atlantic, the Indian and the Pacific Oceans (Rossi et al. 1998a; Rocha-Olivares et al. 2000; Livi et al. 2011). Moreover, divergent populations have been identified along the coast of China, and among harbours and coastal locations between Pacific Russia and the Philippines (Rossi et al. 1998a; Jamandre et al. 2009; Shen et al. 2011; Durand et al. 2012b; Sun et al. 2012; Whitfield et al. 2012). The mtDNA sequence divergence among *M. cephalus* lineages, its paraphyletic relationship, its preference for coastal habitats and limited dispersal capability has led to consider *M. cephalus* as a species complex (Durand et al. 2012b; Whitfield et al. 2012). To date, 14 highly divergent mtDNA lineages have been identified, in which three lineages distributed in coastal areas from the northwest Pacific were also supported by nuclear microsatellite DNA markers (Shen et al. 2011; Durand et al. 2012b). However, there have not been any distinct morphological traits nor direct evidence of reproductive incompatibility reported among the *M. cephalus* haplogroups, making it difficult to determine whether *M. cephalus* is truly a species complex, or it has a very large genetically effective population size and highly structured populations.

Mugil cephalus is commonly known in New Zealand as Kānae or grey mullet. It was a resource widely used by pre-European Māori (the indigenous people of New Zealand Aotearoa) and one of the first major European fisheries in the North Island, starting in the late 1800 (Paulin and Paul 2006). New Zealand *M. cephalus* spawn in the sea during the summer. Its larvae drift in ocean currents for over a month before post-flexion larvae and juveniles move into estuarine nursery areas in the North Island (Hartill 2004). They mature at about 3 years old, have a maximum age of 20 years and reach an average size of 55cm (Hartill 2004; Watson et al. 2005; McKenzie and Vaughan 2008; Morrison et al. 2016). Few studies have focused their efforts on the distribution and population status of this species and there have been no genetic studies. The aim of this study is to investigate the phylogeographic structure and

demographic history of New Zealand *M. cephalus* using DNA sequences from two mtDNA gene regions. Levels of mitochondrial DNA variation in New Zealand's *M. cephalus* are compared to the DNA sequence variation reported from other populations around the Pacific Rim.

Materials and Methods

Data collection

DNA from a total of 311 individuals (116 adults and 154 juveniles from 12 locations around New Zealand, 10 adults from eastern Australia and 31 adults from Chile) were extracted (Appendix Table IV.6; Appendix Table IV.7; Figure IV.1). The DNA was extracted from a tissue biopsy from the caudal peduncle following the phenol/chloroform extraction protocol and purified DNA was stored in TE buffer at 4°C (Sambrook et al. 1989). To compare the phylogeographic history of *M. cephalus* in New Zealand with the rest of the Pacific, sequences of studies from Pacific Rim were downloaded from GenBank and the data set comprised 116 sequences of Cytochrome Oxidase I mitochondrial gene (COX1) and 129 sequences of the control region (D-loop) of the mitochondrial DNA (mtDNA) (Appendix Table IV.6; Figure IV.1). The comparisons between New Zealand and the Pacific Rim analyses had to be conducted separately for COX1 and D-loop because there were no sequences from both genes that came from the same individual.



Figure 0.1 Locations where <u>M. cephalus</u> samples were collected in New Zealand (green dots). Red dots are the sampling locations of COX1 sequences and blue dots represent the sampling locations for D-loop sequences downloaded from GenBank

Amplification and alignment

Amplification of mtDNA gene COX1 was obtained by PCR using specific primers developed using AmplifiX v1.7.0 (Jullien 2013). Four complete mitochondrial genomes of *M. cephalus* were retrieved from GenBank and were used to design the COX1 primers: AP002930 and AP002931 (Miya et al. 2001), KM368340 (Shen et al. 2015), and NC_003182 (Miya et al. 2001). We targeted partial sequences of ~650bp length from the Tyrosine tRNA gene as forward primer (MugceCO1F-Forward primer 5' CAC AAA GAC ATC GGC ACC CTC TAT 3') to mid COX1 sequence as reverse primer (MugceCO1R-Reverse primer 5' TGG GTG GCC AAA GAA TCA GAA CAG 3').

D-loop sequences were amplified with the forward primer Mul-Pro (5' CCA AGG CCA GGA TTT TTA CGT 3'; Rocha-Olivares et al. 2000) and a reverse primer designed from AP002930 (Miya et al. 2001), KM368340 (Shen et al. 2015), and NC_003182 (Miya et al. 2001) complete mtDNA genomes in GenBank. The reverse primer Mugce DLR-Reverse (5' GGC TCT AAG GTT TTC CTG TTT CCG 3') was developed in AmplifiX v1.7.0. The PCR volume used to amplify the targeted sequences was 25µl of MgCl₂ 2mM, 0.6µM of each primer, 0.6µM of dNTP's, 0.4 µg/µl of bovine serum albumin, 1µl of 10x PCR buffer (200 mM Tris HCl (pH 8.4), 500 mM KCl), 1 unit of Taq DNA-polymerase (Bioline), and between 50-80 ng/µl of template DNA. Thirty-five PCR cycles were run with 5 minutes of initial soak at 95°C, 94°C for 30s, 60°C for 30s, 72°C for 90s and a final elongation at 72°C for 30 minutes. Products were purified with a 1:1 mix of ExoSap-IT (Affymetrix, USA) and 10x PCR buffer (200 mM Tris HCl (pH 8.4), 500 mM KCl) from Bioline. Amplified sequences were sent to Macrogen Inc. (Korea) for sequencing. Cytochrome Oxidase I and D-loop sequences were deposited in GenBank with accession numbers KX639485-KX639489 and KX639435-KX639484 respectively (Appendix Table IV.7). The resultant chromatograms were visualized using Geneious v8 (Kearse et al. 2012), and consensus sequences were used for the alignment.
Sequences were aligned in Clustal W (Thompson et al. 1994), checked by eye and trimmed to equal lengths in Geneious v8. The best substitution model for COX1 and D-loop sequences was calculated with JModelTest v2.1.4 (Posada 2008), following the lowest likelihood ratio under the Bayesian Information Criterion (BIC).

Lineage divergence

Two phylogenetic trees with the concatenated sequences of COX1 and D-loop were built in order to identify possible M. cephalus populations in New Zealand. A Bayesian Inference (BI) phylogenetic tree was built in Mr. Bayes v3.2.2 (Huelsenbeck and Ronquist 2001). The BI analyses were performed with 5×10^6 generations in two simulations with one cold and three heated parallel chains. The sample frequency was 100, the parameter values were saved every 5000 rounds and 10,000 trees were discarded as burn-in. The analysis finished when stationarity was reached. A Maximum Likelihood phylogenetic tree was built in RAxML v.8 (Stamatakis 2014) and implemented in CIPRES (Cyberinfrastructure for Phylogenetic Research) (Miller et al. 2010). The model used was GTR+ G+ I for the construction of both phylogenetic trees to account for the variability in the substitution rate of the sequences coming from both mtDNA markers when concatenated and with no partitions. Support values were calculated with 1000 bootstrap iterations and only values > 89% were reported. Both phylogenetic trees were compared to recognize supported groups and only those are reported. A second ML phylogenetic tree was constructed using COX1 sequences from New Zealand and the COX1 sequences downloaded from GenBank to place New Zealand in a Pacific Rim context. The ML tree was built in RAxML v.8 and implemented in CIPRES. The model used was GTR+ G+ I to account for the variability in the substitution models used in previous studies. Support values were calculated with 1000 bootstrap iterations.

Divergence time and cryptic lineages discovery

To designate the optimal divergence time a general mixed Yule-coalescence (GYMC) model with the concatenated sequences of COX1 and D-loop was used (Monaghan et al. 2009).

This analysis is commonly used to delimitate species and to find cryptic entities that show genetic variation that can be explained by speciation models rather than coalescence models (Talavera et al. 2013). However, it is a reliable analysis to define the divergence time when there is no information on the fossil record or other evidence to calibrate the nodes of a phylogenetic tree and coalescent genealogies to the most recent common ancestor (MRCA) (Talavera et al. 2013). The GMYC model could provide an approach that it is not biased by the high variation in haplotype sequences to estimate the divergence time (Monaghan et al. 2009; Talavera et al. 2013).

The GMYC models are especially robust in determining the time of divergence between unique haplotypes because it assess the genealogy to the MRCA using a Wright-Fisher model modified for mtDNA haplotypes in which each node or step to the MRCA contains information of the mutation rate and effective population size (Kingman 2000; Rosenberg and Nordborg 2002). In the Wright-Fisher model each new generation is formed by randomly sampling Nparents and with full replacement from the current generation; hence, generations are discrete (Kingman 2000). Going back in time, lineages coalesce whenever two or more individuals were produced by the same parent in which the offspring inherits the haplotypes of the parents unless there is a mutation that occurs with some probability each generation (Rosenberg and Nordborg 2002). The GMYC model draws the genealogy until the nodes to the MRCA match the nodes drew by the phylogenetic approach under Yule's speciation model (Monaghan et al. 2009). While the coalescent approach started with *n* haplotypes picked at random to coalesce resulting in *n*-1 ancestral haplotypes, in the Yule model at any given time the then extanthaplotypes (n) are equally likely to speciate giving rise to n+2 haplotypes and the extinction of the ancestral haplotype (Steel and McKenzie 2001; Gernhard et al. 2008). In the coalescent approach the mutation rate varies in each node; under the Yule speciation model the probability of speciation or rate of speciation is constant over time (Steel and McKenzie 2001; Gernhard et al. 2008).

The GYMC model estimates the boundary of divergent lineages based on the linearization of the genetic distances of the coalescent nodes and Yule's speciation events, *i.e.* draws a threshold when the intra-haplogroup variation matched or exceeded the interhaplogroup variation under Yule's speciation model (Monaghan et al. 2009; Talavera et al. 2013). A fully resolved tree based on Yules' speciation model was used as prior to build a distance cluster and identify entities that met the assumption of Yule's speciation (Kekkonen and Hebert 2014). To build the tree, single haplotypes of the concatenated sequences were used and a fully resolved ultrametric tree was constructed with three independent runs using Bayesian Inference and Yule's speciation model as a prior in BEASt v1.8.2 (Drummond and Rambaut 2007). To account for all possible source of variation in the unique sequences, the nucleotide substitution model used was the GTR+G+I that also allows different nucleotide substitution rates of the concatenated sequences. The base frequencies were estimated from the data set with four gamma categories and no codon partitions. To allow for variability in the mutational rate over the tree, an uncorrelated relaxed lognormal clock model was used with $2x10^{-7}$ years as the mutational rate (Ward et al. 2005; Durand et al. 2012b). The ucld.mean parameter was set to uniform prior distribution with zero as the lower and 10 as the upper boundary. The length of the MCMC chain was 400 000 000 sampling every 4000 until the ESS values converged (ESS>200). To confirm convergence of the Markov chains, the ESS values of each run were evaluated in Tracer v1.8.2 (Drummond and Rambaut 2007). The resulting trees (states) were combined and re-sampled every 4000 trees on LogCombiner v1.8.2 (Drummond and Rambaut 2007). The remaining trees were used to build the maximum clade credibility tree with a 0.05 posterior probability limit and node heights of target tree constructed in TreeAnnotator v1.8.2 (Drummond and Rambaut 2007). Both single- and multiple-threshold GMYC analyses were conducted in R (R Core Team. 2012) using the APE (Paradis et al. 2004), and SPLITS v1.0 (Ezard et al. 2009) packages.

Neighbour Joining clustering

Two Neighbour Joining (*NJ*) trees were constructed to find the pairwise differences among lineages between New Zealand and Pacific locations. The *NJ* trees were built in MEGA v6 (Tamura et al. 2013); using Kimura 2-parameter substitution model including transitions + transversions, with uniform rates among sites, three codon partitions for COX1 sequences and none for D-loop sequences. To test the support of each lineage cluster, 10, 000 bootstrap replicates were used for each *NJ* tree.

Genetic structure and diversity

Genetic diversity measures by mtDNA marker and locality were calculated in DNAsp v5.1 (Librado and Rozas 2009), as number of haplotypes, number of segregating sites, haplotype diversity (h), nucleotide diversity (π), and Tajima's D and Fu's F deviation from neutrality. In order to estimate the variation given by the sampling effort and uneven sample sizes, an analytical method of rarefaction was performed in Analytical Rarefaction v1.3 (UGA Stratigraphy Lab Website; http://www.uga.edu/~strata/software). Rarefaction curves were plotted using the number of samples against the cumulative frequency of the haplotypes of each mtDNA marker. To estimate haplotype network relationships per locality, a median-joining network (Bandelt et al. 1999), and pie charts of frequency distribution of haplotypes were calculated in PopArt (Leigh and Bryant 2015) with default settings. An Analysis of Molecular Variance (AMOVA), ϕ_{ST} values with their corresponding probability with Bonferroni correction for multiple analyses were calculated in Arlequin v3.5 (Excoffier and Lischer 2010). The AMOVA allows for the assessment of the genetic differences in three levels of hierarchy: among regions, among localities and within localities. To test the hypothesis of Isolation-By-Distance (IBD) in New Zealand, a Mantel test with 10,000 permutations was also calculated in Arlequin v3.5.

Mitochondrial DNA and microsatellite DNA structure comparison

The population genetic structure analysis of *M. cephalus* in New Zealand based on nuclear microsatellite DNA markers (msatDNA), identified the recent admixture of four ancestral genetically distinct populations (Chapter II). In order to test whether there was a significant level of similarity between the clusters found with allele frequencies (msatDNA) and the possible mtDNA haplogroups, an assignment analysis for the mtDNA sequences was performed using the software Geneclass v2 (Piry et al. 2004) to calculate the probability of individual assignment to the locality of origin or sampling location. The Rannala and Mountain Bayesian method was used (Rannala and Mountain 1997). Likelihood estimations were calculated with L_home. The probability of the null hypothesis that each individual is from the source haplogroup was estimated using 1,000,000 simulations (Monte Carlo resampling) under Paetkau's algorithm (Paetkau et al. 2004). The assignment probabilities were then compared to the highest cluster probability of the individual ancestry from STRUCTURE v2.3.4 (Pritchard et al. 2000). Structure assigns multi-locus genotypes to an ancestral cluster and calculates the maximum likelihood of the number of clusters that contribute to each individual ancestry and can explain the allele frequency distribution (see Chapter II).

Demographic history

Possible events of population expansion were evaluated calculating mismatch distributions per marker using Arlequin v3.5. Models of constant population size and population growth-decline were tested in DNAsp v5.1. Theta parameters θ_0 and θ_1 of effective population-size shifts estimated from mismatch distributions, and deviation from neutrality (Tajima's *D* and Fu's F statistics) under the infinite-sites model were calculated in Arlequin v3.5. To reconstruct the demographic history and calculate the change in the effective population size over time, BEASt v1.8.2 software was used to calculate Bayesian Skyline Plots (BSPs) for each marker and for the concatenated sequences. The substitution model used for each marker was determined with the lowest likelihood ratio in BI calculated with JModelTest

v2.1.4 (see above). For the concatenated sequences the GTR+G+I substitution model was used to allow for variability in the substitution rate along the sequence. Three independent runs were performed with MCMC length of 100 000 000 and sampled every 1000 and the resulting states were combined and re-sampled every 1000 trees on LogCombiner v1.8.2. The skyline plots with 95% confidence levels were drawn in Tracer v1.8.2, only after the chains converged (ESS values > 200).

Migration and connectivity patterns

Migration ratios and the number of effective migrants among sampling locations were assessed in Migrate-n v3.6 (Beerli 2009), using a Metropolis-Hastings algorithm to sample multi-dimensional distributions of Theta (θ) with Maximum likelihood (ML) and with Bayesian inference (BI) approaches (Beerli and Felsenstein 1999, 2001). One hundred seeds, 50, 000,000 MCMC iterations, a thinning of 1,000,000 and a sampling interval of 1,000 were used. Migrate-n uses a coalescence approach to determine the migration rates between locations using the genealogy to the most recent common ancestor (MRCA). The genealogy is built including all segregating sites in which each step to the MRCA is defined as an ancestral generation and it is drawn as a node on a line that merges all the way back to the MRCA (Kingman 2000). Migrate-n calculates the probability that two individuals in a particular generation have a common ancestor (Beerli 2009). The probabilities are calculated integrating all possible relationships in each ancestral generation following the genealogical tree. To estimate the number of immigrants per generation, the migration rate (M) and effective population size (θ) per generation were scaled to the mutational rate (m) (Beerli 2009). In each node or generation, the effective population size or theta (θ) is calculated and to obtain the value of θm (theta scaled to the mutational rate), theta is multiplied by the mutation rate (m) that is usually inferred from the data set, and a constant that depends on the ploidy level and the inheritance of the DNA markers (χ) (Beerli 2009). In mtDNA haplotypes $\chi = 1$ (single copy from only one parent). The immigration rate in each generation (M) is calculated using θm and multiplying it by the number of migrants and dividing it by the mutation rate (Beerli 2009). The result reflects the number of variants brought to the population each generation that are not a consequence of mutation. Migrate-n integrates all migration rates (*M*) and effective population sizes per generation (θm) to estimate the number of migrants since the MRCA (Beerli 2009). The results were represented graphically with the R package circlize (Gu et al. 2014); in which the width of the curved lines in the circular plot represents the number of migrants and is dependent on the total volume of migrants across all regions.

Results

New Zealand

A total of 241 samples were amplified for the COX1 and 94 samples amplified from the D-loop region (Table IV.1). For the analyses done with concatenated sequences only 94 samples that were successfully amplified for both markers were used. After alignment and trimming the remaining COX1 sequences length was 502bp with five variable sites and five distinct haplotypes were identified. The D-loop region sequence length was 864bp and 176 polymorphic sites were identified within 50 distinct haplotypes (Table IV.1). A tandem repeat was found on the 3' end (((ATA)₂TT)₂₋₁₂), close to the Phenylalanine transfer RNA (tRNA-Phe). Similar tandem repeats have been found in other M. cephalus populations in the Gulf of Mexico and northwest Pacific (Rocha-Olivares et al. 2005; Jamandre et al. 2009, 2014). In this study, the tandem repeat was removed from the analysis to avoid large variance, limit homoplasy and avoid bias in the demographic events that led to the observed whole sequence polymorphisms, nucleotide diversity and haplotype diversity (Jamandre et al. 2009). The nucleotide substitution evolutionary model that best fits COX1 sequences was HKY, with uniform distribution and kappa = 2.4056 (ti/tv = 1.2728). D-loop region best nucleotide substitution model was HKY+I+G with p-inv = 0.4850, a gamma shape distribution of 0.7990and kappa = 14.0727 (ti/tv = 5.9376).

Lineage divergence

The phylogenetic analysis of concatenated sequences supported the presence of two well-defined-divergent haplogroups of *M. cephalus* in New Zealand (Figure IV.2a). The average percentage of differences between both haplogroups (referred to as L1 and L2 from now on) based in the linear pairwise comparisons between sequences was 17.2%. After the linearization of the genetic distances only the multiple threshold model was significant and identified 12 variable entities grouped in nine possible clusters (Appendix Table IV.8; Appendix Figure IV.8). Three entities were well supported and showed variation that could confidently be explained by a speciation model under Yule's algorithm (p > 0.95; Figure IV.2; Appendix Figure IV.8). The divergence time under the multiple threshold approach of the GMYC model identified the three events at ~50,000, ~27,000 and ~16,000 years ago (Appendix Table IV.8; Figure IV.2b).

Genetic structure and diversity

The genetic variation in COX1 was low (h = 0.2342; $\pi = 0.0013$) compared with the Dloop region (h = 0.9723; $\pi = 0.05$; Table IV.1). Cytochrome oxidase 1 haplotypes were in mutation-drift equilibrium (Tajima's D = -1.3002, p > 0.05; Table IV.1), which means that the amount of pairwise differences and segregating sites in the sequences were not significantly different from zero. There were few high frequency haplotypes reflected in the low levels of haplotype diversity (N = 5; Table IV.1). Fu's *Fs* is a measure that uses haplotype diversity; it estimates the probability to find a random sample with a number of haplotypes equal or smaller than the observed value assuming neutrality and the observed haplotype diversity (Fu 1997). The Fu's *Fs* values for COX1 were negative in this study (Fu's *Fs* = -2.5312 p < 0.02; Table IV.1), which is a characteristic signal for bottlenecks or selective sweeps followed by recent population expansion (Galtier et al. 2000). D-loop was non-significant for Tajima's *D* and Fu's *Fs* neutrality assessments (p > 0.05 and p > 0.02 respectively), there was no evidence of population expansion and the positive value of Fu's *Fs* reflecting recent bottleneck was not significant (Fu's *Fs* = 0. 4096 p > 0.02; Table IV.1). There were only five distinct haplotypes in COX1 sequences. The median-joining network showed a star pattern consistent with recent population expansion after the effect of selective sweeps or a bottleneck (Figure IV.3a). The star-pattern showed one highly frequent haplotype from which the other haplotypes are linked with one to two mutational steps difference (Figure IV.3a). The median-joining network of D-loop sequences showed a split of 113 mutational steps between two haplogroups that matched L1 and L2 from the phylogenetic analysis (Figure IV.2; Figure IV.3a).

;; S= segregating	ary event; $\vartheta_1 = the$	
 n = number of sample 	ion before an evolutio	Vot applicable.
uences in New Zealanu	of the effective populat	cant (p < 0.05). N/A = I
and D-loop region seq	liversity;	values that are signifi
om COX1 partial gene	rsity; π = nucleotide d	ı event. In red are the
s of the sequences fro	es; h = haplotype dive	after the evolutionary
Basic genetic statistic	= Number of haplotyp	effective population
Table 0.1	sites; Nh =	size of the

							COX1							
Region	Locality	Latitude	Longitude	u	S	μŊ	Ч	π	Tajima's D	p-value	Fu's Fs	p-value	θ_0	θ_1
	Bay of Plenty	-37.945267	177.003155	37	7	ю	0.4193	0.0024	-0.1329	0.1	0.7128	0.1	0	2.4501
East	Hauraki Gulf	-37.007437	-175.591492	35	4	4	0.2655	0.0017	-1.1136	0.1	-0.9425	0.1	0.00176	3.1106
	Bay of Islands	-35.237716	174.099419	34	ŝ	4	0.1711	0.0009	-0.4267	0.1	-0.6746	0.1	0	6666
	Ninety Mile Beach	-35.039854	173.164722	5	2	7	0.6	0.003	1.6859	0.1	1.7337	0.02	0.00176	8.6743
	Hokianga	-35.407473	173.432797	29	1	7	0.1921	0.001	-1.0697	0.1	-0.1383	0.1	0	1.8775
	Dargaville	-35.981893	173.757307	10	٢	б	0.3777	0.0021	-1.8391	0.05	-2.317	0.02	0.4517	6666
West	Pouto	-36.301823	174.083383	10	9	ю	0.6	0.0001	0.8044	0.1	0.8775	0.1	0	2.3719
	Kaipara Harbour	-36.341742	174.22599	42	7	ю	0.2195	0.0011	-1.4456	0.1	-2.8805	0.02	0	4.9456
	Waikato	-37.557986	175.124493	11	-	7	0.1818	0.0009	-0.4451	0.1	0.4974	0.1	0	252.5
	Porirua	-41.097363	174.900105	12	-	2	0.303	0.0016	0.3328	0.1	0.7855	0.1	0.00176	3.3575
17	Lower Hutt	-41.233764	174.899579	11	-	7	0.3272	0.0017	0.511	0.1	0.8595	0.1	0	3.6334
South	South Island	-41.243534	174.040317	5	0	-	0	0	N/A	N/A	N/A	N/A	0	0
	Tot	al		241	S	5	0.2342	0.0013	-1.3002	0.1	-2.5312	0.02	0.038	16690.0767
							D-loop							
Region	Locality	Latitude	Longitude	u	S	μŊ	μ	μ	Tajima's D	p-value	Fu's Fs	p-value	θ_0	θ_1
	Bay of Plenty	-37.945267	177.003155	5	8	4	0.9	0.004	-0.5264	0.1	-0.5533	0.1	0	52.5781
East	Bay of Islands	-35.237716	174.099419	10	136	7	0.9111	0.0371	-1.7998	0.05	-2.2492	0.02	0	18.698
	Ninety Mile Beach	-35.039854	173.164722	4	132	4	-	0.1009	1.9457	0.1	2.1313	0.1	0	6666
	Hokianga	-35.407473	173.432797	10	20	×	0.9555	0.0101	0.9732	0.1	0.9199	0.1	0.00176	22.932
	Dargaville	-35.981893	173.757307	10	134	6	0.9777	0.0385	-1.6277	0.1	-2.1414	0.1	0.00352	26.2139
West	Pouto	-36.301823	174.083383	6	136	6		0.0633	0.20007	0.1	1.1522	0.1	0	26.389
	Kaipara Harbour	-36.341742	174.22599	10	137	6	0.9777	0.0817	1.9724	0.1	1.6601	0.02	11.539	20.914
	Waikato	-37.557986	175.124493	10	135	6	0.9777	0.0371	-1.7744	0.05	-2.2907	0.02	2.9777	22.666
	Porirua	-41.097363	174.900105	10	136	~	0.9333	0.0722	0.6716	0.1	1.4935	0.1	0	523.425
Couth	Lower Hutt	-41.233764	174.899579	11	139	11	-	0.0687	0.9558	0.1	1.371	0.1	17.956	37.5658
IIIII	South Island	-41.243534	174.040317	5	33	4	0.9	0.0175	-0.4919	0.1	-0.5307	0.1	0	56.9527
	Tot	al		94	176	50	0.9723	0.05106	0.6716	0.1	0.4096	0.1	50.28	1000





Haplogroup L1 showed a star-like pattern with eight haplotypes. There was one highly frequent haplotype and few low frequency and highly divergent haplotypes (10-13 mutational steps; Figure IV.3b). Haplogroup L2 showed 42 low frequency haplotypes with several highly divergent haplotypes characteristic of populations in constant population size over long periods of time (Figure IV.3b). The highly frequent haplotype from COX1 sequences was found in all sampled locations; two haplotypes were restricted to the East Coast (haplotypes two and five); and one haplotype was restricted to norther areas from Whakatane in East Coast to Hauraki Gulf in the West (haplotype four; Figure IV.4a). For D-loop sequences, haplotypes from Haplogroup L1were distributed across the North Island (from Wellington to Ninety Mile Beach (upper west coast) and the Bay of Islands (upper east coast), and were not found in Whakatane (Figure IV.4b). Haplotypes from haplogroup L2 were spread from the upper South Island Ninety Mile Beach, Bay of Islands and Bay of Plenty (Figure IV.4b). There was an overlap in the geographic distribution of both mitochondrial haplogroups; however, haplogroups L2was more diverse and was the predominant haplogroup in New Zealand (Figure IV.4b). The analysis of molecular variances (AMOVA) showed no significant differences among geographic regions and among locations, within locations and within locations respective to the total variation (Table IV.2). There were highly significant differences within and between haplogroups (p = 0.0001); however, there were no significant differences within haplogroups averaged to the total variation (p > 0.05; Table IV.2). The assessment of genetic differentiation with ϕ_{ST} values showed no significant differences among locations indicating high levels of gene flow with the exception of Hokianga and Porirua that showed significant differences (p < p0.01; Table IV.3).

Mitochondrial DNA and microsatellite DNA structure comparison

The assignment analysis of the expected mtDNA haplotype frequencies under the Rannala and Mountain (1997) algorithm showed that only 56.3% of the individuals were assigned to the mtDNA haplogroup of origin, L1 or L2 (Table IV.9). The individuals that were

clustered together in the allele frequencies of msatDNA markers analyses (Figure II.6), did not match how the individuals were clustered based on mtDNA variation (Table IV.9). The assignment probabilities were low in comparison with the cluster probabilities from STRUCTURE, reflecting a different genetic structure between msatDNA markers and mtDNA markers.

Demographic history

Mismatch distributions showed a uniform distribution for COX1 haplotype frequencies which was in accordance with the neutrality test and population expansion (Table IV.1; Figure IV.5). D-loop haplotype frequencies showed a multimodal distribution. A multimodal distribution of haplotype frequencies means a strong genetic structure within locations. It is the characteristic pattern of secondary contacts between isolated populations or highly divergent populations following a sympatric speciation pattern (Avise 2000). The patterns shown in plotting the pairwise differences for the SSD and the Raggedness Index were not significantly different from the constant population size model (Figure IV.5). Analysing deviations of neutrality by sampling location there were few exceptions to neutrality (Table IV.1), with only the Kaipara Harbour showing recent population expansion signatures in both mtDNA markers (Table IV.1). The BSPs showed constant population size followed by recent population expansion (Appendix Figure IV.11). The time of the recent events of population expansion did not match the time of divergence found with the GMYC model (Figure IV.2; Appendix Figure IV.11). The BSP of COX1 sequences showed a bottleneck followed by population expansion ~1, 500 years ago; the BSP of D-loop sequences showed constant population size just before 15, 000 years ago where they showed a population expansion (Appendix Figure IV.11). The concatenated sequences showed a similar pattern as D-loop; however, the population expansion happened more recently (before 10,000 years ago; Appendix Figure IV.11). Haplogroup L1 showed a slow population expansion followed by a sudden increase in the effective population size between 1,000 and 1,500 years ago 9 Appendix Figure IV.11). Haplogroup L2 showed a

similar pattern as D-loop in which the population expansion occurred just before 5, 000 years ago (Appendix Figure IV.11).

Bayesian skyline plots are not sensitive to population subdivision because they average the demographic histories of the population sub-groups, making them not reliable when each population sub-group has a different demographic history (Hey and Nielsen 2004, 2007; Marko et al. 2010). In this particular study BSPs are not the most suitable analysis to describe the demographic history of *M. cephalus*.

Table IV.2 Analysis of Molecular Variance among sampling locations, geographic regions and haplogroups L1 and L2 in New Zealand. φ_{SC} = variation among localities and between haplogroups in the case of L1 and L2 comparisons; φ_{ST} = variation within localities and within haplogroups; φ_{CT} = variation within localities respective to the total variation and within haplogroups respective to the total variation in the case of L1 and L2 comparisons. Significant values are shown in red.

	COX1		D-loop		Haplogroups L1 a	nd L2 comparisor	1
Variation type	Population differentiation statistic	p-value	Population differentiation statistic	p-value	Variation type	Population differentiation statistic	p-value
Among localities	$\varphi_{SC} = -0.02056$	0.9	$\varphi_{SC} = 0.04543$	0.15	Between haplogroups	$\varphi_{SC} = 0.18059$	0.0001
Within localities	$\varphi_{ST} = -0.01964$	0.8	$\varphi_{ST}=0.4231$	0.17	Within haplogroups	$\varphi_{ST} = 0.86717$	0.0001
Within localities respective to the total variation	$\varphi_{CT} = 0.00091$	0.8	$\varphi_{CT} = -0.00327$	0.42	Within haplogroups respective to the total variation	$\varphi_{CT} = 0.83789$	0.5

Migration and connectivity patterns

Approximately 51% of the haplotypes were successfully assigned to the location of origin. There were elevated average migration ratios (Table IV.4; Appendix Table IV.10). The connectivity patterns showed that the localities with higher volume of immigrants are Bay of Islands in the eastern region, the South Island in the southern region and Waikato and 90 Mile Beach in the western region (Table IV.4; Figure IV.6). Kaipara Harbour, Pouto and Dargaville are the locations with higher volume of emigrants (Table IV.4; Figure IV.6). Hokianga and Porirua showed levels of significant differentiation ($\phi_{ST} = 0.025$; p < 0.05; Table IV.3) and extensive gene flow was shown among locations in all the other pairwise comparisons of ϕ_{ST} (Table IV.3).

a) COX1



b) D-loop



Figure IV.3 a) Median-Joining network of the five haplotypes found with COX1 in New Zealand. b) Median-Joining network of the 50 haplotypes found with D-loop in New Zealand.

	Bay of Plenty	Bay of Islands	90 Mile Beach	Hokianga	Dargaville	Pouto	Kaipara Harbour	Waikato	Porirua	Lower Hutt	South Island
Bay of Plenty	*										
Bay of Islands	0.47213+-0.0051	*									
90 Mile Beach	0.12712+-0.0033	0.06603+-0.0028	*								
Hokianga	0.07989+-0.0024	0.32898+-0.0048	0.05831+-0.0021	×							
Dargaville	0.35442+-0.0052	0.62865+-0.0051	0.07762+-0.0026	0.48886+-0.0056	*						
Pouto	0.23701+-0.0038	0.56628+-0.0047	0.34106+-0.0051	0.12494+-0.0035	0.40689+-0.0048	*					
Kaipara Harbour	0.20790+-0.0038	0.27067+-0.0052	0.61707+-0.0049	0.08187+-0.0024	0.24146+-0.0039	0.62360+-0.0052	*				
Waikato	0.26136+-0.0047	0.51074+-0.0049	0.06920+-0.0023	0.15682+-0.0037	0.91733+-0.0030	0.26552+-0.0044	0.10098+-0.0028	*			
Porirua	0.23087+-0.0045	0.30532+-0.0043	0.28334+-0.0041	0.02584+-0.0016	0.17008+-0.0034	0.81269+-0.0037	0.62083+-0.0056	0.10385+-0.0031	*		
Lower Hutt	0.38095+-0.0047	0.43431+-0.0052	0.29255+-0.0046	0.09811+-0.0026	0.32027+-0.0049	0.82705+-0.0035	0.49401 + -0.0049	0.18998+-0.0043	0.99089+-0.0010	*	
South Island	0.34957+-0.0048	0.71636+-0.0041	0.16048+-0.0036	0.28878+-0.0042	0.54410+-0.0044	0.45936+-0.0045	0.22166+-0.0041	0.39541+-0.0048	0.29779+-0.0044	0.41283+-0.0053	*

Table 0.2 Population genetic fixation indices per location (φ_{ST} +/- s.d.). Significant values (p < 0.05) are in red.



Figure IV.4 Map of the distribution of haplotype frequencies in New Zealand. a) Frequency distributions of haplotypes from COX1. b) Frequency distribution of D-loop haplotypes clustered in haplogroups L1 and L2.



Figure IV.4 Cont.

Table IV.4 Migration rates based on simulated effective population sizes and summing all possible genealogies over all possible branch lengths following the coalescence approach by (Kingman 2000), *as set up in Migrate-n* (Beerli and Felsenstein 1999).

	90 Mile Beach	Bay of Islands	Dargaville	Hokianga	Lower Hutt	Kaipara Harbour	Porirua	Pouto	South Isalnd
90 Mile Beach	249.25	547.51	463.18	423.05	483.78	406.2	533.11	340.45	471.03
Bay of Islands	420.91	340.5	470.76	488.37	473.96	481.8	519.64	464.58	577.98
Dargaville	618.88	549.83	850.5	501.7	558.83	560.7	563.6	604.42	639.12
Hokianga	612.76	529.77	510.38	665.5	514.55	484.9	550.37	644.11	653.26
Lower Hutt	520.04	558.53	478.62	500.32	1374.25	601.7	449.25	502.1	520.9
Kaipara Harbour	618.02	527.96	698.37	588.76	432.62	1084	520.28	514.44	576.59
Porirua	509.58	591.58	390.96	565.12	432.9	503.8	739.25	461.19	656.12
Pouto	495.78	622.76	535.26	435.15	578.54	592.2	500.38	958.75	531.76
South Isalnd	657.54	505.16	470.65	458.19	485.34	546.5	574.77	616.83	353
Waikato	617.12	599.25	543.21	585.66	532.88	500.7	469.46	449.39	482.92
Whakatane	445.6	599.73	584.27	513.1	384.52	437.5	463.29	448.58	450.84







Figure IV.6 Contemporary mtDNA connectivity patterns of New Zealand <u>M. cephalus.</u> Only the flows that are representing a migration volume of more than 51% are shown. The colour pattern follows regional distribution: Blue = South Island and southern North Island; orange/yellow = west coast North Island; green = east coast North Island. The outer circle segments below the locality name represent the total number of migrants moving to and from the locality. The inner bars show the total volume of migration encoded in the circle segments. To define directionality, in each location emigrants are represented with the same colour of the location and immigrants are coloured as the locality of origin. The width of the curved line is the volume of movements where the exact volume of movements can be visualized only at the beginning and end points. The width of the curved line depends on number of migrants relative to the total volume of migrants across all regions indicated by the outside circle segments.

Pacific Rim

A total of 111 COX1 published sequences downloaded from GenBank, 10 sequences amplified from Australian samples and 30 sequences collected from Chile were compared with those from New Zealand. Additionally, 130 D-loop published sequences downloaded from GenBank, were also used (Table IV.5). Six lineages for COX1 and four lineages for D-loop were identified (Table IV.5). New Zealand shared two COX1 haplotypes with Australia; however, New Zealand's D-loop lineages were restricted to New Zealand in geographic distribution (Table IV.5; Figure IV.7; Appendix Figure IV.10). Chile shows a similar pattern where no COX1 haplotypes are shared with other locations of the Pacific Rim (Table IV.5; Figure IV.7; Appendix Figure IV.9). New Zealand showed the lowest percentage of differences (2.7%) with a haplogroup from the Pacific distributed from Russia to Australia based on a *NJ* tree under GTR+G+I evolutionary model of COX1 sequences (NWP2; Appendix Figure IV.9). The *NJ* tree of D-loop sequences under the same substitution model showed the lowest percentage of differences (22%) between haplogroup L2 and a the same Pacific haplogroup as COX1 (NWP2, Appendix Figure IV.10).

Table IV.5 Basic genetic statistics of the haplotype lineages/groups from COX1 partial gene and D-loop region sequences in the Pacific Rim. n= number of samples; S= segregating sites; Nh = Number of haplotypes; h = haplotype diversity; π = nucleotide diversity; θ_0 = the size of the effective population before an evolutionary event; θ_1 the size of the effective population after the evolutionary event. In red are the values that are significant (p < 0.05). N/A = Not applicable. NWP2 = North-West Pacific 2 lineage identified in Shen et al. (2011).

						COX1					
Haplogroup/lineage	n	S	Nh	h	π	Tajima's D	p-value	Fu's Fs	p-value	θ0	θ_1
L1	55	104	11	0.3622	0.01388	-2.5031	0.0001	-25.0597	0	0	0.3897
L2	196	134	26	0.4896	0.00909	-2.1747	0.0003	-25.3634	0	0	9.6023
NWP2	23	16	12	0.8498	0.00354	-2.1201	0.0049	-27.3341	0	0	99999
NWP1	29	30	17	0.8079	0.0045	-2.5596	0.0002	-26.7871	0	0	99999
NWP3	13	3	5	0.6923	0.00189	-0.5895	0.468	-23.0717	0	0	99999
Chile	36	19	6	0.3937	0.00562	-1.2862	0.086	-26.4085	0	0	0.5677
Total (Average)	352	51	13	0.5992	0.01325	-1.7837	0.094	-25.6781	0	0	50001.25
						D-loop					
Haplogroup/lineage	n	S	Nh	h	π	Tajima's D	p-value	Fu's Fs	p-value	θ₀	θ_1
L1	17	6	6	0.515	0.00114	-1.63094	0.08	-2.848	0.042	0.007	1.0989
L2	78	64	29	0.919	0.01054	-1.43594	0.04	-5.142	0.003	0.9	11.367
NWP2	47	143	37	0.981	0.03081	-0.90775	0.18	-7.442	0.1	0.0035	55.643
Pacific IV	82	158	59	0.969	0.00739	-2.47747	0	-64.869	0	0	99999
Total (Average)	224	338	123	0.976	0.1462	-1.6873	0.06	-0.425	0.054	0.2276	25017

Pacific Rim lineages based on COXI haplotypes showed low to mid values of nucleotide diversity ($\pi = 0.0018$ -0.013; Table IV.5), and low values of haplotype diversity with the exception of the lineages NWP2 and NWP1 (h = 0.849 and h = 0.807, respectively; Table

IV.5). The haplogroups identified with D-loop sequences showed low values of nucleotide diversity ($\pi = 0.001-0.03$; Table IV.5), and high values of haplotype diversity (h = 0.91-0.98; Table IV.5), with the exception of haplogroup L1 (h = 0.51; Table IV.5). Lineages NWP1, NWP2 and NWP3 have been identified in previous studies as highly divergent lineages that could be cryptic species of *M. cephalus* that require taxonomic revision (Shen et al. 2011; Durand et al. 2012b). The three lineages NWP1, NWP2 and NWP3 were initially identified in China Sea waters (Shen et al. 2011). One study tried to resolve the phylogenetic relationships of the mullets including *M. cephalus* based on partial COX1, Cyt *b* and 16s mtDNA sequences (Durand et al. 2012b). The phylogenetic analysis showed that lineages NWP1 and NWP3 were monophyletic and restricted to China Sea locations. The lineage NWP2 was paraphyletic with sequences of *M. cephalus* individuals sampled in locations from Russia to Australia (Durand et al. 2012b). New Zealand haplogroups L1 and L2 were grouped with NWP2 in the ML phylogenetic analysis and formed a paraphyletic group (Appendix Figure IV.12). This same group has been seen in previous phylogenetic analysis in which Australian and New Zealand samples from Patea formed a paraphyletic group with NWP2, samples that were also used in this study as part of the sequences downloaded from GenBank (Durand et al. 2012b).







Figure IV.7 Cont.

Discussion

Two genetically distinct haplogroups of *M. cephalus* were identified in New Zealand (L1 and L2). There was no correlation between the genetic distances and the geographic distances, nor a genetic partition among geographic regions. High levels of gene flow were found among the sampled locations and both haplogroups appeared to be in mutation-drift equilibrium. The average level of genetic differences between the concatenated sequences of L1 and L2 haplogroups (17.2%) was higher than the typical ~15% of interspecific variation previously reported in phylogenetic assessments with 16s rRNA, COX1 and Cytochrome *b* mtDNA markers, among closely related species within the genus *Mugil* (Durand et al. 2012b).

GMYC model more than species delimitation

The GMYC model has been used mainly for species delimitation because it identifies the threshold when the genetic variation among individuals can be explained by Yule's speciation model (Monaghan et al. 2009; Pontin and Cruickshank 2012; Fujisawa and Barraclough 2013; Talavera et al. 2013). Morphological characters are then used to support the divergent groups as species candidates (Talavera et al. 2013). Without the support of morphological characters, the divergent entities that have genetic variation that could be explained by Yule's speciation process would remain as highly divergent lineages and possible candidates for new species (Talavera et al. 2013). However, the GMYC model is not sensitive to the effects of bottlenecks or selective sweeps, migration and founder effects that could also result in highly divergent lineages and possible candidates for new species. The strength of the GMYC model is that can accurately estimate the time to the MRCA with little information to calibrate the nodes using a phylogenetic and coalescence approach (Kingman 2000; Steel and McKenzie 2001; Rosenberg and Nordborg 2002; Gernhard et al. 2008).

The phylogenetic and the coalescence approach when run independently need node calibration to accurately estimate the nodes to the MRCA. To avoid this constraint, the GMYC approach allows to approximately estimate the time when an ancestral generation with a given

mutational rate equals the probability of speciation in which the GMYC model uses unique haplotypes, and assumes neutrality where the variation between sequences is caused only by mutation (Monaghan et al. 2009; Pontin and Cruickshank 2012; Fujisawa and Barraclough 2013; Talavera et al. 2013). Species that show evidence of genetic drift or selection will violate this assumption in which the divergence time generated by the GMYC model reflects the time of bottlenecks and population expansion events (Pons et al. 2006, 2011; Monaghan et al. 2009; Papadopoulou et al. 2011; Esselstyn et al. 2012; Falniowski and Szarowska 2012; Parnmen et al. 2012; Pontin and Cruickshank 2012; Kekkonen and Hebert 2014). In his study, both mtDNA markers showed evidence of bottlenecks followed by rapid population expansion. A possible explanation to the high levels of variation is the founder effect of two successful colonisations of New Zealand by M. cephalus individuals from allopatric populations from the Pacific (see below). For *M. cephalus* in New Zealand without any other evidence to support the high levels of mtDNA variation, the GMYC model rather than identifying cryptic species or species candidates, identified the time since genetic drift has been acting after the arrival of M. cephalus migrants from Pacific allopatric populations where the time threshold showed the most recent contact between founders and residents rather than the split of the allopatric populations. The first *M. cephalus* founders we could identify arrived to New Zealand not before ~50,000 years ago and the contact of the second wave of founders giving raise to haplogroups L1 and L2 happened between 20, 000 and 16, 000 years ago.

Population connectivity and gene flow

In New Zealand, population genetic and phylogeographic studies based on mtDNA markers have reported that the levels of genetic differentiation in populations from coastal species, are correlated to geographical breaks such as East to West, Isolation by Distance patterns (IBD) and latitudinal gradients (see Gardner et al. 2010 for a review). *Mugil cephalus* in New Zealand showed two genetically distinct haplogroups in sympatry and extensive gene

flow among locations where approximately 49% of the individuals sampled in this study were identified as migrants.

The historical analysis of the *M. cephalus* fishery in Kaipara Harbour suggested the presence of two types of *M. cephalus*. One, the "clean-gut" *M. cephalus* that migrates daily in and out the estuary and feeds in coastal environments, and a second M. cephalus referred to as "settler" or "muddy fish" that is resident to the estuary and may migrate towards freshwater systems (Paulin and Paul 2006). Recently, the analysis of Local Ecological Knowledge (LEK) of the Kaipara Harbour area reported two different kinds of M. cephalus which may be distinguishable by similar migratory behaviour as reported from historical information (Paulin and Paul 2006; Morrison et al. 2014a). In this study, there was no genetic differentiation among locations regardless if the fish was sampled in a coastal, estuarine or freshwater locations. The only significant differences were found between Hokianga and Porirua (two estuarine locations); however, this study is limited in finding the cause for the observed levels of differentiation. One factor causing significant differences between Hokianga and Porirua could be the environmental conditions in which coastal current patterns may limit fish migrations between the two locations; also, different environmental conditions may result in a selective pressure limiting the distribution extent of *M. cephalus* (*i.e* Blel et al. 2010). Whether the cause, more research is needed to test those two environmental hypothesis. It is common for M. cephalus to undergo daily migrations from coastal waters into estuaries, rivers and lakes (see Whitfield et al. 2012 for a review). The tidal currents that flood New Zealand's estuaries allow *M. cephalus* and other marine species to reach shallow areas of high productivity to feed or shelter from predators (McDowall 1976; Francis et al. 2011). Schools of M. cephalus have been observed in open coastal areas in New Zealand including exposed surf beaches and coastal rocky reef systems (Morrison et al. 2014a, 2016a).

New Zealand is not the only area where schools of *M. cephalus* can migrate long distances. Nuclear DNA markers were used to identify a migratory population of *M. cephalus* that moves between China and Taiwan during the spawning season. Moreover, a tagging study in North Carolina, USA, showed that less than 10% of *M. cephalus* individuals migrated ~860 km in a two-month period to reproduce (Huang et al. 2001; Bacheler et al. 2005; Ke et al. 2009). It is possible that schools of *M. cephalus* have short migratory movements in and out of estuaries followed by intermittent longer migrations along the coast.

Estuaries are used as nursery areas by range of marine species and it has been observed that species richness increases towards the muddier upper reaches of New Zealand estuaries (Francis et al. 2011). Common New Zealand species found in such areas are: *M. cephalus*, the kahawai *Arripis trutta* and the Snapper *Pagrus auratus* which support important commercially and customarily fisheries (McDowall 1976; Watson et al. 2005; Francis et al. 2011). The role of estuaries as nursery areas for *M. cephalus*, its short migrations in and out of estuaries and the long coastal migratory patterns of this species, could be responsible for the observed levels of gene flow among locations of New Zealand's North Island. It is not known whether *M. cephalus* individuals moving among estuaries may reproduce in spawning grounds of different regions or they are strictly associated with a particular spawning area.

In Chinese waters, the genetic diversity and genetic distances among *M. cephalus* mtDNA lineages were a consequence of an East China Sea and South China Sea brake assessed with COX1 and D-loop sequences (Jamandre et al. 2009; Liu et al. 2009a, 2010; Sun et al. 2012). The Cytb phylogeographic assessment of *M. cephalus* in the West Pacific identified three populations in which seasonal coastal currents flowing through the marginal seas among China, Japan and Taiwan might facilitate the dispersal of *M. cephalus* larvae, allowing secondary contacts and range expansion after periods of isolation during glacial ages (Ke et al. 2009). Similar pattern was found with COX1 sequences that showed *M. cephalus* undergoing

several events of isolation due to glaciation and sea-level change followed by postglacial secondary contacts in the China Sea, resulting in three highly divergent lineages supported with msatDNA markers isolated since 200, 000 years ago. Intermittent migrations of *M. cephalus* due to changes in coastal and oceanic currents have allowed bouts of individuals from different Pacific populations to reach and colonise New Zealand at least twice in the las 50,000 years.

Lineage divergence and New Zealand colonisation

The genetic variation between the haplogroups L1 and L2 was high and similar to the levels of variation found among species of the genus *Mugil* (Durand et al. 2012b). Some authors have suggested that *M. cephalus* is a species complex because of the high levels of mtDNA genetic variation in *M. cephalus* lineages and the cosmopolitan distribution of the species (Jamandre et al. 2009, 2014; Shen et al. 2011; Durand et al. 2012a,b; Whitfield et al. 2012). Highly divergent lineages from different geographic origins have been found with Cyt*b*, 16s, COX1 and D-loop sequences (Crosetti et al. 1993; Rocha-Olivares et al. 2000; Rossi et al. 2004; Fraga et al. 2007; Heras et al. 2007; Livi et al. 2011). However, it has been almost impossible to test for genetic incompatibility using mtDNA because it is a non-recombining genome and uniparentally inherited.

The number of species or cryptic entities could be overestimated as a result of low female dispersal, when male dispersal is the primary mechanism for allelic exchange between populations (Avise 2000; Irwin 2002; Hickerson et al. 2010). While it is possible that L1 and L2 in New Zealand are two cryptic species in sympatry, there was no additional evidence to support this hypothesis. High levels of gene flow were identified in this study and admixture of four populations was found with microsatellite DNA (msatDNA) markers (Chapter II). Without collaborating evidence of genetic isolation from other sources, such as nuclear msatDNA markers, niche specialization or distinctive morphological characters, the two-species hypothesis has little support. The variation found in mtDNA markers is then evidence of the effect lineage sorting and genetic drift rather than isolation.

One study has reported three genetically distinct populations of *M. cephalus* in sympatry (NWP1, NWP2 and NWP3) in the China Sea using data from msatDNA and partial sequences from the mtDNA COX1 gene (Shen et al. 2011). The three groups identified phylogenetically based on COX1 partial sequences, were supported by cluster probabilities of ancestry based on allele frequency variation of nuclear msatDNA markers (Shen et al. 2011). The isolation of the three populations was the result of the separation of the Japan and China Seas when the sea level dropped during a glaciation 200,000 years ago (Shen et al. 2011). No study has reported or tested for reproductive incompatibility, niche specialization or found distinctive morphological traits to support cryptic species within *M. cephalus*. The support for cryptic speciation within *M. cephalus* has only been tested phylogenetically with three mtDNA markers (16s rRNA, COX1 and Cytb), and using msatDNA marker ancestries under the non-admixture model that support three of the mtDNA groups found (Shen et al. 2011; Durand et al. 2012a; Whitfield et al. 2012).

The most plausible explanation for the level of genetic differences between L1 and L2 and their overlapping distribution is that New Zealand was colonised by two allopatric *M. cephalus* populations from the Pacific. The founder individuals from the allopatric populations that come into contact will retain high levels of mtDNA differentiation but nuclear genetic diversity will be homogenised through recombination (Driscoll 1998; Grosberg and Cunningham 2001; Marko and Hart 2012). The lack of recombination of mtDNA allows it to be used to identify where individuals from previous allopatric populations now overlap (Zecca et al. 2011; Polgar et al. 2014). Distinct mtDNA lineages in an area that has been colonised more than once will remain until there is complete lineage sorting by genetic drift or selection, which, depending of the effective population size, could be for a long period of time (Avise 2004; Zecca et al. 2011). In New Zealand, no transition zones between L1 and L2 were identified because *M. cephalus* mitochondrial haplogroups showed high levels of gene flow

among the sampled locations, suggesting that there has not been enough time for genetic drift to completely erode the signature contact of the two colonisation events. There is a phylogeographic pattern of mitochondrial genetic variation and geographic distribution of cryptic lineages in *M. cephalus* of the Northwest Pacific, which has been explained in a similar way (Ke et al. 2009; Shen et al. 2011). The higher sea levels during the inter-glacial period of the Plio-Pleistocene (~5mya to ~12,000 years) has allowed intermittent contact among *M. cephalus* populations between the Japan and China 200,000 years ago.

The pattern of multiple colonisation events from genetically distinct source populations has been reported for the breeding populations of the brook charr Salvelinus fontinalis in Mistassini Lake, Canada. Distinct haplogroups that originated from different ancestral populations have an overlapping geographic distribution (Fraser and Bernatchez 2005). The population overlap occurs in feeding areas, with seasonal migrations for breeding into the rivers connected to the lake. Some populations are anadromous (migration from marine to freshwater environments to spawn), others migrate from the lake to in-flow rivers, and a few move to outflow rivers to spawn. Cichlids in the species complex Amphilophus spp. from the Nicaraguan crater lakes are comprised of distinct genetic units (Barluenga and Meyer 2010). The large Nicaragua and Managua lakes have acted as source populations for the small crater lakes in the surrounding areas, and have the most genetically diverse populations. Founder haplogroups diversified rapidly after each of the colonisation events. The genetically distinct populations found in each crater lake were the result of the effect of genetic drift and the number of founder haplotypes retained since the time of the colonisation event (between 20,000 and 5,400 years ago). Some of the haplotypes in the larger lakes source populations were still present, at a low frequency, in the colonised crater lakes (Barluenga and Meyer 2010; Elmer et al. 2014). The contemporary haplotype frequencies in the New Zealand M. cephalus population would have changed since the original founder event due to the effects of genetic drift (Hewitt 1999, 2000).

New mutations have been accumulated and without the effect of subsequent successful colonisation events, *M. cephalus* in New Zealand seems to be comprised of haplotypes not found in other areas of the Pacific with the exception of two COX1 haplotypes from Australia.

Migration is successful when a reproducing group becomes established in the colonised environment (Buckley et al. 2013; Waters et al. 2013a). It is possible that there were more than two migration events of M. cephalus to New Zealand but these others did not establish successfully or they integrated into the local population and their mtDNA has disappeared through genetic drift. Density-dependant processes such as density blocking and competitive exclusion by the existing local population could have prevented more recent migrants from establishing (Waters et al. 2013b). The founder individuals will exclude the effects of subsequent migration events, where secondary dispersers fail to settle due to limited spaces to occupy, and migrants will fail to coexist due to competition (Ostfeld and Canham 1995; Tomaiuolo et al. 2007; Waters et al. 2013b). The main factors that may affect the successful establishment of new populations could be grouped into three types. First, density-blocking and competitive exclusion processes; secondly, genetic incompatibility and thirdly, the time required for genetic drift to change the haplotype frequencies in mtDNA, or for allele frequencies to be homogenized in the case of bi-parentally inherited markers (Hallatschek and Nelson 2008; Hedgecock and Pudovkin 2011; Buckley et al. 2013; Waters et al. 2013b). Reproductive behaviour differences between migrants and residents could lead to a reproductive barrier and subsequently strong effects of genetic drift due to small effective population size in migrants. For example, the Atlantic herring *Clupea harengus* in the Baltic Sea is composed of two seemly sympatric populations. The reproductive barrier between the two populations is a shift in the spawning season. One population spawns in summer and one during winter. The change is a result of several colonisation events of an extant population from the Atlantic adapted to spawn during winter, and the heritability of this behaviour (Bekkevold et al. 2005, 2007; Jørgensen et al. 2005).

Mugil cephalus in New Zealand compared to the Pacific

The L1 haplogroup of New Zealand's *M. cephalus* shared two haplotypes with the east coast of Australia. A more comprehensive sampling between Australia and New Zealand could reveal one of two scenarios: either a single haplogroup that extends from eastern Australia to the east coast of New Zealand, or could reveal that haplotypes from Australia could have colonised New Zealand. Since then, New Zealand and Australian populations have subsequently diverged by new mutations arising and there has not been enough time for genetic drift to remove all the founder haplotypes, similar to the Nicaraguan cichlids (Hewitt 1999, 2000; Barluenga and Meyer 2010; Elmer et al. 2014). In either case, Australia is likely to be the source population of New Zealand's L1 haplogroup. More extensive sampling effort is required to test the hypotheses of a single Australian-New Zealand haplogroup or two allopatric populations. The haplotypes from the L2 haplogroup were found only in New Zealand and are more related to a haplogroup distributed in the West Pacific from Russia to Australia (NWP2; Shen et al. 2011). The presence of L1 and L2 in New Zealand could be described as two independent migration events from two different allopatrically-isolated populations of M. cephalus from the North-West Pacific to New Zealand. One migration event followed a route from Australia and one that is likely to have followed a different route from a Pacific population. Haplotypes from the Waikato River mouth and Patea dam, have been previously used for phylogenetic analyses of *M. cephalus* where New Zealand was phylogenetically grouped with Russian and Chinese haplotypes (NWP2; Durand et al. 2012b). Based on the available samples and the locations of the Pacific surveyed in this study, the geographic distribution of L1 seems to be limited to New Zealand and Australia and L2 is just found in New Zealand. The lineage Chile showed haplotypes that are also unique compared to the rest of the Pacific. There were not enough samples surveyed from neighbouring countries to have

a reliable phylogeographic assessment for Chilean waters. It is possible that the geomorphology and variation in sea-surface currents along the South-East Pacific may have structured *M*. *cephalus* populations in a similar way as the Northwest and Southwest-Pacific areas. A hypothesis that remains to be tested.

Conclusions

Two mtDNA distinct haplogroups of *M. cephalus* were identified in New Zealand in sympatry. There was extensive gene flow among the sampled locations where *M. cephalus* undergoes migrations among estuaries which allows L1 and L2 haplogroups to have a wide and overlapping distribution. The genetic variability between L1 and L2 haplogroups was probably the result of two colonisations from Pacific allopatric populations. The first colonisation occurred not before ~50,000 years ago where founder haplotypes for L2 arrived to New Zealand. The second colonisation occurred between 20, 000 and 16, 000 years ago where the founder haplotypes for L1 colonised and established in New Zealand following a route through Australia. The levels of mtDNA variation between both haplogroups using COX1 and D-loop concatenated sequences was a similar level to what has been reported for among *Mugil* species; however, to date there has not been evidence to suggest that the New Zealand haplogroups are reproductively isolated groups or new species. The samples used and locations surveyed for this study showed only two shared haplotypes with Australia; which leads to the conclusion that New Zealand haplogroups of *M. cephalus* are distinct and unique groups of mullets in the Pacific.

Appendix

Table IV.6 <u>Mugil cephalus</u> mitochondrial DNA sequences that were downloaded from GenBank from Cytochrome Oxidase I partial gene and D-loop region to perform the Pacific Rim Analysis. a) Cytochrome Oxidase I; b) D-loop region

a) Cytochrome Oxidase I

Country	Location	Latitude N	Longitude E	GenBank Accession number	Source
				EU200483	Kartavtsev et al. 2007.
				EU392243	Unpublished
Russia	Vladivostok	43.04889	131.690454	EU392244	Semina et al. 2008
				EU392245	Unpublished
				EU392246	
				HM628583	
				HM628584	
				HM628585	
				HM628586	
				HM628587	
				HM628588	
				HM628589	
				HM628590	
	China Sea	39.888388—23.03391	119.879977—116.979587	HM628591	
				HM628592	
China				HM628593	Sun et al. 2012
				HM628594	5un et ul. 2012
				HM628595	
				HM628596	
				HM628597	
				HM628598	
				HM628599	
				HM628600	
				HM628601	
				HM628602	
				HM628603	
				HM628604	

Country	Location	Latitude N	Longitude E	GenBank Accession number	Source	
				HM628605		
				HM628606		
				HM628607		
				HM628608		
				HM628609		
				HM628610		
				HM628611		
				HM628612		
				HM628613		
				HM628614		
				HM628615		
	Shantou	23.344239	116.617813	GU260697	Shen et al. 2011	
	De cel Discer	22 402921	112 740202	Q060553	Durand et al. 2012	
	Pearl River	22.493821	113.740293	GU260692	Shen et al. 2011	
				GU260696	Shen et al. 2011	
				EU595084		
				EU595085		
				JN242565		
				JN242566	Zhang and Hannar 2008	
	South China Sea	21.091	112.553	JN242567	Ling and Hanner 2008.	
				JN242568	Unpublisnea	
				JN242569		
				JN242570		
				JN242571		
Country	Location	Latitude N	Longitude E	GenBank Accession number	Source	
---------	-----------------	------------	-------------	--------------------------	---------------------------	--
				GU260668		
				GU260669		
				GU260670		
				GU260672		
				GU260673		
	Vaaluna	25 1521	121 756021	GU260675	Shap at al 2011	
	Keelung	23.1321	121.750021	GU260677	Shell <i>et al</i> . 2011	
				GU260682		
				GU260686		
				GU260687		
				GU260688		
				GU260693		
	Taipei	25.047167	121.491873	JQ060552	Durand et al. 2012	
	Gaoping Estuary	22.598362		JQ060539		
			120.438544	JQ060540	Durand et al. 2012	
Taiwan				JQ060541		
Taiwaii				GU260664		
				GU260679		
				GU260680		
				GU260681		
	Gaoping River	22 598362	120 / 385//	GU260684	Shen at al. 2011	
		22.570502	120.+303++	GU260685		
				GU260690		
				GU260691		
				GU260694		
				GU260695		
				GU260665		
				GU260666		
	Kaohsiung	22 618267	120 263082	GU260667	Shen et al 2011	
	Ruonstang	22.010207	120.203002	GU260671		
				GU260674		
				GU260676		
Japan	Chiba	35,59372	140.056564	AP002930	Miya <i>et al.</i> 2001	
Japan	Cinua	55.57512	1+0.030304	AP002931	Wilya et ul. 2001	

Country	Location	Latitude N	Longitude E	GenBank Accession number	Source	
				GU260678	Shen et al. 2011	
				HQ149082		
	Yokosuka	35.27987	139.692642	HQ149083		
				JF952795	Zhang and Hanner 2011.	
				JF952796	Unpublished	
	Kagoshima	31.521507	130.536945	JQ060557	Durand et al. 2012	
	Okinawa	26.656335	128.066808	GU260683	Shen et al. 2011	
Philippines	Philippines	18.659959	121.360215	GU260689	Shen et al. 2011	
Fiji	Viti Levu	-17.806488	177.946566	JQ060567	Durand et al. 2012	
				JQ060566		
				JQ060569		
	Kone	-21.064717	164.860926	JQ060570	Durand et al. 2012	
				JQ060571		
New Caledonia				JQ072904		
	Noumea	-22.249066	166.447356	JQ060568	Durand et al. 2012	
	Southwestern Lagoon	-22.298296	166.434846	JQ060572	Durand et al. 2012	
	North Pine River	-27.270364	153.000659	HM006970	Page and Hughes, 2010. Unpublished	
Australia	Harvey Estuary	-32.664352	115.669699	JQ060559	Durand et al. 2012	
	Swansea Channel	-33.077986	151.63951	JQ060558	Durand et al. 2012	
Country E Country E Y Y K O Philippines P Fiji V P Fiji V New Caledonia N K N K N N N N N N N N N N N N N	Port Hacking	-34.070263	151.126124	JQ060565	Durand et al. 2012	
Nam Zaalan d	Waikato River Mouth	-37.382204	174.72949	JQ060556	Drama d et al 2012	
New Zealand	Patea Dam	-39.546569	174.566648	JQ060554	Durand <i>et al.</i> 2012	
				JQ060549		
Hawaii	Hawaii	21.294846	-157.877289	JQ060550	Durand et al. 2012	
				JQ060551		
Ecuador	Isla Santa Cruz, Galanagos	-0 754727	-90 312008	JX559532	Durand et al 2012	
Leuador	Isia Santa Ciuz, Galapagos	-0.754727	-90.312000	JX559533		
Peru	Lima	-12.065054	-77.169143	HQ149714	Shen et al. 2011	
Chile	Coast of Ionique	-20 229675	-70 148892	JQ060563	Durand et al 2012	
Child	Coast of Iquique	-20.229073	-70.140092	JQ060564		

b) Control Region

Country	Location	Latitude N	Longitude E	GenBank Accession Number	Source
				EU663629	
				EU663630	
				EU663631	
				EU663632	
				EU663633	
				EU663634	
				EU663635	
				EU663636	
	Oiredae	26.015279	120.391724	EU663637	
				EU663638	Lamon day of al 2000
China				EU663639	
Ciillia	Qiliguao	50.015278		EU663640	Jamandre et al. 2009
				EU663641	
				EU663642	
				EU663643	
				EU663644	
				EU663645	
				EU663646	
				EU663647	
				EU663648	
				EU663649	
				EU663650	

Country	Location	Latitude N	Longitude E	GenBank Accession Number	Source
				EU663651	
				EU663652	
				EU663653	
				EU663654	
				EU663655	
				EU663656	
				EU663657	
				EU663658	
	Pearl River	22.493821	113.740293	EU663659	Jamandre et al. 2009
				EU663660	
				EU663661	
				EU663662	
				EU663663	
				EU663664	
				EU663665	
				EU663666	
				EU663667	
				EU663691	
				EU663692	
				EU663693	
				EU663694	
				EU663695	
				EU663696	
				EU663697	
				EU663698	
т :	17 1	25 1521	101 75 (001	EU663699	L 1 . L 2000
Taiwan	Keelung	25.1521	121.756021	EU663700	Jamandre <i>et al.</i> 2009
				EU663701	
				EU663702	
				EU663703	
				EU663704	
				EU663705	
				EU663706	
				EU663707	
				EU663708	

Country	Location	Latitude N	Longitude E	GenBank Accession Number	Source
				EU663709	
				EU663710	
				EU663711	
				EU663712	
				EU663713	
				EU663714	
				EU663715	
				EU663716	
				EU663717	
				EU663718	
				EU663719	
				EU663720	
				EU663721	
				EU663722	
				EU663723	
				EU663724	
	Kaohsiung	22,618267	120 263082	EU663725	Iamandre <i>et al</i> 2009
	inconstantg	22.010207	120.200002	EU663726	
				EU663727	
				EU663728	
				EU663729	
				EU663730	
				EU663731	
				EU663732	
				EU663733	
				EU663734	
				EU663735	
				EU663736	
				EU663/3/	
				EU003/38	
T	C1 1	25 50272	140 056564	AP002930	Miya <i>et al.</i> 2001
Japan	Cniba	35.59372	140.036364	KW1308340	Snen <i>et al.</i> 2014
				NC_003182	Miya <i>et al</i> . 2001

Pokosuka 35.27987 139.692642 EU663670 EU663671 EU663673 BU663674 EU663673 Pokosuka 35.27987 139.692642 EU663674 Jamandre et al. 2009 EU663675 EU663676 EU663676 EU663677 Pokosuka EU663676 EU663676 EU663677 EU663676 EU663679 EU663679 EU663681 EU663681 EU663683 EU663683 EU663684 EU663682 EU663683 EU663683 EU663682 EU663683 EU663683 EU663683 EU663682 EU663684 EU663683 EU663684 EU663684 EU663685 EU663686 EU663684 EU663684 EU663686 EU663686 EU663684 EU663686 EU663686 EU663686 EU663686 EU663686 EU663687 EU663686 EU663686 EU663686 EU663686 EU663686 EU663686 EU663686 EU663687 EU663688 EU663688 EU663688 EU663688 EU663688 EU663688 EU663	Country	Location	Latitude N	Longitude E	GenBank Accession Number	Source
Pokosuka 35.27987 139.692642 EU663671 EU663673 EU663673 EU663673 EU663673 EU663675 EU663675 EU663676 EU663676 EU663675 EU663676 EU663675 EU663676 EU663676 EU663676 EU663686 EU663686 EU663686 EU663686 EU663686 EU663686 EU663686 EU663686 EU663687 EU663688 EU663686 EU663688 EU663686 EU663686 EU663689 EU663740 EU663740 EU663740 EU663740 EU663740					EU663668	
Yokosuka 35.27987 139.692642 EU663670 EU663674 EU663674 Jamandre et al. 2009 EU663676 EU663674 Jamandre et al. 2009 EU663676 EU663676 EU663676 EU663676 EU663676 EU663676 EU663676 EU663676 EU663676 EU663678 EU663678 EU663686 EU663688 EU663688 EU663688 EU663684 EU663687 EU663685 EU663688 EU663686 EU663687 EU663686 EU663688 EU663686 EU663689 EU663688 EU663689 EU663688 EU663689 EU663688 EU663689 EU663740 EU663740 EU663740					EU663669	
Yokosuka 35.27987 139.692642 EU663671 EU663672 EU663674 Jamandre et al. 2009 EU663675 EU663676 EU663676 EU663677 EU663676 EU663678 EU663678 EU663682 EU663684 EU663684 EU663682 EU663685 EU663684 EU663686 EU663686 EU663689 EU66379 EU663680 EU663689					EU663670	
Yokosuka 35.27987 139.692642 EU663673 Jamandre et al. 2009 EU663675 EU663675 EU663676 EU663676 EU663676 EU663677 EU663677 EU663679 EU663679 EU663678 EU663679 EU663679 EU663684 EU663680 EU663681 EU663685 EU663684 EU663684 EU663686 EU663685 EU663686 EU663686 EU663686 EU663686 EU663689 EU663740 EU663740					EU663671	
Yokosuka 35.27987 139.692642 EU663673 Jamandre et al. 2009 EU663674 Ub63676 EU663676 EU663678 EU663678 EU663678 EU663678 EU663678 EU663680 EU663680 EU663681 EU663681 EU663683 EU663683 EU663683 EU663684 EU663684 EU663685 EU663686 EU663686 EU663686 EU663687 EU663686 EU663688 EU663686 EU663686 EU663687 EU663687 EU663687 EU663687 EU663740 EU663740 EU663740					EU663672	
Okinawa 35.27987 139.692642 EU663674 Jamandre et al. 2009 EU663675 EU663676 EU663676 EU663678 EU663679 EU663680 EU663681 EU663683 EU663684 EU663685 EU663684 EU663685 EU663686 EU663686 EU663684 EU663686 EU663686 EU663684 EU663686 EU663686 EU663686 EU663686 EU663686 EU663687 EU663686 EU663686 EU663687 EU663687 EU663687 EU663689 EU6636740 EU663740 EU663740		Velseeulse			EU663673	
EU663675 EU663676 EU663678 EU663679 EU663680 EU663681 EU663681 EU663681 EU663684 EU6636		Токозика	35.27987	139.692642	EU663674	Jamandre et al. 2009
EU663676 EU663677 EU663679 EU663680 EU663680 EU663681 EU663681 EU663683 EU663683 EU663683 EU663684 EU663684 EU663684 EU663685 EU663687 EU663686 EU663687 EU663687 EU663687 EU663687 EU663687 EU663687 EU663687 EU663740 EU663740					EU663675	
EU663677 EU663678 EU663679 EU663680 EU663680 EU663681 EU663683 EU663684 EU663684 EU663685 EU663685 EU663685 EU663686 EU663686 EU663686 EU663687 EU663689 EU663689 EU663739 EU663740 EU663740					EU663676	
EU663678 EU663679 EU663680 EU663680 EU663681 EU663682 EU663684 EU663684 EU663685 EU663685 EU663685 EU663685 EU663685 EU663685 EU663686 EU663686 EU663686 EU663686 EU663687 EU663689 EU663690 EU663690 EU663740					EU663677	
EU663679 EU663680 EU663680 EU663681 EU663682 EU663683 EU663684 EU663685 EU663685 EU663686 EU663686 EU663686 EU663687 EU663688 EU663689 EU663689 EU663739 EU663739					EU663678	
EU663680 EU663681 EU663682 EU663683 EU663683 EU663684 EU663684 EU663685 EU663686 EU663686 EU663686 EU663686 EU663687 EU663689 EU663689 EU663689 EU663689 EU6636740 EU663740					EU663679	
Okinawa 26.656335 128.066808 EU663681 EU663684 EU663684 EU663686 EU663686 EU663686 EU663686 EU663686 EU663686 EU663687 EU663688 EU663689 EU663689 EU6636739 EU663740 EU663740 EU663741					EU663680	
Okinawa 26.656335 128.066808 EU663682 EU663684 EU663685 Jamandre et al. 2009 EU663686 EU663686 EU663687 EU663689 EU6636739 EU663740					EU663681	
Okinawa 26.656335 128.066808 EU663683 EU663685 Jamandre et al. 2009 EU663686 EU663686 EU663686 EU663687 EU663689 EU663690 EU663740 EU663740			26.656335		EU663682	
Okinawa 26.656335 128.066808 EU663684 EU663685 Jamandre et al. 2009 EU663686 EU663686 EU663686 EU663687 EU663689 EU663689 EU663690 EU663740 EU663740 EU663740				128.066808	EU663683	Jamandre et al. 2009
Okinawa 26.656335 128.066808 EU663685 Jamandre et al. 2009 EU663686 EU663686 EU663687 EU663687 EU663688 EU663688 EU663689 EU663690 EU663739 EU663740 EU663741 EU663741 EU663741					EU663684	
EU663686 EU663687 EU663689 EU663690 EU663739 EU663740 EU663741		Okinawa			EU663685	
EU663687 EU663688 EU663689 EU663690 EU663739 EU663740 EU663741					EU663686	
EU663688 EU663689 EU663690 EU663739 EU663740 EU663741					EU663687	
EU663689 EU663690 EU663739 EU663740 EU663741					EU663688	
EU663690 EU663739 EU663740 EU663741					EU663689	
EU663739 EU663740 EU663741					EU663690	
EU663740					EU663739	
EU662741					EU663740	
20003/41					EU663741	
EU663742					EU663742	
EU663743					EU663743	
Bhilingings Luzon 18 650050 121 260215 EU663744 Lamondra et al. 2000	Dhilinnings	Luzon	19 650050	121 260215	EU663744	Lomondro et al 2000
Euzon 18.639939 121.360213 EU663745	Philippines	Luzon	18.039939	121.300213	EU663745	Jamandre et al. 2009
EU663746					EU663746	
EU663747					EU663747	
EU663748					EU663748	
EU663749					EU663749	
EU663750					EU663750	

Country	Location	Latitude N	Longitude E	GenBank Accession Number	Source
				EU663751	
				EU663752	
				EU663753	
				EU663754	

Table IV.7 Haplotype identity list and basic statistics of haplotypes from <u>M. cephalus</u> mitochondrial DNA. 90MB = 90 Mile Beach; BOI = Bay of Islands; DARG = Dargaville; HOKI = Hokianga; HUTT = Lower Hutt, Wellington; KAI = Kaipara Harbour; PORI = Porirua; POUTO = Pouto; SI = South Island; WAIK = Waikato; WHAK = Whakatane. a) Cytochrome Oxidase I; b) D-loop region.

a) Cytochrome Oxidase I

Number of Haplotypes	Haplotype Diversity	Polymorphic sites	Sequence length (bp)	Nucleotide diversity (π)	Tajima's D	Fu and Li's D	Fu and Li's F	Fu's F
5	0.2342	5	502	0.00123	$-1.30026 \ p > 0.1$	-2.5312 <i>p</i> < 0.05	$-2.51536 \ p > 0.05$	$-2.884 \ p > 0.04$

List of Haplotypes

Haplotype	Haplotype Frequency	Individuals that share the haplotype	GenBank Accession Numbers
H1	17	BOP3347, KAI2423, PORI6, HUTT3, HUTT11, BOI3, BOP3355, KAI2169, HAGU6020, BOP3351, PORI9, DARG1, WAIK7, KAI2712, KAI2715, HAGU048, BOP1790	KX639485
Н2	210	90MB1, DARG6, BOI8, DARG2, BOI5, BOI7, BOI2, BOP5, 90MB5, BOI1, BOP4, HAGU062, HOKI4170, SI3, SI1, SI5, BOI4, KAI4910, BOI3915, KAI2710, KAI2711, KAI2452, KAI2426, KAI2424, KAI2716, KAI2422, KAI2418, KAI2417, KAI2170, HAGU6172, BOP3354, BOP3352, HOKI4169, HOKI4171, KAI2165, KAI2167, KAI4911, POUTO3, HAGU6173, KAI2717,	KX639486

Haplotype	Haplotype Frequency	Individuals that share the haplotype	GenBank Accession Numbers
		HAGU6175, HOKI4175, HOKI4174,	
		KAI9, PORI7, PORI8, POUTO1,	
		KAI37, KAI30, KAI18, HOKI10,	
		DARG10, SI2, SI4, HOKI2, WAIK5,	
		BOP2, BOP3, WAIK10, WAIK9,	
		DARG5, DARG7, DARG8, WAIK6,	
		BOI2833, BOI2836, BOI2839,	
		HOKI3087, BOI2841, HAGU053,	
		HAGU056, HAGU050, HAGU047,	
		BOP3350, BOI3866, BOI3865,	
		HAGU054, HAGU057, HAGU049,	
		BOP1867, HAGU0XX, HAGU6037,	
		HAGU055, HAGU052, BOI2840,	
		HOKI5683, BOI2838, HOKI5687,	
		BOI2832, HAGU6769, BOP1871,	
		BOP1873, BOP1875, BOP1902,	
		HAGU6017, BOP1903, BOP1904,	
		HAGU6018, BOP1870, HAGU6174,	
		BOP1900, HAGU6777, HOKI3084,	
		BOP1899, BOP1897, HOKI3082,	
		KAI4918, BOP1895, BOI3921,	
		BOI3920, POUTO6, POUTO7, BOP1,	
		POUTO9, HOKI1, POUTO5, POUTO4,	
		PORI5, PORI2, HOKI8, HOKI7,	
		HOKI6, HOKI3, DARG3, HUTT10,	
		HUTT1, HUTT5, PORI1, HUTT6,	
		BOI9, WAIK2, HAGU067, KAI31,	
		PADA060554, HOKI4, HOKI5691,	
		HUTT9, 90MB2, WAIK3, WAIK4,	
		KAI41, POUTO8, PORI4, POUTO10,	
		WAIK8, 90MB4, HUTT7, HUTT2,	
		WAIK060556, BOI10, HUTT8,	
		BOI2835, KAI4905, BOI7246,	
		BOP1901, KAI2168, HAGU6768,	
		DARG9, POUTO2, PORI10, KAI13,	

Haplotype	Haplotype Frequency	Individuals that share the haplotype	GenBank Accession Numbers
		PORI3, KAI42, KAI26, BOI3914,	
		BOI7242, KAI2714, KAI2421,	
		BOP1732, KAI2420, BOP1733,	
		BOP1786, BOP1868, KAI2419,	
		HOKI5692, BOP3356, HAGU061,	
		HOKI3085, HOKI3080, HAGU2163,	
		HAGU6179, HAGU6038, HAGU6039,	
		90MB3, BOI2834, HAGU6781,	
		KAI4909, HAGU6165, KAI4920,	
		BOP1898, HOKI5, WAIK1, HOKI4176,	
		KAI2166, BOI6, KAI2162, BOI3917,	
		HOKI3081, BOI3919, BOI3916,	
		BOI3913, BOI3872, BOP1735,	
		KAI4912, PADA060555, BOP1876,	
		KAI4913, HAGU6776, HUTT4,	
		HOKI3088	
H3	1	BOI3870	KX639487
		DARG4, HAGU066, BOP1869,	
ЦЛ	10	НОКІЗ086, НОКІ9, ВОРЗЗ49,	KY620488
114	12	BOI3918, BOP1736, BOP1865,	KAUJ7400
		HAGU051, HOKI4172, KAI11	
H5	1	HAGU6166	KX639489

b) D-loop region

Number of Haplotypes	Haplotype Diversity	Polymorphic sites	Sequence length (bp)	Nucleotide diversity (π)	Tajima's D	Fu and Li's D	Fu and Li's F	Fu's F
50	0.9723	176	864	0.05	$0.67163 \ p > 0.1$	0.09415 <i>p</i> > 0.1	$0.40969 \ p > 0.1$	$0.263 \ p > 0.1$

List of Haplotypes

Haplotype	Haplotype Frequency	Individuals that share the haplotype	GenBank Accession Numbers
H1	5	90MB01, DARG10, HUTT10, KAI31, WAIK01	KX639435

Haplotype	Haplotype Frequency	Individuals that share the haplotype	GenBank Accession Numbers
H2	4	90MB02, DARG05, WAIK02, WAIK09	KX639436
Н3	1	90MB03	KX639437
H4	1	90MB04	KX639438
Н5	3	BOI01, HOKI08, POUTO09	KX639439
H6	4	BOI02, BOI05, BOI06, POUTO04	KX639440
H7	1	BOI03	KX639441
H8	1	BOI04	KX639442
Н9	7	BOI07, BOI09, HOKI02, HOKI07, PORI07, SI01, SI05	KX639443
H10	3	BOI08, DARG02, WAIK10	KX639444
H11	10	BOI10, DARG09, HUTT07, KAI26, KAI42, PORI03, PORI04, PORI10, POUTO10, WAIK08	KX639445
H12	2	DARG01, PORI06	KX639446
H13	6	DARG03, DARG07, HUTT04, PORI01, POUT007, WHAK01	KX639447
H14	1	DARG04	KX639448
H15	1	DARG06	KX639449
H16	3	DARG08, HOKI01, HOKI03	KX639450
H17	1	HOKI04	KX639451
H18	2	HOKI05, SI02	KX639452
H19	3	HOKI06, HUTT06, KAI18	KX639453
H20	1	HOKI09	KX639454
H21	2	HOKI10, HUTT09	KX639455
H22	1	HUTT01	KX639456
H23	1	HUTT02	KX639457
H24	1	HUTT03	KX639458
H25	1	HUTT05	KX639459
H26	1	HUTT08	KX639460
H27	1	HUTT11	KX639461
H28	1	KAI09	KX639462
H29	1	KAI11	KX639463
H30	1	KAI13	KX639464
H31	1	KAI30	KX639465

Haplotype	Haplotype Frequency	Individuals that share the haplotype	GenBank Accession Numbers
H32	2	KAI37, POUTO05	KX639466
H33	1	KAI41	KX639467
H34	1	PORI02	KX639468
H35	1	PORI05	KX639469
H36	1	PORI08	KX639470
H37	1	PORI09	KX639471
H38	2	POUTO01, WAIK06	KX639472
H39	1	POUTO02	KX639473
H40	1	POUTO03	KX639474
H41	1	POUTO06	KX639475
H42	1	SI03	KX639476
H43	1	SI04	KX639477
H44	1	WAIK03	KX639478
H45	1	WAIK04	KX639479
H46	1	WAIK05	KX639480
H47	1	WAIK07	KX639481
H48	2	WHAK02, WHAK03	KX639482
H49	1	WHAK04	KX639483
H50	1	WHAK05	KX639484

Table IV.8 Variable entities and clusters calculated with the General Mixed Yule Coalescence model. The significant model was with multiple threshold that represents three events where the variation among haplotypes (coalescence time) could be explained by Yule's speciation model predicting a total of nine highly variable clusters. The time is negative and refers to years in the past.

Parameters	Values
Likelihood of null model	-397.0602
Maximum likelihood of GMYC model	-393.728
Likelihood ratio	6.664387
Probability of likelihood ratio test	0.0357
Number of ML clusters	9
Confidence interval	6-12
Number of ML entities	12

Confidence interval	7-17
	-50,509.18
Threshold time (years)	-27,328.26
	-16,176.15

Table IV.9 Assignment probabilities of individuals used in Chapters II and IV. Columns L1 and L2 refer to the assignment probabilities to the Haplogroup of origin according to the bayesian approach by Rannala and Mountain (1997). The probabilities highlighted in red represent the probability based on msatDNA allele frequencies that did not matched the haplogroup of origin. A total of 56.3% of the individuals were assigned correctly to the haplogroup of origin. Columns Cluster 1-4, show the clustering probabilities of msatDNA allele frequencies based on the Fst comparisons of ancestry and expected allele frequencies with the bayesian approach as performed in STRUCTURE. The higher probabilities are highlighted in yellow.

Haplogroups	Individuals	L1	L2	Cluster 1	Cluster 2	Cluster 3	Cluster 4
L1	WAIK08	0.012	0.253	0.981	0	0.016	0.002
	BOI10	0.384	0.279	0.57	0	0.424	0.005
	KAI13	0.039	0.256	0.448	0.156	0.395	0
	KAI26	0.092	0.168	0.459	0.118	0.423	0
	KAI41	0.123	0.425	0.451	0.137	0.412	0
	KAI42	0.7	0.512	0.457	0.12	0.423	0
	DARG09	0.011	0.039	0.211	0	0.362	0.427
	POUTO02	0.281	0.608	0.246	0.031	0.601	0.123
	POUTO10	0.123	0.392	0.3	0.028	0.545	0.127
L2	PORI03	0.011	0.01	0.265	0.447	0.173	0.115
	PORI04	0.298	0.478	0.203	0.55	0.142	0.105
	PORI10	0.04	0.064	0.188	0.576	0.1	0.137
	WAIK01	0.713	0.851	0.979	0	0.019	0.002
	WAIK02	0.076	0.116	0.981	0	0.016	0.003
	WAIK03	0.42	0.728	0.987	0	0.012	0.001
	WAIK04	0.647	0.639	0.979	0	0.02	0.001
	WAIK05	0.011	0.229	0.983	0	0.015	0.002
	WAIK06	0.082	0.079	0.982	0	0.015	0.002
	WAIK07	0.177	0.677	0.985	0	0.013	0.001

	WAIK09	0.01	0.121	0.98	0	0.017	0.003
	WAIK10	0.095	0.02	0.987	0	0.01	0.002
	BOI01	0.203	0.498	0.864	0	0.132	0.004
	BOI02	0.744	0.596	0.796	0	0.199	0.006
	BOI03	0.52	0.795	0.579	0	0.416	0.005
	BOI04	0.123	0.549	0.857	0	0.136	0.007
	BOI05	0.607	0.773	0.626	0	0.37	0.004
	BOI06	0.742	0.768	0.778	0	0.214	0.008
	BOI07	0.269	0.329	0.762	0.001	0.23	0.008
	BOI08	0.391	0.522	0.711	0	0.28	0.008
	BOI09	0.773	0.831	0.724	0	0.273	0.003
	KAI09	0.226	0.352	0.426	0.109	0.466	0
-	KAI11	0.524	0.541	0.551	0.109	0.339	0
	KAI18	0.391	0.38	0.532	0.144	0.324	0
	KAI30	0.145	0.225	0.519	0.164	0.317	0
	KAI31	0.531	0.679	0.391	0.151	0.458	0
-	KAI37	0.47	0.325	0.454	0.165	0.381	0
	DARG01	0.037	0.059	0.434	0	0.338	0.228
	DARG02	0.175	0.479	0.472	0	0.285	0.243
	DARG03	0.221	0.27	0.438	0	0.351	0.21
-	DARG04	0.281	0.176	0.358	0	0.556	0.086
	DARG05	0.124	0.099	0.352	0	0.153	0.495
-	DARG06	0.85	0.495	0.372	0	0.421	0.207
	DARG07	0.12	0.467	0.368	0	0.167	0.466
	DARG08	0.316	0.105	0.513	0.002	0.268	0.218
	DARG10	0.689	0.957	0.356	0.008	0.377	0.259
	POUTO1	0.207	0.074	0.282	0.029	0.508	0.181
	POUTO3	0.184	0.294	0.294	0.029	0.557	0.12

POUTO4	0.534	0.957	0.217	0.024	0.661	0.098
POUTO5	0.248	0.548	0.254	0.029	0.596	0.121
POUTO6	0.705	0.643	0.231	0.037	0.601	0.131
POUTO7	0.176	0.035	0.287	0.033	0.58	0.1
POUTO8	0.124	0.168	0.313	0.042	0.539	0.106
POUTO9	0.566	0.896	0.281	0.036	0.563	0.12
PORI01	0.082	0.014	0.169	0.647	0.101	0.083
PORI02	0.023	0.011	0.216	0.55	0.087	0.147
PORI05	0.188	0.909	0.229	0.469	0.191	0.111
PORI06	0.029	0.024	0.225	0.549	0.1	0.126
PORI07	0.216	0.3	0.183	0.614	0.099	0.104
PORI08	0.567	0.723	0.267	0.425	0.143	0.165
PORI09	0.724	0.277	0.174	0.599	0.136	0.091
HOKI01	0.389	0.562	0.634	0.008	0.163	0.195
HOKI02	0.394	0.573	0.55	0.002	0.31	0.138
HOKI03	0.254	0.629	0.616	0.001	0.214	0.169
HOKI04	0.652	0.46	0.33	0.001	0.299	0.37
HOKI05	0.1	0.022	0.663	0	0.207	0.13
HOKI06	0.054	0.266	0.49	0.001	0.191	0.318
HOKI07	0.06	0.004	0.264	0.001	0.64	0.095
HOKI08	0.172	0.095	0.476	0	0.137	0.386
HOKI09	0.778	0.832	0.297	0.005	0.388	0.311
HOKI10	0.394	0.052	0.198	0.002	0.698	0.102
SI1	0.399	0.71	0.37	0.004	0.531	0.095
SI2	0.612	0.503	0.352	0.004	0.535	0.11
SI3	0.554	0.706	0.315	0.005	0.581	0.098
SI4	0.082	0.054	0.354	0.006	0.521	0.118
SI5	0.438	0.749	0.347	0.005	0.547	0.101

WHAK1	0.778	0.827	0.185	0.161	0.543	0.11	
WHAK2	0.043	0.009	0.472	0.107	0.336	0.086	
WHAK3	0.289	0.253	0.219	0.11	0.523	0.148	
WHAK4	0.382	0.672	0.255	0.089	0.563	0.093	
WHAK5	0.769	0.971	0.166	0.149	0.581	0.104	

Table IV.10 Estimation of migration ratio among geographical regions based on D-loop sequences for the 12 New Zealand locations.

New Zealand region	Average rate of population expansion per generation.				
South	0.54797				
West	0.49747				
East	0.66876				
	Average proportion of migrants between populations per generation				
East-South	0.39426				
West-South	0.33592				
South-East	0.78937				
West-East	0.83414				
South-West	0.4269				
East-West	0.30498				



Figure IV.8 Posterior probability branch support analysis for the GMYC model that shows the entities that are well defined and which variation could be explained by Yule's speciation model. Only three entities have a significant posterior probability branch support (p > 0.95) and those are the candidate nodes where the time threshold could be confidently located. The red blocks are the nine clusters predicted by the GMYC model.



Figure IV.9 Neighbour-Joining (NJ) tree of COX1 sequences from New Zealand and the Pacific Rim. The NJ tree was constructed based on HKY nucleotide substitution model and pair-wise comparisons between sequences. From right to left, haplogroup L2 haplotypes are only found in New Zealand. Haplogroup/lineage NWP1 distributes from Russia to Australia. Haplogroup NWP2 is comprised of haplotypes distributed from Russia to the Philippines. Haplogroup L1 are haplotypes that are closely related to NWP2 and are distributed only in Australia and New Zealand. Haplogroup Chile are haplotypes only found in Chile. Haplogroup NWP3 is comprised by haplotypes distributed from Russia to New Caledonia. Haplogroups NWP1, NWP2 and NWP3 have been previously identified in the Pacific (Shen et al., 2011; Durand et al., 2012c). The bar charts on the right represent the number of haplotypes in the NJ tree that come from the Pacific location in the X-axis.



Figure IV.10 Neighbour-Joining (NJ) tree of the D-loop sequences from New Zealand and the Pacific Rim. The NJ was based on HKY + G+ I nucleotide substitution model and pair-wise comparisons between sequences. From right to left, haplogroup L2 and L1 were comprised of haplotypes only found in New Zealand. Haplogroup NWP2 are haplotypes distributed from China to Philippines. Haplogroup Pacific IV is distributed from China to Japan. The bar charts on the right represent the number of haplotypes in the NJ tree that come from the Pacific location in the X-axis.



Figure IV.11 Skyline plots of New Zealand sequences. Each skyline plot was run once, the size of the MCMC was 100^{6} - 400^{6} and ESS values were above 200. The skyline plot of COX1 sequences based was calculated with the HKY nucleotide substitution model. The skyline plot of D-loop sequences was calculated with the HKY + G + I nucleotide substitution model. The skyline plot of concatenated sequences was calculated with the GTR + G + I nucleotide substitution model. The skyline plots for L1 and L2 were calculated using concatenated sequences and the same nucleotide substitution model as the concatenated skyline plot. N = effective population size.



Figure IV.12 Phylogenetic analysis based on COX 1 sequences of <u>M. cephalus</u> from the Pacific Rim including New Zealand. Collapsed branches correspond to previous lineages identified in Shen et al. 2011., the lineage from Chile and L2.

General Discussion

Overview on Mugil cephalus in New Zealand

The overall goals of population genetics are to describe species genetic diversity patterns and to identify the factors that could act as barriers to gene flow. As the amount of population genetics research increases and the level of resolution obtained with new molecular markers improves, we have realized that those patterns and processes are more complex and entangled than previously thought. This has been particularly evident in marine species. Most marine species have very large population sizes, potentially disperse over long distances and tend to cross a range of complex environments, which means the influence and interaction among genetic drift, selection and gene flow can be quite different from terrestrial species. Mugil cephalus is thought to contain several cryptic species due to the variability of its mitochondrial DNA and its preference for coastal habitats. This association with shallow water habitats seems difficulty to reconcile with its world-wide distribution, which suggests it has the tendency for oceanic dispersal, but its populations are known to contain highly divergent lineages in different regions of the world (Rocha-Olivares et al. 2000; Rossi et al. 2004; Fraga et al. 2007; Heras et al. 2007; Livi et al. 2011; Durand et al. 2012b). In this study, we found that *M. cephalus* in New Zealand is more complex than previously thought. Its genetic structure is a pattern of distinct populations despite significant gene flow and temporal shifts in the genetic diversity due to seasonal reproductive success. It is a species that has cosmopolitan sources that most likely colonised New Zealand in different episodes. Although a large proportion of genetic variation could be explained by reproductive behaviour and intermittent inter-oceanic migrations, the effects of anthropogenic activities should not be ruled out. Understanding the effects of human-based activities in addition to M. cephalus population genetic structure, reproductive behaviour and migration patterns, will ensure a sustainable management of *M. cephalus* fishery in New Zealand.

Population genetic structure of *M. cephalus* in New Zealand.

The patterns in the genetic structure and connectivity of *M. cephalus* in New Zealand using nuclear microsatellite DNA (msatDNA) showed recent admixture of four genetically distinct ancestral populations with a sympatric distribution (Chapter II). The genetic differences were not correlated with linear geographic distances or geophysical attributes. Instead, low but significant levels of gene flow among locations were found. The msatDNA markers used showed linkage not as result of physical proximity, but a consequence of a complex demographic history in which M. cephalus in New Zealand could be the result of multiple source populations from other regions of the world. The low levels of gene flow, high proportion of non-migrants and the chaotic genetic structure showed by M. cephalus populations in New Zealand, could be explained by seasonal reproductive variation, different spawning grounds and philopatry (Hedgecock 1994; Thorrold et al. 2001; Jones et al. 2005; Svedäng et al. 2007; Hedgecock and Pudovkin 2011; Bentzen and Bradbury 2016; Bonanomi et al. 2016). The chaotic distribution of genetic diversity is also reflected in significant shifts in *Ne* among locations. Five locations showed strong effects of bottlenecks that have previously been identified as different putative stocks based on catch per unit effort (CPUE) analyses, and the same locations were used to test for temporal changes in genetic diversity (see below) (McKenzie and Vaughan 2008).

Sweepstakes Reproductive Success of M. cephalus in New Zealand

The differences between the levels of genetic variation found in juveniles and adults were consistent with a sweepstakes reproductive success (SRS) in New Zealand *M. cephalus* (Chapter III). There were significant differences in the allele frequencies and effective population sizes between juveniles and adults but the study was limited to identify the specific environmental factors driving the seasonal bottlenecks in the juveniles. Two spawning groups of *M. cephalus* in New Zealand were identified based in cluster analyses based on the variation of the genetic distances and the assessment of the variation in the effective population sizes or

number of breeders required to explain the observed allele frequencies in each group of juveniles compared to the adults. An early spawning group that starts spawning in October and a late spawning group that finish the spawning season in March. The method to assess the *Ne* based on *He* excess showed that the SRS pattern alone could not explain the total variation found in the *Ne* between juveniles and adults.

The locations used in the SRS assessment are areas with high fishing pressure and were identified as putative fishery stocks based on CPUE (McKenzie and Vaughan 2008). Whether the unexplained variation in the *Ne* between juveniles and adults in estuarine nursery areas for the GMU1 management area, is the consequence of fishing pressure and other human-based activities (e.g. pollution or nitrogen loading), or a combination of both, remains uncertain and further work is needed. Species that fit within the SRS hypothesis maintain their overall level of genetic diversity by the accumulated contribution made by various overlapping generations to the adult breeding group. If a cohort fails to recruit into the breeding group, there could be a loss or underrepresentation of alleles transmitted to the next generation. Healthy nursery environments contribute directly to the maintenance of genetic diversity in *M. cephalus* populations. The erosion of genetic diversity because of a failed cohort could add pressure on an already low population size and this may become compounded by the effects of climate change.

The fact that *M. cephalus* has a specific reproductive behaviours and ecological flexibility allowing it to occupy a range of niches, could be other reasons for why there might be a chaotic genetic structure and gene flow among non-adjacent areas. Philopatry would help to explain the overlapping geographic distribution of genetically differentiated groups and higher gene flow between more distant rather than close locations. It is possible that the different *M. cephalus* populations identified in Chapter II, have specific spawning grounds spread throughout the North Island of New Zealand, facilitating their dispersal to reach distant

locations. Philopatry in *M. cephalus* would disrupt the random mating contributing to the levels of genetic differentiation. Philopatry in marine fish species and possibly *M. cephalus* is most likely to be the consequence of a combination of chemical cues and geomagnetic imprinting (*i.e.* Putman et al. 2013). Although this is somewhat problematic for explaining the pattern for species that passively drift in the water column as eggs or recently hatched larvae (Bentzen and Bradbury 2016). Seasonal changes in the current patterns could also affect dispersal and allow distant locations to be genetically similar and neighbour population being genetically distinct.

Phylogeographic patterns of *M. cephalus* in New Zealand and the Pacific

Mugil cephalus was thought to be a species complex comprised of ~14 cryptic species (Durand et al. 2012b). This thesis research of the *M. cephalus* populations in New Zealand has added to a handful of other studies that have come out recently to challenge that suggestion. We now know that *M. cephalus* in New Zealand has a chaotic genetic structure and there is temporal variation in the genetic diversity between juveniles and adults. The analysis of mtDNA sequences found that there were two highly divergent haplogroups for *M. cephalus* in New Zealand, which were in sympatry (haplogroup L1 and haplogroup L2; Chapter IV). It was concluded that these were most likely the result of two colonisation events from allopatric populations in the Pacific. The comparison of mtDNA and msatDNA data showed that the two haplogroups were intermixed; there were no morphological characters or other evidence to suggest that both haplogroups were reproductively isolated or to support the suggestion that they were separate species. The haplogroup L1 showed few highly divergent low frequency haplotypes and one high frequency haplotype. It was less diverse than L2 and it shared two haplotypes with Australia suggesting Australia might be the source population of a relatively recent colonisation (~20, 000 to ~16, 000 years ago). The haplogroup L2 was more diverse and with several low frequency haplotypes, suggesting a longer period of time accumulating mutations after the founder haplotypes arrived to New Zealand. Haplogroup L2 was distinct

from other areas of the Pacific and was closely related to haplogroups distributed from Russia to Australia. Haplogroup L2 was probably a result of the first colonisation event ~50,000 years ago. The level of mtDNA variation in *M. cephalus* populations in New Zealand was similar to those reported between some species of the genus *Mugil*. The key factors that might have enabled these two founder events are difficult to determine. However, the effect of environmental factors such as ocean conditions over the geographic distances, variation of major oceanographic current patterns or salinity/temperature levels between estuaries, combined with migrations from feeding to spawning grounds, have been considered important in other *M. cephalus* studies of populations in the Pacific and the Mediterranean Sea (Rossi et al. 1998b; Huang et al. 2001; Blel et al. 2010; Shen et al. 2011).

Highly divergent mtDNA lineages does not equal cryptic species

Mitochondrial DNA (mtDNA) is maternally inherited and has a different mutational rate from nuclear DNA markers (Avise 2004). The lack of recombination of the mitochondrial genome has shown that the events that shaped the frequency of mtDNA sequences tens to hundreds of thousands of years ago, can still be identified in contemporary populations (Avise 2004; Hickerson et al. 2010). The level of sequence variation and differentiation found between populations might be similar to the level reported between some species; however, it is not necessarily evidence that the populations are different species. Species with very large population sizes will retain high levels of genetic diversity because genetic drift – a force of evolution that eliminates genetic diversity – is weak (Palumbi 1994).

Colonisations and secondary contacts could also be responsible for high levels of difference among mtDNA sequences (Waters et al. 2013b). It only takes a few migrants per generation to maintain genetic connectivity. Gene flow between populations will homogenize nuclear allele frequencies and any historical significance of past differentiation will be lost

(Palumbi 1994). However, mtDNA does not undergo recombination and the DNA sequences of haplotypes are unable to be homogenized. This means the patterns of mtDNA can persist for many more generations than they will in nuclear loci (Avise 2000). The distinct mtDNA haplotypes carried by the migrants or founders to a new population will remain until there is complete lineage sorting by genetic drift or selection (Hallatschek and Nelson 2008). The time it takes for genetic drift or selection to change the haplotype frequencies is dependent on the effective population size of founders and residents and it could take a long period of time (Avise 2004; Zecca et al. 2011). Not enough time for genetic drift to sort the lineages could explain distinct mtDNA haplogroups with significant levels of gene flow that do not match the populations contributing to each individual ancestry found with other nuclear markers such as msatDNA markers.

The changes in environmental conditions can produce new barriers to gene flow, which can have a temporary effect on the genetic structure of species, for example, periods of isolation between populations which are followed by migration events such as in postglacial recolonisation (Waters 2011). Intermittent bouts of migrants from source populations with differentiated haplogroups, can produce patterns of population genetic variation in the foundered group that may resemble a species complex (Zecca et al. 2011; Reid et al. 2016). Without obvious diagnostic features to support the assertion of a separate species, such as niche specialization or diagnosable genetic incompatibility, it is most reasonable to conclude that the genetic variation found within the populations is a consequence of isolated sources, large genetically effective population sizes and sporadic migration events (*i.e.* Grant and Grant 2009; Pettengill and Moeller 2012; Reid et al. 2016). *Mugil cephalus* populations in the Pacific Ocean have shown population genetic patterns that are associated with intermittent migration events and there is no compelling evidence to support the suggestion that the New Zealand population is comprised of more than one species. Delimiting species based only on the number of differences between non-recombining DNA sequences can be problematic. The best situation is to find fixed genetic differences that strongly suggest genetic incompatibility, which are also associated with other attributes such as morphological characters. DNA sequence differences in non-recombining genomes (mtDNA for example), could be used to support new species hypotheses, but not when using genetic data from highly structured populations. In this case it would be difficult, if not impossible, to distinguish species differences from the partitioning of genetic variation within a species that will develop across a landscape/seascape (De Queiroz 2005). The demographic history of each population combined with the slow effect of genetic drift in lineage sorting is the source of the non-recombining genome sequence differences rather than species diversification (Kingman 2000; Wilson et al. 2003; Sargsyan and Wakeley 2008).

The phylogenetic species concept has been criticised when used in this situation because it fails to properly test for reproductive incompatibility (Irwin, 2002). When a phylogenetic approach is used to delimit species (typically based on mtDNA markers) it may overestimate the number of species if females have low dispersal distances and only paternally inherited genes are being regularly exchanged between populations (Avise 2000; Irwin 2002; Hickerson et al. 2010). *Mugil cephalus* mtDNA studies are a good example of this case. Depending on which mtDNA marker was used, the phylogenetic relationships were different (see Jamandre et al. 2009, 2014, Durand et al. 2012b,a; Whitfield et al. 2012). When nuclear DNA such as msatDNA markers were used in a complementary way to a mtDNA phylogenetic analysis, isolated populations were found in the China Sea; however, there was evidence of gene flow between seemingly isolated mtDNA lineages in the Mediterranean (Blel et al. 2010; Shen et al. 2011; Durand et al. 2012a,b). The simplicity of the phylogenetic species concept means it will fail to recognise that species responses to the environment may differ between

locations, putting them on different evolutionary trajectories (Simpson 1951; Hennig 1966; Donoghue 1985).

Highly structured populations of *M. cephalus* were found in New Zealand and the different genetic structures found between mtDNA and msatDNA markers might be caused by admixture of migrants that have a cosmopolitan origin. There was no evidence that could be used to adjudicate on any speculation about genetic incompatibility among *M. cephalus* populations. Therefore, it was concluded that *M. cephalus* in New Zealand is most likely not a species complex, instead it is composed of divergent populations that were established from cosmopolitan sources.

Mitochondrial DNA and colonisation wave of haplotypes

Populations of marine species capable of long-distance migrations face environmental limitations to a successful colonisation (Palumbi 1994). Habitat suitability and available spaces to occupy are two factors that can limit the success of migrants as they could affect the capability of the individuals to reproduce (establish) in the new site (Buckley et al. 2013; Waters et al. 2013b). However, there are other important factors that are not commonly taken into account when it comes to migrants of marine populations establishing in new areas.

The sampling effect of migration offers opportunity for low frequency or rare alleles in the source population to be a significant portion of the founder group. This pattern has been described as a "wave" where low frequency genes are carried or "surf the wave" along a portion of the high frequency genes from the source population (Hallatschek and Nelson 2008). This pattern is called "gene-wave fronts" and it helps to explain the genetic differentiation levels in founder expanding populations (Edmonds et al. 2004; Hallatschek and Nelson 2008). Migrants are exposed to the effect of environmental and behavioural variables that could disrupt their opportunities to settle in new areas or to reproduce with the resident population (Edmonds et al. 2004). Not all migration events are successful and it might take several migration episodes before a founder population is successfully established (Buckley et al. 2013). The gene-wave fronts and the intermittent success of migration events will typically cause the source and the founder populations to have significantly different gene frequencies despite successful migratory events.

Mugil cephalus are known for their ability for oceanic-level migrations (see Whitfield et al. 2012 for a review). Postglacial re-colonisation events and secondary contacts between allopatric populations have been identified in *M. cephalus* populations from the northwest Pacific (Shen et al. 2011). The mtDNA analysis in this study (Chapter IV), showed that New Zealand was colonised by *M. cephalus* at least twice from allopatric populations from other areas in the Pacific, resulting in two genetically distinct haplogroups. Different low frequency haplotypes were surfing each gene-wave front, which resulted in two highly divergent haplogroups that persisted for several thousands of years after the colonisation event. The genetically distinct haplogroups were a vestige of New Zealand colonisation by *M. cephalus* where there has not been enough time for genetic drift to completely sort the lineages.

Sympatric speciation model?

Speciation models have two main components, the spatial distribution of populations and genetic isolation. The allopatric speciation model refers to how new species arise after random mating within a single population is disrupted by the appearance of gene flow barriers that sunder a population into two populations which genetically diverge into reproductively incompatible groups (Avise and Ayala 2009). The peripatric speciation model occurs when a subset of individuals colonise a new area without further migratory events from the source population, or they occupy a new niche and mate only with the population that share their niche preferences, ultimately leading to genetic isolation (Avise and Ayala 2009). The parapatric speciation model is where, despite a continuity in the distribution range of a species and no apparent barriers to gene flow, the extremes of the distribution range are genetically isolated (Avise and Ayala 2009). The gene exchange occurs at the centre of the species distribution and is the core of species population genetic structure following an isolation-by-distance pattern or IBD (Slatkin and Maddison 1990; Bradbury and Bentzen 2007; White et al. 2010).

The sympatric speciation model occurs when a new species evolves from a single ancestral species inhabiting the same geographic region (Avise and Ayala 2009). In the sympatric speciation model, genetic isolation is evident after the random mating is disrupted due to selection or mutation. Selection after the apparition of mutants could lead to ecologically important genomic regions that tend to resist genetic homogenization because it lowers fitness in some individuals, while genetic homogenization continues among most other parts of the genome, which can produce separate species while in sympatry (Avise and Ayala 2009).

Understanding which speciation model fits the natural patterns of genetic structure in species is often dependent on the study area characteristics and more importantly the researcher's approach. To really comprehend how speciation in marine species occurs, the sampling localities need to be removed as a grouping variable, which will help to define populations as genetically distinct units instead of where they come from. This will unavoidably lead to a change in how speciation is defined. A way to avoid the bias brought to our research in the sampling design used to address our specific questions and test our hypothesis, is to use variables that will group individuals together that are independent from our sampling design limitations (Dean et al. 2015).

In this study, the main limitation was the broad distribution of *M. cephalus* in New Zealand, and panmixia was the null hypothesis in most of the analyses used in this study. The way to avoid this geographic constraint, was to define *M. cephalus* populations based on

185

whether individuals form a genetically distinct group irrespective of their sample location. The conclusion was to consider *M. cephalus* populations in New Zealand as having overlapping distributions, in which the levels of genetic differences and genetic diversity did not follow a simple spatial pattern. The intermittent migrations found in *M. cephalus* populations in the Pacific have possibly allowed it to maintain a cosmopolitan distribution (Shen et al. 2011; Whitfield et al. 2012).

In the last 50,000 years the main ocean surface and bottom circulation patterns have been in constant change. The changes in the circulation patterns could have allowed M. *cephalus* and other marine species populations to undergo secondary contacts of previously isolated populations, reach distant areas and to possibly colonise new areas. Between ~100, 000 and ~40,000 years ago, the rise in sea level could have allowed isolated populations of M. *cephalus* to utilise new migration routes (Pichevin et al. 2009; Morley et al. 2014; Pena and Goldstein 2014). Approximately 20,000 years ago the bottom oceanic circulation was reversed starting in the south Atlantic rather than north Atlantic as it is today (Negre et al. 2010).

This change in direction of the oceanic bottom currents is likely to have had an effect on *M. cephalus* populations changing the distribution of nutrients and optimal spawning temperatures. The bottom ocean circulation has major consequences for the climate and the persistence of many marine species. High density water masses sink in the north Atlantic between Greenland and Norway, full of nutrients, carbon and oxygen. From there they travel as a bottom oceanographic current towards the Indian Ocean and then split to the North and South Pacific. This high nutrient, carbon and oxygen loaded water mass used to sink in the south Pacific, west of Africa, ~20,000 years ago (Negre et al. 2010). The principal effect was in the upwelling currents that bring cold and nutrient-rich water from the bottom of the ocean, affecting temperature and the distribution of marine biota (Negre et al. 2010). The environmental changes in the last ~100,000 years produced transient barriers to gene flow and *M. cephalus* in New Zealand and the Pacific would fit the allopatric speciation model with secondary contacts.

The four genetically distinct populations of *M. cephalus* found in New Zealand with an overlapping distribution, could also be understood as peripatric populations occupying different reproductive niches (two spawning groups). A chaotic genetic structure in *M. cephalus* populations in New Zealand was identified as a result of SRS. Different reproductive successes of the adult group of *M. cephalus* in New Zealand means not all of the adults are successfully contributing to the species gene pool, and there are different successful reproducers in different spawning seasons. Two spawning groups of adults were identified, an early spawning group and a late spawning group; and also is possible that there are different spawning groups. With an identifiable barrier to gene flow, sympatry in *M. cephalus* populations in New Zealand would not be entirely applicable.

Chaos in a seemingly stable environment and genetic incompatibility

Spatially chaotic genetic patterns are the result of temporal genetic variation between early-life history stages and adults (Hedgecock 1994; Hedgecock and Pudovkin 2011). *Mugil cephalus* had a chaotic genetic structure with strong effects of Sweepstakes-Reproductive-Success (SRS). The reasons for differential reproductive success among the adults is different spawning behaviours, philopatry and in part responses to environmental conditions. The genetic variability in populations over small time scales do not translate in independent reproductive units and the conditions required for speciation. While speciation is possible, it is unlikely because gene flow is still occurring between populations without evidence of selection or drift, where genetic variation will tend to be homogenized.

Barriers to gene flow are constantly changing especially over long periods of time (e.g. Stefanni and Knutsen 2007; Zecca et al. 2011; Reid et al. 2016). Hence, unless we identify a

reproductive incompatibility among *M. cephalus* haplogroups, populations or individuals, we are only able to conclude that it is a single species with highly structured populations. There is an expectation that as two populations diverge they will eventually accumulate enough genetic differences to become genetically incompatible becoming different species. However, some widely-distributed species with a chaotic population genetic structure, such as *M. cephalus*, seem to maintain their reproductive cohesion and remain as a single species due to large effective population sizes.

Number of loci and sample size limitations

Widely discussed in population genetic studies is the balance between sample size and the number of loci that can reflect unbiased population genetic variation. The use of low variability markers and low small sample sizes invariably will lead to underestimation of the demographic and evolutionary processes that are shaping the species genetic structure.

The problem lies in accurately identify the population genetic structure and the most frequent methods is Wright's F-statistics or F_{ST} (Meirmans and Hedrick 2011). F-statistics was originally developed as an inbreeding coefficient assuming loci to be bi-allelic; since then F_{ST} analogues such as Nei's G_{ST} (Nei 1987), F'_{ST} (Weir and Cockerham 1984), Slatkin's R_{ST} (Slatkin 1995), have been implemented to address multi-allelic loci using a ratio of genetic variances, comparing expected heterozygosity within and among populations under different evolutionary models (Meirmans and Hedrick 2011). F-statistics measures are constrained to relating the amount of genetic variation among populations to the total genetic variation over all populations in which the significant levels of differentiation are determined by the within-population genetic variation (Meirmans and Hedrick 2011). R-statistics developed by Slatkin (1995), is not constrained to the within-population genetic variation because the population sizes and number of migrants are unbiased over the mutation rates following a step-wise

mutation model (Slatkin 1995). However, when the influence of mutation is small, R_{ST} estimates are not reliable because the new alleles are not adding significant variation to the population (Meirmans and Hedrick 2011). Using F_{ST} combined with other F-statistics analogues would allow a better assessment of the population genetic structure and the influence of demographic processes, selection, genetic drift and migration.

Each F-statistics will require different sample sizes and different amount of variable loci to significantly reflect population genetic variation. If sample sizes or the number of variable loci are low, the loci might show biased effects of bottleneck, strong genetic drift and inbreeding (Piry et al. 1999). Large number of variable loci will increase the within-population genetic variation and could lead to an overestimation of the number of genetically distinct groups. Increasing the within-population genetic variation by adding variable loci to the population genetic analysis, could reflect the variation among genealogies of groups of individuals rather than a true estimation of the population genetic structure by comparing levels of differentiation among populations. The result is an overestimated F_{ST} and an overestimation of the number of populations in the species distribution range (Luikart et al. 1998; Kalinowski 2005; Ryman and Palm 2006). While some studies had a view that high statistical power is found in large sample sizes and many polymorphic loci (the bigger the better), and the lack of significant differentiation is due to low number of polymorphisms in the loci used, this not always true (Ryman and Palm 2006). It is important to know the marker that are used to assess the population genetic structure of a species to avoid waste of computation efforts due to high number of polymorphisms in each locus; the variation in each locus is not infinite and can usually be assessed with allele-discovery curves which reflect the number of individuals needed to find a new allele that could significantly contribute to genetic differentiation (Rioux-Paquette 2012).

There are different ways to assess the statistical power of the sample size and loci used, POWSIM being one of the most common (Ryman and Palm 2006). Unfortunately, the statistical power of the loci in population genetic analyses is not always quantitatively evaluated making "the bigger the better" an intuitive understatement. POWSIM uses a predefined level of genetic differentiation which is the smallest significant F_{ST} value found between populations across all the data set. The smallest significant F_{ST} value is used in simulations with different sample sizes, number of loci and allele frequencies to estimate the alpha error (the probability to accept the null hypothesis of genetic homogeneity when is false) for any hypothetical degree of true differentiation between the simulated populations (Ryman and Palm 2006). In this study there was used a low number of loci compared to other population genetic studies that suggest the use of 10-12 variable loci and sample sizes and ~20 individuals to obtain a reliable $F_{ST} < 0.05 > 0.01$ (Kalinowski 2005). However, the sample sizes and loci used showed enough statistical power to detect significant "true" variation among populations (Chapter II).

Conclusions and future directions

Understanding how populations evolve over time and how genetic diversity is maintained is of primarily importance to understanding evolution, and it can be helpful for informing decisions about management. This thesis research has shown that historical and contemporary population patterns are more complex than were expected in *M. cephalus* populations. *Mugil cephalus* has a cosmopolitan distribution despite its preference for and dependence on the coastal environment. It is most likely that intermittent inter-oceanic migrations homogenized populations and reduced the likelihood of complete reproductive isolation. In New Zealand, the observed population genetic structure was geographically chaotic, admixture of four distinct populations was found and it was the first study to report different reproductive success as consequence of seasonal bottlenecks in juveniles for this
species. Adults spawned at different times and two spawning population were identified. It is possible that specific spawning grounds for different *M. cephalus* populations, human-based activities and reproductive behaviours such as philopatry could also play an important role in the chaotic distribution of allele frequencies.

The next steps would be to first, conduct regular surveys and attempt to identify the extent of the variation of genetic structure and test for correlations with environmental variables. Second, identify the specific spawning grounds to test for philopatry. And third, investigate whether human-based activities near estuaries have reduced the productive nursery grounds. Equally important is to understand the effect of fishing pressure in the genetic structure of this species. More investment into historical and contemporary connectivity patterns and genetic population structure, will improve our understanding about the movements and spatial delimitation of species populations and help managing fish stocks.

The findings of this research will make easier to identify *M. cephalus* stocks, which can be managed separately and will help to re-define the GMU management areas. Genetically differentiated populations should be managed differently to ensure the sustainability of the resource. The differences in the *Ne* identified in *M. cephalus* populations complementarily to the reproductive behaviour of the adults (different spawning groups and philopatry), and the seasonal changes in the *Ne* found in the juveniles should be taken in consideration for the management of each GMU area. This thesis research described the spatial and temporal genetic variation in *M. cephalus* in New Zealand and it has the potential for making an important contribution to securing a sustainable future for the kānae or grey mullet.

References

- Achaz, G. 2009. Frequency spectrum neutrality tests: One for all and all for one. Genetics 183:249–258.
- Anderson, E. C., and J. C. Garza. 2009. Estimation of population size with molecular genetic data. NOAA. Tech. Memo. 64.
- Anon. 1989. Effects of Commercial Fishing on the Fisheries of the Manukau Harbour and Lower Waikato River. Auckland.
- Antao, T., A. Lopes, R. J. Lopes, A. Beja-Pereira, and G. Luikart. 2008. LOSITAN: A workbench to detect molecular adaptation based on a F st -outlier method. BMC Bioinformatics 9:1–5.
- Aspinwall, N. 1974. Genetic Analysis of North American Populations of the Pink Salmon , Oncorhynchus gorbuscha, Possible Evidence for the Neutral Mutation-Random Drift Hypothesis. Evolution (N. Y). 28:295–305.
- Attenbrow, V. 2010. Sydney's Aboriginal past: Investigating the archaeological and historical records. UNSW Press, Sydney, NSW.
- Avise, J. C. 2004. Molecular markers, natural history, and evolution. Sinauer Associates, Inc., Sunderland.
- Avise, J. C. 2000. Phylogeography: the history and formation of species. Harvard University Press, Cambridge, Massachusetts.
- Avise, J. C., and F. J. Ayala. 2009. In the light of Evolution. Volume III: Two Centuries of Darwin. National Academy of Sciences, Washington, DC.
- Azuma, Y., Y. Kumazawa, M. Miya, K. Mabuchi, and M. Nishida. 2008. Mitogenomic evaluation of the historical biogeography of cichlids toward reliable dating of teleostean divergences. BMC Evol. Biol. 8:215.
- Bacheler, N. M., R. a. Wong, and J. a. Buckel. 2005. Movements and Mortality Rates of Striped Mullet in North Carolina. North Am. J. Fish. Manag. 25:361–373.
- Bailey, G., and J. Parkington. 1988. The Archaeology of prehistoric coastlines. Cambridge University Press, New york.
- Ball, A. O., G. R. Sedberry, M. S. Zatcoff, R. W. Chapman, and J. L. Carlin. 2000. Population structure of the wreckfish *Polyprion americanus* determined with microsatellite genetic markers. Mar. Biol. 137:1077–1090.
- Bandelt, H. J., P. Forster, and A. Röhl. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16:37–48.
- Bargelloni, L., J. A. Alarcon, M. C. Alvarez, E. Penzo, A. Magoulas, C. Reis, and T. Patarnello. 2003. Discord in the family Sparidae (Teleostei): Divergent phylogeographical patterns across the Atlantic-Mediterranean divide. J. Evol. Biol. 16:1149–1158.
- Barluenga, M., and A. Meyer. 2010. Phylogeography, colonization and population history of the Midas cichlid species complex (*Amphilophus* spp.) in the Nicaraguan crater lakes. BMC Evol. Biol. 10:326.
- Barluenga, M., K. N. Stölting, W. Salzburger, M. Muschick, and A. Meyer. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. Nature 439:719–723.
- Béarez, P., D. Jackson, and N. Mollaret. 2015. Early Archaic Fishing (12,600–9,200 cal yr BP) in the Semiarid North Coast of Chile. J. Isl. Coast. Archaeol. 10:133–148.

- Beaumont, M. A., and R. A. Nichols. 1996. Evaluating Loci for Use in the Genetic Analysis of Population Structure. Proc. R. Soc. London B Biol. Sci. 263:1619–1626.
- Beerli, P. 2009. How to use Migrate or why are Markov chain Monte Carlo programs difficult to use?
 BT Population Genetics for Animal Conservation. P. *in* G. Bertorelle, M. W. Bruford, H. C. Hauffe, A. Rizzoli, and C. Vernesi, eds. Cambridge University Press, Cambridge.
- Beerli, P., and J. Felsenstein. 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics 152.
- Beerli, P., and J. Felsenstein. 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. Proc Nat Acad Sci 98.
- Bekkevold, D. 2002. Male reproductive competition in spawning aggregations of cod (*Gadus morhua*, L.). Mol. Ecol. 11:91–102.
- Bekkevold, D., C. André, T. G. Dahlgren, L. A. W. Clausen, H. Mosegaard, G. R. Carvalho, T. B. Christensen, E. Norlinder, D. E. Ruzzante, D. Bekkevold, C. Andrt, T. G. Dahlgren, L. A. W. Clausen, E. Torstensen, H. Mosegaard, G. R. Carvalho, T. B. Christensen, I. E. Norlinder, and D. E. Ruzzante. 2005. Environmental correlates of population differentiation in Atlantic Herring. Evolution (N. Y). 59:2656–2668.
- Bekkevold, D., L. A. W. Clausen, S. Mariani, C. Andre, T. B. Christensen, and H. Mosegaard. 2007. Divergent origins of sympatric herring population components determined using genetic mixture analysis. Mar. Ecol. Prog. Ser. 337:187–196.
- Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste, and F. Bonhomme. 2004. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Université de Montpellier, Montpellier.
- Benson, E. P. 2012. The Worlds of the Moche on the North Coast of Peru. University of Texas Press, Project MUSE, Austin.
- Bentzen, P., and I. R. Bradbury. 2016. Don't bet against the natal homing abilities of marine fishes. Mol. Ecol. 25:2691–2692.
- Bernal-Ramirez, J. H., G. J. Adcock, L. Hauser, G. R. Carvalho, and P. J. Smith. 2003. Temporal stability of genetic population structure in the New Zealand snapper, *Pagrus auratus*, and relationship to coastal currents. Mar. Biol. 142:567–574.
- Betancur-R, R., P. Arturo Acero, H. Duque-Caro, and S. R. Santos. 2010. Phylogenetic and morphologic analyses of a coastal fish reveals a marine biogeographic break of terrestrial origin in the Southern Caribbean. PLoS One 5:1–10.
- Blaber, S. J. M. 2013. Fishes and fisheries in tropical estuaries: The last 10 years. Estuar. Coast. Shelf Sci. 135:57–65.
- Blaber, S. J. M. 1976. The food and ecology of Mugilidaein the ST. Lucia lake system. Biol. J. Linn. Soc. 8:267–277.
- Blel, H., J. Panfili, B. Guinand, P. Berrebi, K. Said, and J. D. Durand. 2010. Selection footprint at the first intron of the Prl gene in natural populations of the flathead mullet (*Mugil cephalus*, L. 1758). J. Exp. Mar. Bio. Ecol. 387:60–67.
- Bolnick, D. I., and B. M. Kirkpatrick. 2007. Sympatric speciation: models and empirical evidence. Annu. Rev. Ecol. Evol. Syst. 38:459–487.
- Bonanomi, S., N. Overgaard Therkildsen, A. Retzel, R. Berg Hedeholm, M. W. Pedersen, D. Meldrup, C. Pampoulie, J. Hemmer-Hansen, P. Grønkjær, and E. E. Nielsen. 2016. Historical DNA documents long-distance natal homingin marine fish. Mol. Ecol. 25:2727–2734.

Borowski, Z., M. Swistocka, M. Matosiuk, P. Mirski, K. Krysiuk, M. Czajkowska, A. Borkowska, and

M. Ratkiewicz. 2016. Purifying selection, density blocking and unnoticed mitochondrial DNA diversity in the red deer, Cervus elaphus. PLoS One 11:1–17.

- Bradbury, I. R., and P. Bentzen. 2007. Non-linear genetic isolation by distance: Implications for dispersal estimation in anadromous and marine fish populations. Mar. Ecol. Prog. Ser. 340:245–257.
- Briggs, J. C. 1960. Fishes of Worldwide (Cirumtropical) Distribution. Copeia 1960:171–180.
- Broquet, T., F. Viard, and J. M. Yearsley. 2013. Genetic drift and collective dispersal can result in chaotic genetic patchiness. Evolution (N. Y). 67:1660–1675.
- Buchheister, A., C. F. Bonzek, J. Gartland, and R. J. Latour. 2013. Patterns and drivers of the demersal fish community of chesapeake bay. Mar. Ecol. Prog. Ser. 481:161–180.
- Buckley, H. L., A. M. Paterson, and R. H. Cruickshank. 2013. The founder space race : a response to Waters et al . Trends Ecol. Evol. 28:189–190.
- Campton, D. E., and B. Mahmoudi. 1991. Allozyme variation and population structure of striped mullet (*Mugil cephalus*) in Florida. Copeia 1991:485–492.
- Cardona, L. 2000. Effects of Salinity on the Habitat Selection and Growth Performance of Mediterranean Flathead Grey Mullet *Mugil cephalus* (Osteichthyes, Mugilidae). Estuar. Coast. Shelf Sci. 50:727–737.
- Carvalho, G. R., and L. Hauser. 1994. Molecular genetics and the stock concept in fisheries. Rev. Fish Biol. Fish. 4:326–350.
- Chang, C. W., and Y. Iizuka. 2012. Estuarine use and movement patterns of seven sympatric Mugilidae fishes: The Tatu Creek estuary, central western Taiwan. Estuar. Coast. Shelf Sci. 106:121–126.
- Chang, C. W., W. N. Tzeng, and Y. C. Lee. 2000. Recruitment and hatching dates of grey mullet (*Mugil cephalus* L.) juveniles in the Tanshui estuary of northwest Taiwan. Zool. Stud. 39:99–106.
- Christie, M. R., D. W. Johnson, C. D. Stallings, and M. A. Hixon. 2010. Self-recruitment and sweepstakes reproduction amid extensive gene flow in a coral-reef fish. Mol. Ecol. 19:1042–1057.
- Christie, M. R., M. L. Marine, R. a French, R. S. Waples, and M. S. Blouin. 2012. Effective size of a wild salmonid population is greatly reduced by hatchery supplementation. Heredity. 109:254–260.
- Clancy, G. P. 2005. The diet of the Osprey (*Pandion haliaetus*) on the north coast of New South Wales. Emu 105:87–91.
- Cordell, H. J. 2002. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum. Mol. Genet. 11:2463–2468.
- Cornuet, J. M., and G. Luikart. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014.
- Cowen, R. K. 2007. Population Connectivity in Marine Systems. Oceanography 20:14-21.
- Cowen, R. K. 2006. Scaling of Connectivity in Marine Populations. Science 311:522-527.
- Cox, N. J., M. L. Frigge, D. L. Nicolae, P. Concannon, C. L. Hanis, G. I. Bell, and A. Kong. 1999. Loci on chromosomes 2 (NIDDM1) and 15 interact to increase susceptibility to diabetes in Mexican Americans. Nat. Genet. 21:213–215.
- Crosetti, D., J. C. Avise, F. Placidi, A. R. Rossi, and L. Sola. 1993. Geographic Variability in the Gray Mullet Mugil-Cephalus - Preliminary-Results of Mtdna and Chromosome Analyses. Aquaculture 111:95–101.
- Cushman, S. A., A. J. Shirk, and E. L. Landguth. 2013. Landscape genetics and limiting factors.

Conserv. Genet. 14:263-274.

De Queiroz, K. 2005. Different species problems and their resolution. BioEssays 27:1263–1269.

- De Silva, S. S., and M. J. S. Wijeyaratne. 1977. Studies on the biology of young grey mullet, *Mugil cephalus* L. II. Food and Feeding. Aquaculture 12:157–167.
- Dean, A., Ma. Morris, J. Stufken, and D. Bingham. 2015. Handbook of Design and Analysis of Experiments.
- Do, C., R. S. Waples, D. Peel, G. M. Macbeth, B. J. Tillett, and J. R. Ovenden. 2014. NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. Mol. Ecol. Resour. 14:209–214.
- Donoghue, M. J. 1985. A Critique of the Biological Species Concept and Recommendations for a Phylogenetic Alternative. Bryologist 88:172–181.
- Driscoll, D. A. 1998. Genetic structure, metapopulation processes and evolution influence the conservation strategies for two endangered frog species. Biol. Conserv. 83:43–54.
- Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7:214.
- Durand, J. D., W. J. Chen, K. N. Shen, C. Fu, and P. Borsa. 2012a. Genus-level taxonomic changes implied by the mitochondrial phylogeny of grey mullets (Teleostei: Mugilidae). Comptes Rendus - Biol. 335:687–697.
- Durand, J. D., K. N. Shen, W. J. Chen, B. W. Jamandre, H. Blel, K. Diop, M. Nirchio, F. J. Garcia de León, A. K. Whitfield, C. W. Chang, and P. Borsa. 2012b. Systematics of the grey mullets (Teleostei: Mugiliformes: Mugilidae): Molecular phylogenetic evidence challenges two centuries of morphology-based taxonomy. Mol. Phylogenet. Evol. 64:73–92.
- Earl, D. A., and B. M. Von Holdt. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4:359–361.
- Edmonds, C. a, A. S. Lillie, and L. L. Cavalli-Sforza. 2004. Mutations arising in the wave front of an expanding population. Proc. Natl. Acad. Sci. 101:975–979.
- Elliott, M., and A. K. Whitfield. 2011. Challenging paradigms in estuarine ecology and management. Estuar. Coast. Shelf Sci. 94:306–314.
- Elliott, M., A. K. Whitfield, I. C. Potter, S. J. M. Blaber, D. P. Cyrus, F. G. Nordlie, and T. D. Harrison. 2007. The guild approach to categorizing estuarine fish assemblages: A global review. Fish Fish. 8:241–268.
- Elmer, K. R., S. Fan, H. Kusche, M. L. Spreitzer, A. F. Kautt, P. Franchini, and A. Meyer. 2014. Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. Nat Commun 5:5168.
- Eschmeyer, W., and R. Fricke. 2011. Catalog of fishes, electronic version. http://research.calacademy.org/research/ichthyology/catalog/fishcatmain.asp.
- Esselstyn, J. A., B. J. Evans, J. L. Sedlock, F. a. Anwarali Khan, and L. R. Heaney. 2012. Single-locus species delimitation: a test of the mixed Yule-coalescent model, with an empirical application to Philippine round-leaf bats. Proc. R. Soc. B Biol. Sci. 279:3678–3686.
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. 14:2611–2620.
- Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10:564–567.

- Ezard, T., T. Fujisawa, and T. G. Barraclough. 2009. Splits: SPecies' LImits by Threshold Statistics. R package version 1.0-11/r29. http://R-Forge.R-project.org/projects/splits/.
- Fahrig, L., and G. Merriam. 1985. Habitat patch connectivity and population survival. Ecology 66:1762–1768.
- Falniowski, A., and M. Szarowska. 2012. Sequence-based species delimitation in the Balkan Bythinella Moquin-Tandon, 1856 (Gastropoda: Rissooidea) with general mixed yule coalescent model. Folia Malacol. 20:111–120.
- Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164:1567–1587.
- Fitzpatrick, B. M., J. A. Fordyce, and S. Gavrilets. 2008. What, if anything, is sympatric speciation? J. Evol. Biol. 21:1452–1459.
- Flowers, J. M., S. C. Schroeter, and R. S. Burton. 2002. The recruitment sweepstakes has many winners: genetic evidence from the sea urchin Strongylocentrotus purpuratus. Evolution 56:1445–1453.
- Fraga, E., H. Schneider, M. Nirchio, E. Santa-Brigida, L. F. Rodrigues-Filho, and I. Sampaio. 2007. Molecular phylogenetic analyses of mullets (Mugilidae, Mugiliformes) based on two mitochondrial genes. J. Appl. Ichthyol. 23:598–604.
- Francis, M. P., M. A. Morrison, J. Leathwick, and C. Walsh. 2011. Predicting patterns of richness, occurrence and abundance of small fish in New Zealand estuaries. Mar. Freshw. Res. 62:1327– 1341.
- Francis, M. P., M. A. Morrison, J. Leathwick, C. Walsh, and C. Middleton. 2005. Predictive models of small fish presence and abundance in northern New Zealand harbours. Estuar. Coast. Shelf Sci. 64:419–435.
- Frankel, O. H. 1974. Genetic conservation: our evolutionary responsibility. Genetics 78:53–65.
- Fraser, D. J., and L. Bernatchez. 2005. Allopatric origins of sympatric brook charr populations: Colonization history and admixture. Mol. Ecol. 14:1497–1509.
- Fraser, D. J., M. M. Hansen, S. Ostergaard, N. Tessier, M. Legault, and L. Bernatchez. 2007. Comparative estimation of effective population sizes and temporal gene flow in two contrasting population systems. Mol. Ecol. 16:3866–3889.
- Fu, Y. X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking, and background selection. Genetics 147.
- Fujisawa, T., and T. G. Barraclough. 2013. Delimiting species using single-locus data and the generalized mixed yule coalescent approach: A revised method and evaluation on simulated data sets. Syst. Biol. 62:707–724.
- Fury, C. A., and P. L. Harrison. 2011. Seasonal variation and tidal influences on estuarine use by bottlenose dolphins (*Tursiops aduncus*). Estuar. Coast. Shelf Sci. 93:389–395.
- Gaither, M. R., B. W. Bowen, T.-R. Bordenave, L. A. Rocha, S. J. Newman, J. A. Gomez, L. van Herwerden, and M. T. Craig. 2011. Phylogeography of the reef fish *Cephalopholis argus* (Epinephelidae) indicates Pleistocene isolation across the indo-pacific barrier with contemporary overlap in the coral triangle. BMC Evol. Biol. 11:189.
- Galtier, N., F. Depaulis, and N. H. Barton. 2000. Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. Genetics 155:981–987.
- Gardner, J., J. J. Bell, H. Constable, D. Hannan, P. Ritchie, and G. Zuccarello. 2010. Multi-species coastal marine connectivity: a literature review with recommendations for further research. New Zeal. Aquat. Environ. Biodivers. Report, MPI 1–47.

- Garza, J. C., and E. G. Williamson. 2001. Detection of reduction in population size using data from microsatellite loci. Mol. Ecol. 10:305–318.
- Gernhard, T., K. Hartmann, and M. Steel. 2008. Stochastic properties of generalised Yule models, with biodiversity applications. J. Math. Biol. 57:713–735.
- Gobalet, K. W., T. A. Wake, and K. L. Hardin. 2005. Archaeological Record of Native Fishes of the Lower Colorado River : How To Identify Their Remains. West. North Am. Nat. 65:335–344.
- Górski, K., C. De Gruijter, and R. Tana. 2015. Variation in habitat use along the freshwater-marine continuum by grey mullet Mugil cephalus at the southern limits of its distribution. J. Fish Biol. 87:1059–1071.
- Goudet, J. 1995. FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. J. Hered. 86:485–486.
- Goudet, J. 1999. PCA-GEN, Version 1.2. http://www.unil.ch/izea/softwares/pcagen.
- Grant, P. R., and B. R. Grant. 2009. The secondary contact phase of allopatric speciation in Darwin's finches. Proc. Natl. Acad. Sci. 106:20141–20148.
- Grant, W. S., and B. W. Bowen. 1998. Shallow Population Historiesin Deep Evolutionnary Lineages of Marine Fishhes: Insights From Sardines and Anchovies and Lesson for Coservatio. J. Hered. 89:415–426.
- Grosberg, R., and C. W. Cunningham. 2001. Genetic Structure in the Sea: From Populations to Communities. Mar. Community Ecol. (Eds. M. D. Bertness, S. Gaines M. E Hay) Sinauer As:61–84.
- Gu, Z., L. Gu, R. Eils, M. Schlesner, and B. Brors. 2014. Circlize implements and enhances circular visualization in R. Bioinformatics 30:2811–2812.
- Guo, S. W., and E. A. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361–372.
- Hallatschek, O., and D. R. Nelson. 2008. Gene surfing in expanding populations. Theor. Popul. Biol. 73:158–170.
- Harrison, I. J. 1995. Mugilidae: Lisas. In: Guia FAO para Identification de Especies para lo Fines de la Pesca. Pacifico Centro-Oriental. Pp. 1293–1298 in W. Fischer, F. Krupp, W. Schneider, C. Sommer, K. E. Carpenter, and V. H. Niem, eds. Vol. III Part 2. FAO.
- Hartill, B. 2004. Characterisation of the commercial flatfish, grey mullet, and rig fisheries in the Kaipara Harbour. New Zeal. Fish. Assess. Rep. 2004/1:23 p.
- Hastings, A., and S. Harrison. 1994. Metapopulation dynamics and genetics. Annu. Rev. Ecol. Syst. 25:167–188.
- Hauser, L., G. J. Adcock, P. J. Smith, J. H. B. Ramiréz, and G. R. Carvalho. 2002. Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). Pnas 99:11742–7.
- Hauser, L., and G. R. Carvalho. 2008. Paradigm shifts in marine fisheries genetics: Ugly hypotheses slain by beautiful facts. Fish Fish. 9:333–362.
- Hedgecock, D. 1994. Temporal and Spatial Genetic Structure of Marine Animal Populations. CalCOFI Reports 35:73–81.
- Hedgecock, D., P. H. Barber, and S. Edmands. 2007a. Genetic Approaches to Measuring Connectivity. Oceanography 20:70–79.
- Hedgecock, D., S. Launey, A. I. Pudovkin, Y. Naciri, S. Lapègue, and F. Bonhomme. 2007b. Small

effective number of parents (Nb) inferred for a naturally spawned cohort of juvenile European flat oysters *Ostrea edulis*. Mar. Biol. 150:1173–1182.

- Hedgecock, D., and A. I. Pudovkin. 2011. Sweepstakes reproductive success in highly fecund marine fish and shellfish: a review and commentary. Bull. Mar. Sci. 87:971–1002.
- Hellberg, M. E. 2009. Gene Flow and Isolation among Populations of Marine Animals. Annu. Rev. Ecol. Evol. Syst. 40:291–310.
- Hendry, A. P., and T. Day. 2005. Population structure attributable to reproductive time: Isolation by time and adaptation by time. Mol. Ecol. 14:901–916.
- Hennig, W. 1966. Phylogenetic Systematics. University of Illinois Press.
- Heras, S., M. I. Roldán, M. G. Castro, and M. B. Cousseau. 2007. *Mugil cephalus*: cosmopolitan species or species complex? Rapp. Comm. int. Mer Médit. 38:499.
- Hewitt, C. L., J. Willing, A. Bauckham, a. M. Cassidy, C. M. S. Cox, L. Jones, and D. M. Wotton. 2004. New Zealand marine biosecurity: Delivering outcomes in a fluid environment. New Zeal. J. Mar. Freshw. Res. 38:429–438.
- Hewitt, G. 1999. Post-glacial re-colonization of European biota. Biol. J. Linn. Soc. 68:87-112.
- Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. Nature 405:907–913.
- Hey, J., and R. Nielsen. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proc. Natl. Acad. Sci. 104:2785–2790.
- Hey, J., and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudobscura* and *D. persimilis*. Genetics 167:747–760.
- Hickerson, M. J., B. C. Carstens, J. Cavender-Bares, K. A. Crandall, C. H. Graham, J. B. Johnson, L. Rissler, P. F. Victoriano, and A. D. Yoder. 2010. Phylogeography's past, present, and future: 10 years after Avise, 2000. Mol. Phylogenet. Evol. 54:291–301. Elsevier Inc.
- Hickey, A. J. R., S. D. Lavery, D. A. Hannan, C. S. Baker, and K. D. Clements. 2009. New Zealand triplefin fishes (family Tripterygiidae): Contrasting population structure and mtDNA diversity within a marine species flock. Mol. Ecol. 18:680–696.
- Hoffman, E. a, F. W. Schueler, and M. S. Blouin. 2004. Effective population sizes and temporal stability of genetic structure in *Rana pipiens*, the northern leopard frog. Evol. Int J Org Evol. 58:2536–2545.
- Hogan, J. D., R. J. Thiessen, and D. D. Heath. 2010. Variability in connectivity indicated by chaotic genetic patchiness within and among populations of a marine fish. Mar. Ecol. Prog. Ser. 417:263– 275.
- Horne, J. B., L. Van Herwerden, S. Abellana, and J. L. Mcilwain. 2013. Observations of migrant exchange and mixing in a coral reef fish metapopulation link scales of marine population connectivity. J. Hered. 104:532–546.
- Hsu, C., Y. Han, and W. Tzeng. 2007. Evidence of Flathead Mullet *Mugil cephalus* L . Spawning in Waters Northeast of Taiwan. Zool. Stud. 46:717–725.
- Huang, C. S., C. F. Weng, and S. C. Lee. 2001. Distinguishing two types of gray mullet, *Mugil cephalus* L. (Mugiliformes: Mugilidae), by using glucose-6-phosphate isomerase (GPI) allozymes with special reference to enzyme activities. J. Comp. Physiol. - B Biochem. Syst. Environ. Physiol. 171:387–394.
- Huang, W., A. Massouras, Y. Inoue, J. Peiffer, M. Ràmia, A. Tarone, L. Turlapati, and T. Zichner.

2014. Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. Genome Res. 24:1193–1208.

- Hubisz, M. J., D. Falush, M. Stephens, and J. K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. Mol. Ecol. Resour. 9:1322–1332.
- Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.
- Hughes, J. M., D. J. Schmidt, J. I. Macdonald, J. A. Huey, and D. A. Crook. 2014. Low interbasin connectivity in a facultatively diadromous fish: Evidence from genetics and otolith chemistry. Mol. Ecol. 23:1000–1013.
- Hung, C.-M., and D. Shaw. 2006. The Impact of Upstream Catch and Global Warming on the Grey Mullet Fishery in Taiwan: A Non-cooperative Game Analysis. Mar. Resour. Econ. 21:285–300.
- Hurst, R., J. Devine, and A. Mckenzie. 2009. Fish abundance and climate trends in New Zealand. Environment.
- Hwang, S.-Y., C.-L. Kuo, and S. Tanaka. 1990. Stock Assessment by Grey Mullet, *Mugil cephalus* in Taiwan by Cohort Analysis. Nippon Suisan Gakkaishi 56:1955–1963.
- Hyde, J. R., K. E. Underkoffler, and M. A. Sundberg. 2014. DNA barcoding provides support for a cryptic species complex within the globally distributed and fishery important opah (*Lampris guttatus*). Mol. Ecol. Resour. 14:1239–1247.
- Iacchei, M., T. Ben-Horin, K. A. Selkoe, C. E. Bird, F. J. García-Rodríguez, and R. J. Toonen. 2013. Combined analyses of kinship and FST suggest potential drivers of chaotic genetic patchiness in high gene-flow populations. Mol. Ecol. 22:3476–3494.
- Irwin, D. E. 2002. Phylogeographic Breaks Without Geographic Barriers To Gene Flow. Evolution (N. Y). 56:2383.
- Jamandre, B. W., J. D. Durand, and W. N. Tzeng. 2014. High sequence variations in mitochondrial DNA control region among worldwide populations of flathead mullet *Mugil cephalus*. Int. J. Zool. 2014.
- Jamandre, B. W., J. D. Durand, and W. N. Tzeng. 2009. Phylogeography of the flathead mullet *Mugil cephalus* in the north-west Pacific as inferred from the mtDNA control region. J. Fish Biol. 75:393–407.
- Jiggins, C. D. 2006. Sympatric Speciation: Why the Controversy? Curr. Biol. 16:333–334.
- Johnson, D. W., M. R. Christie, J. Moye, and M. A. Hixon. 2011. Genetic correlations between adults and larvae in a marine fish: Potential effects of fishery selection on population replenishment. Evol. Appl. 4:621–633.
- Johnson, M. S. ., and R. Black. 2012. Pattern Beneath the Chaos : The Effect of Recruitment on Genetic Patchiness in an Intertidal Limpet. Evolution 38:1371–1383.
- Johnson, M. S., and R. Black. 1982. Chaotic genetic patchiness in an intertidal limpet, *Siphonaria* sp. Mar. Biol. 70:157–164.
- Johnson, M. S., and R. Black. 1984. The Wahlund effect and the geographical scale of variation in the intertidal limpet *Siphonaria* sp. Mar. Biol. 79:295–302.
- Jones, G. P., S. Planes, and S. R. Thorrold. 2005. Coral reef fish larvae settle close to home. Curr. Biol. 15:1314–1318.
- Jones, O. R., and J. Wang. 2010. COLONY: A program for parentage and sibship inference from multilocus genotype data. Mol. Ecol. Resour. 10:551–555.

- Jorde, P. E., and N. Ryman. 2007. Unbiased estimator for genetic drift and effective population size. Genetics 177:927–935.
- Jørgensen, H. B. H., M. M. Hansen, D. Bekkevold, D. E. Ruzzante, and V. Loeschcke. 2005. Marine landscapes and population genetic structure of herring (*Clupea harengus* L.) in the Baltic Sea. Mol. Ecol. 14:3219–3234.
- Jost, L. 2008. GST and its relatives do not measure differentiation. Mol. Ecol. 17:4015–4026.
- Joyce, D. a, D. H. Lunt, R. Bills, G. F. Turner, C. Katongo, N. Duftner, C. Sturmbauer, and O. Seehausen. 2005. An extant cichlid fish radiation emerged in an extinct Pleistocene lake. Nature 435:90–95.
- Jullien, N. 2013. AmplifX v1.7.0. http://crn2m.univ-mrs.fr/pub/amplifx-dist.
- Kalinowski, S. T. 2005. Do polymorphic loci require large sample sizes to estimate genetic distances? Heredity 94:33–36.
- Kamath, P. L., M. A. Haroldson, G. Luikart, D. Paetkau, C. Whitman, and F. T. Van Manen. 2015. Multiple estimates of effective population size for monitoring a long-lived vertebrate: An application to Yellowstone grizzly bears. Mol. Ecol. 24:5507–5521.
- Ke, H.-M., W.-W. Lin, and H.-W. Kao. 2009. Genetic diversity and differentiation of gray mullet (*Mugil cephalus*) in the coastal waters of Taiwan. Zoolog. Sci. 26:421–8.
- Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes, and A. Drummond. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.
- Kekkonen, M., and P. D. N. Hebert. 2014. DNA barcode-based delineation of putative species: Efficient start for taxonomic workflows. Mol. Ecol. Resour. 14:706–715.
- Kemp, W. M., W. R. Boynton, J. E. Adolf, D. F. Boesch, W. C. Boicourt, G. Brush, J. C. Cornwell, T. R. Fisher, P. M. Glibert, J. D. Hagy, L. W. Harding, E. D. Houde, D. G. Kimmel, W. D. Miller, R. I. E. Newell, M. R. Roman, E. M. Smith, and J. C. Stevenson. 2005. Eutrophication of Chesapeake Bay: Historical trends and ecological interactions. Mar. Ecol. Prog. Ser. 303:1–29.
- Kesteven, G. L. 1953. Further results of tagging sea mullet, *Mugil cephalus* Linnaeus, on the eastern Australian coast. Mar. Freshw. Res. 4:251–306.
- Kingman, J. F. C. 2000. Perspectives Anecdotal, Historical and Critical Commentaries on Genetics Origins of the Coalescent: 1974–1982. Genetics 156:1461–1463.
- Knowlton, N., J. L. Mate, H. M. Guzman, R. Rowan, and J. Jara. 1997. Direct evidence for reproductive isolation among the three species of the *Montastraea annularis* complex in Central America (Panama and Honduras). Mar. Biol. 127:705–711.
- Kotoulas, G., A. Magoulas, N. Tsimenides, and E. Zouros. 1995. Marked mitochondrial DNA differences between Mediterranean and Atlantic populations of the swordfish, *Xiphias gladius*. Mol. Ecol. 4:473–481.
- Krück, N. C., D. I. Innes, and J. R. Ovenden. 2013. New SNPs for population genetic analysis reveal possible cryptic speciation of eastern Australian sea mullet (*Mugil cephalus*). Mol. Ecol. Resour. 13:715–725.
- Laikre, L., S. Palm, and N. Ryman. 2005. Genetic population structure of fishes: implications for coastal zone management. Ambio 34:111–119.
- Larned, S., M. Unwin, H. Mcmillan, T. Snelder, G. McBride, and P. Verburg. 2015. Analysis of water quality in New Zealand lakes and rivers. Natl. Inst. Water Atmos. Rep. 2015-03:107.

- Lawrence, R., D. M. Evans, A. P. Morris, X. Ke, S. Hunt, M. Paolucci, J. Ragoussis, P. Deloukas, D. Bentley, and L. R. Cardon. 2005. Genetically indistinguishable SNPs and their influence on inferring the location of disease-associated variants. Genome Res. 15:1503–1510.
- Lawson, E., and A. Jimoh. 2010. Aspects of the biology of grey mullet, *Mugil cephalus*, in Lagos lagoon, Nigeria. AACL Bioflux 3:181–194.
- Lehodey, P., J. Alheit, M. Barange, T. Baumgartner, G. Beaugrand, K. Drinkwater, J. M. Fromentin, S. R. Hare, G. Ottersen, R. I. Perry, C. Roy, C. D. van der Lingen, and F. Werner. 2006. Climate variability, fish, and fisheries. J. Clim. 19:5009–5030.
- Leigh, J. W., and D. Bryant. 2015. POPART: Full-feature software for haplotype network construction. Methods Ecol. Evol. 6:1110–1116.
- Lewallen, E. A., R. L. Pitman, S. L. Kjartanson, and N. R. Lovejoy. 2011. Molecular systematics of flyingfishes (Teleostei: Exocoetidae): Evolution in the epipelagic zone. Biol. J. Linn. Soc. 102:161–174.
- Li, G., and D. Hedgecock. 1998. Genetic heterogeneity, detected by PCR-SSCP, among samples of larval Pacific oysters (*Crassostrea gigas*) supports the hypothesis of large variance in reproductive success. Can. J. Fish. Aquat. Sci. 55:1025–1033.
- Librado, P., and J. Rozas. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452.
- Liordos, V., and V. Goutner. 2009. Sexual differences in the diet of great cormorants *Phalacrocorax carbo sinensis* wintering in Greece. Eur. J. Wildl. Res. 55:301–308.
- Liu, J., C. L. Brown, and T. Yang. 2009a. Population genetic structure and historical demography of grey mullet *Mugil cephalus* along the coast of China inferred by analysis of the mitochondrial control region. Biochem. Syst. Ecol. 37:556–566.
- Liu, J. Y., C. L. Brown, and T. B. Yang. 2010. Phylogenetic relationships of mullets (Mugilidae) in China Seas based on partial sequences of two mitochondrial genes. Biochem. Syst. Ecol. 38:647– 655.
- Liu, J. Y., Z. R. Lun, J. Bin Zhang, and T. B. Yang. 2009b. Population genetic structure of striped mullet, *Mugil cephalus*, along the coast of China, inferred by AFLP fingerprinting. Biochem. Syst. Ecol. 37:266–274.
- Livi, S., L. Sola, and D. Crosetti. 2011. Phylogeographic relationships among worldwide populations of the cosmopolitan marine species, the striped gray mullet (*Mugil cephalus*), investigated by partial cytochrome b gene sequences. Biochem. Syst. Ecol. 39:121–131. Elsevier Ltd.
- Luikart, G., W. B. Sherwin, B. M. Steele, and F. W. Allendorf. 1998. Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. Mol. Ecol. 7:963–974.
- Lunney, D., P. Hutchings, and H. Dieter. 2010. The Natural History of Sydney. Royal Zoological Society of New South Wales.
- Lynch, M., and K. Ritland. 1999. Estimation of pairwise relatedness with molecular markers. Genetics 152:1753–1766.
- Maes, G. E., J. M. Pujolar, B. Hellemans, and F. A. M. Volckaert. 2006. Evidence for isolation by time in the European eel (*Anguilla anguilla* L.). Mol. Ecol. 15:2095–2107.
- Manzon, L. A. 2002. The Role of Prolactin in Fish Osmoregulation: A Review. Gen. Comp. Endocrinol. 125:291–310.
- Marais, J. F. K. 1980. Aspects of food intake, food selection, and alimentary cacal morphology of *Mugil cephalus* (Linnaeus, 1958), *Liza tricuspidens* (Smith, 1935), *L. richardsoni* (Smith, 1846), and *L.*

dumerili (Steindachner, 1869). J. Exp. Mar. Bio. Ecol. 44:193-209.

- Marko, P. B., and M. W. Hart. 2012. Retrospective coalescent methods and the reconstruction of metapopulation histories in the sea. Evol. Ecol. 26:291–315.
- Marko, P. B., J. M. Hoffman, S. A. Emme, T. M. McGovern, C. C. Keever, and L. Nicole Cox. 2010. The "expansion-Contraction" model of Pleistocene biogeography: Rocky shores suffer a sea change? Mol. Ecol. 19:146–169.
- Marth, G. T., E. Czabarka, J. Murvai, and S. T. Sherry. 2004. The Allele Frequency Spectrum in Genome-Wide Human Variation Three Large World Populations. Genetics 372:351–372.
- Martucci, O., L. Pietrelli, and C. Consiglio. 1993. Fish otoliths as indicators of the cormorant *Phalacrocorax carbo* diet (Aves, Pelecaniformes). Bolletino di Zool. 60:393–396.
- Maruyama, T., and P. A. Fuerstt. 1985. Population bottlenecks and nonequilibrium models in population genetics. Chapter 11. Number of alleles in a small population that was formed by a recent bottleneck. Popul. (English Ed). 675–689.
- McDonough, C. J., W. A. Roumillat, and C. A. Wenner. 2005. Sexual differentiation and gonad development in striped mullet (*Mugil cephalus* L.) from South Carolina estuaries. Fish. Bull. 103:601–619.
- McDowall, R. M. 1976. The role of estuaries in the life cycles of fishes in New Zealand. Proc. New Zeal. Ecol. Soc. 23:27–32.
- McKenzie, J., L. Paul, and C. Omaolagain. 1999. Length and age composition of commercial grey mullet landings from the west coast setnet fishery (GMU 1), 1997-98.
- McKenzie, J. R., and M. Vaughan. 2008. CPUE analysis and characterisation of grey mullet (*Mugil cephalus*) setnets fisheries in Fishstock GMU1 between 1989 and 2006. New Zeal. Fish. Assess. Rep. 57.
- Meirmans, P. G., and P. W. Hedrick. 2011. Assessing population structure: FST and related measures. Mol. Ecol. Resour. 11:5–18.
- Menotti, and A. O'Sullivan. 2013. The Oxford handbook of wetland archaeology. Oxford University Press.
- Michaelis, H. 1993. Food items of the grey mullet Mugil cephalus in the Banc d'Arguin area (Mauritania). Hydrobiologia 258:175–183.
- Miggiano, E., R. E. Lyons, Y. Li, L. M. Dierens, D. Crosetti, and L. Sola. 2005. Isolation and characterization of microsatellite loci in the striped mullet, *Mugil cephalus*. Mol. Ecol. Notes 5:323–326.
- Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Gatew. Comput. Environ. Work. GCE.
- Miya, M., A. Kawaguchi, and M. Nishida. 2001. Mitogenomic exploration of higher teleostean phylogenies: a case study for moderate-scale evolutionary genomics with 38 newly determined complete mitochondrial DNA sequences. Mol. Biol. Evol. 18:1993–2009.
- Moberg, P. E., and R. S. Burton. 2000. Genetic heterogeneity among adult and recruit red sea urchins, *Strongylocentrotus franciscanus*. Mar. Biol. 136:773–784.
- Monaghan, M. T., R. Wild, M. Elliot, T. Fujisawa, M. Balke, D. J. G. Inward, D. C. Lees, R. Ranaivosolo, P. Eggleton, T. G. Barraclough, and A. P. Vogler. 2009. Accelerated species Inventory on Madagascar using coalescent-based models of species Delineation. Syst. Biol. 58:298–311.

- Morjan, C. L., and L. H. Rieseberg. 2004. How species eveolve collectively: implications of gene flow and selection for the spread and advantgeous alleles. Mol. Ecol. 13:1341–1356.
- Morley, A., Y. Rosenthal, and P. DeMenocal. 2014. Ocean-atmosphere climate shift during the mid-tolate Holocene transition. Earth Planet. Sci. Lett. 388:18–26. Elsevier B.V.
- Morrison, M. 2005. An information review of the natural marine features and ecology of Northland An information review of the natural marine features and ecology of Northland. Natl. Inst. Water Atmos. Rep. AKL2005-30:174.
- Morrison, M. A., E. Jones, D. P. Parsons, and C. Grant. 2014a. Habitats and areas of particular significance for coastal finfish fisheries management in New Zealand: A review of concepts and current knowledge, and suggestions for future research. New Zeal. Aquat. Environ. Biodivers. Rep. 125:205.
- Morrison, M. A., M. L. Lowe, C. Grant, J. Reed, G. Carbines, P. J. Smith, S. J. Bury, and J. Brown. 2014b. Seagrass meadows as biodiversity and productivity hotspots. New Zeal. Aquat. Environ. Biodivers. Rep. 137:147.
- Morrison, M. A., M. L. Lowe, E. G. Jones, L. Makey, U. Shankar, N. Usmar, A. Miller, S. M., and C. Middleton. 2014c. Habitats of particular significance for fisheries management: The Kaipara Harbour. New Zeal. Aquat. Environ. Biodivers. Rep. 129:169.
- Morrison, M. A., M. L. Lowe, D. Parsons, N. Usmar, and I. McLeod. 2009. A review of land-based effects on coastal fisheries and supporting biodiversity in New Zealand. Aquatic Biodiversity and Biosecurity. Aquat. Biodivers. Biosecurity Rep. 37:100.
- Morrison, M. A., J. McKenzie, B. Gillanders, and I. Tuck. 2016. Can otolith chemistry predict the natal origins of grey mullet (*Mugil cephalus*)? New Zeal. Fish. Assess. Report, MPI 15:68.
- MPI. 2016. Fisheries Assessment Plenary May 2016: Stock Assessments and Stock Status. Grey mullet (GMU). Wellington.
- MPI. 2009. Fishery summary grey mullet (Mugil cephalus) GMU. Minist. Prim. Ind. New Zeal. 307–313.
- Muschick, M., M. Barluenga, W. Salzburger, and A. Meyer. 2011. Adaptive phenotypic plasticity in the Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. BMC Evol. Biol. 11:116.
- Neethling, M., C. a Matthee, R. C. K. Bowie, and S. von der Heyden. 2008. Evidence for panmixia despite barriers to gene flow in the southern African endemic, *Caffrogobius caffer* (Teleostei: Gobiidae). BMC Evol. Biol. 8:325.
- Negre, C., R. Zahn, A. L. Thomas, P. Masqué, G. M. Henderson, G. Martínez-Méndez, I. R. Hall, and J. L. Mas. 2010. Reversed flow of Atlantic deep water during the Last Glacial Maximum. Nature 468:84–88.
- Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.
- Nei, M., and F. Tajima. 1981. Genetic drift and estimation of effective population size. Genetics 98:625–640.
- Nomura, T. 2008. Estimation of effective number of breeders from molecular coancestry of single cohort sample. Evol. Appl. 1:462–474.
- Ostfeld, R. S., and C. D. Canham. 1995. Density-Dependent Processes in Meadow Voles: An Experimental Approach. Ecology 76:521–532.
- Oufiero, C. E., and K. R. Whitlow. 2016. The evolution of phenotypic plasticity in fish swimming. Curr. Zool. 62:475–488.

- Ovenden, J. R. 2013. Crinkles in connectivity: Combining genetics and other types of biological data to estimate movement and interbreeding between populations. Mar. Freshw. Res. 64:201–207.
- Paetkau, D., R. Slade, M. Burden, and A. Estoup. 2004. Genetic assignment methods for the direct, real-time estimation of migration rate: A simulation-based exploration of accuracy and power. Mol. Ecol. 13:55–65.
- Palm, S., J. Dannewitz, T. Prestegaard, and H. Wickströ. 2009. Panmixia in European eel revisited: no genetic difference between maturing adults from southern and northern Europe. Heredity 103:82– 89.
- Palumbi, S. R. 1994. Genetic divergence, reproductive isolation, and marine speciation. Annu. Rev. Ecol. Syst. 25:547–572.
- Papadopoulou, A., I. Anastasiou, F. Spagopoulou, M. Stalimerou, S. Terzopoulou, A. Legakis, and A. P. Vogler. 2011. Testing the species--genetic diversity correlation in the Aegean archipelago: toward a haplotype-based macroecology? Am. Nat. 178:241–255.
- Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.
- Parnmen, S., A. Rangsiruji, P. Mongkolsuk, K. Boonpragob, A. Nutakki, and H. T. Lumbsch. 2012. Using Phylogenetic and Coalescent Methods to Understand the Species Diversity in the *Cladia* aggregata Complex (Ascomycota, Lecanorales). PLoS One 7:1–15.
- Parsons, K. E. 1996. The genetic effects of larval dispersal depend on spatial scale and habitat characteristics. Mar. Biol. 126:403–414.
- Paulin, C. D., and L. J. Paul. 2006. The Kaipara mullet fishery : nineteenth-century management issues revisited. Tuhinga 17:1–26.
- Payne, A. I. 1976. The relative abundance and feeding habits of the grey mullet species occuring in an estuary in Sierra Leone, West Africa. Mar. Biol. 35:277–286.
- Peakall, R., and P. E. Smouse. 2012. GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539.
- Pena, L. D., and S. L. Goldstein. 2014. Thermohaline circulation crisis and impacts during the mid-Pleistocene transition. Science 345:318–322.
- Pettengill, J., and D. Moeller. 2012. Phylogeography of speciation: allopatric divergence and secondary contact between outcrossing and selfing Clarkia. Mol. Ecol. 21:4578–4592.
- Pichevin, L. E., B. C. Reynolds, R. S. Ganeshram, L. Cacho, L. Pena, K. Keefe, and R. M. Ellam. 2009. Enhanced carbon pump inferred from relaxation of nutrient limitation in the glacial ocean. Nature 459:1114–1117.
- Pickford, G. E., and J. G. Phillips. 1959. Prolactin, a factor in promoting survival of hypophysectomized killifish in fresh water. Science 130:454–455.
- Piry, S., A. Alapetite, J. M. Cornuet, D. Paetkau, L. Baudouin, and A. Estoup. 2004. GENECLASS2: A software for genetic assignment and first-generation migrant detection. J. Hered. 95:536–539.
- Piry, S., G. Luikart, and J. M. Cornuet. 1999. BOTTLENECK: A computer program for detecting recent reductions in the effective population size using allele frequency data. J. Hered. 90:502–503.
- Polgar, G., L. Zane, M. Babbucci, F. Barbisan, T. Patarnello, L. Rüber, and C. Papetti. 2014. Phylogeography and demographic history of two widespread Indo-Pacific mudskippers (Gobiidae: Periophthalmus). Mol. Phylogenet. Evol. 73:161–176.
- Pollak, E. 1983. A New Method for Estimating the Effective Population Size from Allele Frequency

Changes. Genetics 104:531–548.

- Pons, J., T. G. Barraclough, J. Gomez-Zurita, A. Cardoso, D. P. Duran, S. Hazell, S. Kamoun, W. D. Sumlin, and A. P. Vogler. 2006. Sequence-Based Species Delimitation for the DNA Taxonomy of Undescribed Insects. Syst. Biol. 55:595–609.
- Pons, J., T. Fujisawa, E. M. Claridge, R. Anthony Savill, T. G. Barraclough, and A. P. Vogler. 2011. Deep mtDNA subdivision within Linnean species in an endemic radiation of tiger beetles from New Zealand (genus *Neocicindela*). Mol. Phylogenet. Evol. 59:251–262.
- Pontin, D. R., and R. H. Cruickshank. 2012. Molecular phylogenetics of the genus *Physalia* (Cnidaria: Siphonophora) in New Zealand coastal waters reveals cryptic diversity. Hydrobiologia 686:91– 105.
- Posada, D. 2008. jModelTest: Phylogenetic model averaging. Mol. Biol. Evol. 25:1253-1256.
- Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.
- Proulx, D. 2006. Sourcebook of Nasca Ceramic Iconography: Reading a Culture through Its Art. University of Iowa Press, Project MUSE, Iowa.
- Pujolar, J. M., G. E. Maes, and F. A. M. Volckaert. 2006. Genetic patchiness among recruits in the European eel *Anguilla anguilla*. Mar. Ecol. Prog. Ser. 307:209–217.
- Pusack, T. J., M. R. Christie, D. W. Johnson, C. D. Stallings, and M. A. Hixon. 2014. Spatial and temporal patterns of larval dispersal in a coral-reef fish metapopulation: Evidence of variable reproductive success. Mol. Ecol. 23:3396–3408.
- Putman, A. I., and I. Carbone. 2014. Challenges in analysis and interpretation of microsatellite data for population genetic studies. Ecol. Evol. 4:4399–4428.
- Putman, N. F., K. J. Lohmann, E. M. Putman, T. P. Quinn, A. P. Klimley, and D. L. G. Noakes. 2013. Evidence for geomagnetic imprinting as a homing mechanism in pacific salmon. Curr. Biol. 23:312–316.
- Queller, D. C., and K. F. Goodnight. 1989. Estimating relatedness using genetic markers. Evolution (N. Y). 43:258–275.
- Rannala, B., and J. L. Mountain. 1997. Detecting immigration by using multilocus genotypes. Proc. Natl. Acad. Sci. 94:9197–9201.
- Reid, K., T. B. Hoareau, J. E. Graves, W. M. Potts, S. M. R. dos Santos, A. W. Klopper, and P. Bloomer. 2016. Secondary contact and asymmetrical gene flow in a cosmopolitan marine fish across the Benguela upwelling zone. Heredity 117:1–9.
- Rice, W. E. R. 1989. Analyzing tables of statistical tests. Evolution 43: 223-225.
- Richards, A. R., and P. J. Rago. 1999. A Case History of Effective Fishery Management : Chesapeake Bay Striped Bass. North Am. J. Fish. Manag. 19:356–375.
- Rioux-Paquette, S. 2012. PopGenKit v1.0: Useful functions for (batch) file conversion and data resampling in microsatellite datasets. R Packag. https://cran.r-project.org/web/packages/PopGenKit/index.html.
- Riquet, F., S. Le Cam, E. Fonteneau, and F. Viard. 2016. Moderate genetic drift is driven by extreme recruitment events in the invasive mollusk Crepidula fornicata. Heredity 117:1–9.
- Ritland, K. 1996. Estimators for pairwise relatedness and individual inbreeding coefficients.
- Rocha-Olivares, a., N. Garber, and K. Stuck. 2000. High genetic diversity, large inter-oceanic divergence and historical demography of the striped mullet. J. Fish Biol. 57:1134–1149.

- Rocha-Olivares, A., N. Garber, A. Garber, and K. Stuck. 2005. Structure of the Mitochondrial Control Region and Flanking Trna Genes of *Mugil cephalus*. Hidrobiologica 15:139–149.
- Rosenberg, N. A., and M. Nordborg. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. Nat. Rev. Genet. 3:380–390.
- Rossi, A. R., M. Capula, D. Crosetti, D. E. Campton, and L. Sola. 1998a. Genetic divergence and phylogenetic inferences in five species of Mugilidae (Pisces: Perciformes). Mar. Biol. 131:213–218.
- Rossi, A. R., M. Capula, D. Crosetti, L. Sola, and D. E. Campton. 1998b. Allozyme variation in global populations of striped mullet, *Mugil cephalus* (Pisces: Mugilidae). Mar. Biol. 131:203–212.
- Rossi, A. R., A. Ungaro, S. De Innocentiis, D. Crosetti, and L. Sola. 2004. Phylogenetic analysis of mediterranean mugilids by allozymes and 16S mt-rRNA genes investigation: Are the mediterranean species of Liza monophyletic. Biochem. Genet. 42:301–315.
- Rousset, F. 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. Mol. Ecol. Resour. 8:103–106.
- Rueda, P. S. 2002. Stomach content analysis of *Mugil cephalus* and *Mugil curema* (Mugiliformes: Mugilidae) with emphasis on diatoms in the Tamiahua lagoon, Mexico. Rev. Biol. Trop. 50:245–252.
- Russell, S., C. A. Sullivan, and A. J. Reichelt-Brushett. 2015. Aboriginal consumption of estuarine food resources and potential implications for health through trace metal exposure; A study in Gumbaynggirr Country, Australia. PLoS One 10:1–17.
- Ruzzante, D. E., S. Mariani, D. Bekkevold, C. André, H. Mosegaard, L. A. W. Clausen, T. G. Dahlgren, W. F. Hutchinson, E. M. C. Hatfield, E. Torstensen, J. Brigham, E. J. Simmonds, L. Laikre, L. C. Larsson, R. J. M. Stet, N. Ryman, and G. R. Carvalho. 2006. Biocomplexity in a highly migratory pelagic marine fish, Atlantic herring. Proc. Biol. Sci. 273:1459–64.
- Ryman, N., and S. Palm. 2006. POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. Mol. Ecol. 6:600–602.
- Sadovy, Y., and D. Y.. Shapiro. 1987. Criteria for the Diagnosis of Hermaphroditism in Fishes. Copeia 1987:136–156.
- Sakamoto, T., and S. D. McCormick. 2006. Prolactin and growth hormone in fish osmoregulation. Gen. Comp. Endocrinol. 147:24–30.
- Salvarina, I., M. Koutrakis, and I. Leonardos. 2010. Juvenile feeding habits of Mugilidae species from estuarine systems in north Aegean Sea. 39th CIESM Congr. Commision Int. Mediterr. Sea 39:795.
- Sambrook, J., E. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Cold Spring Harbor, NY.
- Sargsyan, O., and J. Wakeley. 2008. A coalescent process with simultaneous multiple mergers for approximating the gene genealogies of many marine organisms. Theor. Popul. Biol. 74:104–114.
- Schtickzelle, N., and T. P. Quinn. 2007. A metapopulation perspective for salmon and other anadromous fish. Fish Fish. 8:297–314.
- Sedghifar, A., Y. Brandvain, P. Ralph, and G. Coop. 2015. The spatial mixing of genomes in secondary contact zones. Genetics 201:243–261.
- Selkoe, K. A., O. E. Gaggiotti, B. W. Bowen, and R. J. Toonen. 2014. Emergent patterns of population genetic structure for a coral reef community. Mol. Ecol. 23:3064–3079.
- Selkoe, K. A., J. R. Watson, C. White, T. Ben Horin, M. Iacchei, S. Mitarai, D. A. Siegel, S. D. Gaines,

and R. J. Toonen. 2010. Taking the chaos out of genetic patchiness: Seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. Mol. Ecol. 19:3708–3726.

- Selwyn, J. D., J. D. Hogan, A. M. Downey-Wall, L. M. Gurski, D. S. Portnoy, and D. D. Heath. 2016. Kin-Aggregations Explain Chaotic Genetic Patchiness, a Commonly Observed Genetic Pattern, in a Marine Fish. PLoS One 11:e0153381.
- Sharma, R., B. Kumar, R. Arora, S. Ahlawat, A. K. Mishra, and M. S. Tantia. 2016. Genetic diversity estimates point to immediate efforts for conserving the endangered Tibetan sheep of India. Meta Gene 8:14–20.
- Shen, K.-N., C.-Y. Chen, W.-N. Tzeng, J. D. Chen, W. Knibb, and J. D. Durand. 2010. Development and characterization of 13 GT/CA microsatellite loci in cosmopolitan flathead mullet *Mugil cephalus*. Mol. Ecol. Resour. 10:1098e1105.
- Shen, K.-N., B. W. Jamandre, C.-C. Hsu, W.-N. Tzeng, and J.-D. Durand. 2011. Plio-Pleistocene sea level and temperature fluctuations in the northwestern Pacific promoted speciation in the globallydistributed flathead mullet Mugil cephalus. BMC Evol. Biol. 11:83.
- Shen, K.-N., S.-Y. Tsai, C.-H. Chen, C.-D. Hsiao, and J.-D. Durand. 2015. Next generation sequencing yields the complete mitochondrial genome of the largescale mullet, Liza macrolepis (Teleostei: Mugilidae). Mitochondrial DNA 00:1–2.
- Simpson, G. G. 1951. The Species Concept. Evolution 5:285–298.
- Skelly, D. A., P. M. Magwene, and E. A. Stone. 2016. Sporadic, global linkage disequilibrium between unlinked segregating sites. Genetics 202:427–437.
- Slatkin, M. 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462.
- Slatkin, M. 2008. Linkage disequilibrium understanding the evolutionary past and mapping the medical future. Nat. Rev. Genet. 9:477–485.
- Slatkin, M., and L. Excoffier. 1996. Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm. Heredity 76:377–383.
- Slatkin, M., and W. P. Maddison. 1990. Detecting isolation by distance using phylogenies of genes. Genetics 126:249–260.
- Smith, K. A., and K. Deguara. 2003. Formation and annual periodicity of opaque zones in sagittal otoliths of *Mugil cephalus* (Pisces: Mugilidae). Mar. Freshw. Res. 54:57–67.
- Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.
- Steel, M., and A. McKenzie. 2001. Properties of phylogenetic trees generated by yule-type speciation models. Math. Biosci. 170:91–112.
- Stefanni, S., and H. Knutsen. 2007. Phylogeography and demographic history of the deep-sea fish *Aphanopus carbo* (Lowe, 1839) in the NE Atlantic: Vicariance followed by secondary contact or speciation? Mol. Phylogenet. Evol. 42:38–46.
- Sumpton, W. D., J. R. Ovenden, C. P. Keenan, and R. Street. 2008. Evidence for a stock discontinuity of snapper (*Pagrus auratus*) on the east coast of Australia. Fish. Res. 94:92–98.
- Sun, P., Z. Shi, F. Yin, and S. Peng. 2012. Genetic variation analysis of *Mugil cephalus* in China sea based on mitochondrial COI gene sequences. Biochem. Genet. 50:180–91.

Svedäng, H., D. Righton, and P. Jonsson. 2007. Migratory behaviour of Atlantic cod Gadus morhua:

Natal homing is the prime stock-separating mechanism. Mar. Ecol. Prog. Ser. 345:1–12.

- Swearer, S. E., J. S. Shima, M. E. Hellberg, S. R. Thorrold, G. P. Jones, D. R. Robertson, S. G. Morgan, K. a Selkoe, G. M. Ruiz, and R. R. Warner. 2002. Evidence of self recruitment in demersal marine populations. Bull. Mar. Sci. 70:251–271.
- Talavera, G., V. Dincă, and R. Vila. 2013. Factors affecting species delimitations with the GMYC model: Insights from a butterfly survey. Methods Ecol. Evol. 4:1101–1110.
- Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30:2725–2729.
- Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.
- Thomson, J. M. 1963. Synopsis of biological data on the grey mullet *Mugil cephalus* Linnaeus 1758. CSIRO Fish. Oceanogr. Fish. Synopsis 1:1–77.
- Thomson, J. M. 1966. The grey mullets. Pp. 301–355 *in* Oceanography and Marine Biology: An Annual Review. G. Allen and Unwin.
- Thomson, J. M. 1997. The Mugilidae of the world. Mem. Queensl. Museum 41:457–562.
- Thorrold, S. R., C. Latkoczy, P. K. Swart, and C. M. Jones. 2001. Natal homing in a marine fish metapopulation. Science 291:297–299.
- Thorrold, S., D. Zacherl, and L. Levin. 2007. Population Connectivity and Larval Dispersal Using Geochemical Signatures in Calcified Structures. Oceanography 20:80–89.
- Tomaiuolo, M., T. F. Hansen, and D. R. Levitan. 2007. A theoretical investigation of sympatric evolution of temporal reproductive isolation as illustrated by marine broadcast spawners. Evolution 61:2584–2595.
- Torras, X., L. Cardona, and E. Gisbert. 2000. Cascading effects of the flathead grey mullet *Mugil cephalus* in freshwater eutrophic microcosmos. Hydrobiologia 429:49–57.
- Turner, T. F., J. P. Wares, and J. R. Gold. 2002. Genetic effective size is three orders of magnitude smaller than adult census size in an abundant, estuarine-dependent marine fish (*Sciaenops ocellatus*). Genetics 162:1329–1339.
- Van de Putte, A. P., K. Janko, E. Kasparova, G. E. Maes, J. Rock, P. Koubbi, F. A. M. Volckaert, L. Choleva, K. P. P. Fraser, J. Smykla, J. K. J. Van Houdt, and C. Marshall. 2012. Comparative phylogeography of three trematomid fishes reveals contrasting genetic structure patterns in benthic and pelagic species. Mar. Genomics 8:23–34.
- Van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. 2004. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. Mol. Ecol. Notes 4:535–538.
- Varela, A. I., P. A. Ritchie, and P. J. Smith. 2013. Global genetic population structure in the commercially exploited deep-sea teleost orange roughy (*Hoplostethus atlanticus*) based on microsatellite DNA analyses. Fish. Res. 140:83–90.
- von Biela, V. R., C. E. Zimmerman, G. H. Kruse, F. J. Mueter, B. A. Black, D. C. Douglas, and J. L. Bodkin. 2016. Influence of Basin- and Local-Scale Environmental Conditions on Nearshore Production in the Northeast Pacific Ocean. Mar. Coast. Fish. 8:502–521.
- Wang, J. 2017. The computer program structure for assigning individuals to populations: easy to use but easier to misuse. Mol. Ecol. Resour. 17:981–990.

- Waples, R. S. 2005. Genetic estimates of contemporary effective population size: To what time periods do the estimates apply? Mol. Ecol. 14:3335–3352.
- Waples, R. S. 1998. Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. J. Hered. 89:438–450.
- Waples, R. S., and C. Do. 2008. LDNE: A program for estimating effective population size from data on linkage disequilibrium. Mol. Ecol. Resour. 8:753–756.
- Ward, R. D., T. S. Zemlak, B. H. Innes, P. R. Last, and P. D. N. Hebert. 2005. DNA barcoding Australia's fish species. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 360:1847–1857.
- Wasserman, R. J., and N. A. Strydom. 2011. The importance of estuary head waters as nursery areas for young estuary- and marine-spawned fishes in temperate South Africa. Estuar. Coast. Shelf Sci. 94:56–67.
- Waters, J. M. 2011. Competitive exclusion: Phylogeography's "elephant in the room"? Mol. Ecol. 20:4388–4394.
- Waters, J. M., C. I. Fraser, S. C. Banks, and G. M. Hewitt. 2013a. The founder space race : a reply to Buckley et al. Trends Ecol. Evol. 28:190–191.
- Waters, J. M., C. I. Fraser, and G. M. Hewitt. 2013b. Founder takes all : density-dependent processes structure biodiversity. Trends Ecol. Evol. 28:78–85.
- Watson, T., J. Mckenzie, and B. Hartill. 2005. Catch per unit effort analysis of the northern (GMU 1) grey mullet (*Mugil cephalus*) setnet fishery, 1989-2002. New Zeal. Fish. Assess. Rep. 22:1989–2002.
- Weir, B. S., and C. C. Cockerham. 1984. Estimating F-Statistics for the Analysis of Population Structure. Evolution 38:1358–1370.
- Wellband, K. W., and D. D. Heath. 2017. Plasticity in gene transcription explains the differential performance of two invasive fish species. Evol. Appl. 0–2.
- White, C., K. A. Selkoe, J. Watson, D. A. Siegel, D. C. Zacherl, and R. J. Toonen. 2010. Ocean currents help explain population genetic structure. Proc. R. Soc. 277:1685–94.
- White, T. A., H. A. Fotherby, P. A. Stephens, and A. R. Hoelzel. 2011. Genetic panmixia and demographic dependence across the North Atlantic in the deep-sea fish, blue hake (Antimora rostrata). Heredity (Edinb). 106:690–9.
- Whitehead, P. J. P., M.-L. Bauchot, J.-C. Hureau, J. Nielsen, and E. Tortonese. 1986. Fishes of the north-eastern Atlantic and the Mediterranean, volume III. United Nations Educational Scientific and Cultural Organization.
- Whitfield, A. K. 2010. A century of fish research in South African estuaries. African J. Aquat. Sci. 35:211–225.
- Whitfield, A. K., and S. J. M. Blaber. 1978. Food and feeding ecology of piscivorous fishes at Lake St Lucia, Zululand. J. Fish Biol. 13:675–691.
- Whitfield, A. K., J. Panfili, and J. D. Durand. 2012. A global review of the cosmopolitan flathead mullet *Mugil cephalus* Linnaeus 1758 (Teleostei: Mugilidae), with emphasis on the biology, genetics, ecology and fisheries aspects of this apparent species complex. Rev. Fish Biol. Fish. 22:641–681.
- Wilson, G. A., B. Rannala, E. C. Anderson, E. A. Thompson, P. Beerli, J. Felsenstein, P. Beerli, J. Felsenstein, J. M. Bernardo, A. F. M. Smith, L. E. Carmichael, J. A. Nagy, N. C. Larter, C. Strobeck, G. Casella, R. L. Berger, J. M. Cornuet, S. Piry, G. Luikart, A. Estoup, M. Solignac, K. J. Dawson, K. Belkhir, H. Freville, F. Justy, I. Olivieri, O. E. Gaggiotti, F. Jones, W. M. Lee, W. Amos, J. Harwood, D. Gamerman, W. K. Hastings, J. F. C. Kingman, N. Metropolis, A. W.

Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, D. Paetkau, W. Calvert, I. Stirling, C. Strobeck, J. K. Pritchard, M. Stephens, P. Donnelly, B. Rannala, J. L. Mountain, M. Slatkin, N. H. Barton, S. Tavare, R. Vitalis, D. Couvet, S. Wright, and S. Wright. 2003. Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163:1177–91.

Wolinsky, H. 2010. The puzzle of sympatry. Nat. Publ. Gr. 11:830-833.

- Young, G. C., and I. C. Potter. 2002. Influence of Exceptionally High Salinities, Marked Variations in Freshwater Discharge and Opening of Estuary Mouth on the Characteristics of the Ichthyofauna of a Normally-Closed Estuary. Estuar. Coast. Shelf Sci. 55:223–246.
- Zecca, G., G. Casazza, L. Minuto, M. Labra, and F. Grassi. 2011b. Allopatric divergence and secondary contacts in Euphorbia spinosa L: Influence of climatic changes on the split of the species. Organisims Divers. Evol. 11:357–372.
- Zismann, L., V. Berdugo, and B. Kimor. 1975. The food and feeding habits of early stages of grey mullets in the Haifa Bay region. Aquaculture 6:59–75.