Whole genome sequencing reveals distinct population structure in a high gene flow marine species, Australasian snapper (*Chrysophrys auratus*)

Tom Oosting[1], Bastiaan Star[2], Maren Wellenreuther[3], and Peter Ritchie[1]

[1]Victoria University of Wellington
[2]Oslo University
[3]New Zealand Institute for Plant and Food Research Ltd

October 3, 2023

Abstract

The identification of genetically distinct populations is central to the management and conservation of wild populations. Whole-genome-sequencing allows for high-resolution assessment of genetic structure, demographic connectivity and the impacts of selection acting on different parts of the genome. Here, we utilise population genomics to investigate the genetic structure of the Australasian snapper or Tāmure (*Chrysophrys auratus*), an ecologically, economically, and culturally important (taonga) marine fish. We analysed over four million high-quality SNPs obtained by whole-genome sequencing from 382 individuals collected across its New Zealand range. We identified two genetic clusters (an eastern and western cluster) with genetic disjunctions around on either side of the North Island of New Zealand. These genetic clusters do not match the current fisheries management areas. Pairwise-F_{ST} and ADMIXTURE analyses showed the presence of directional gene flow occurring at both genetic disjunctions from the East to the West cluster. We hypothesize that major ocean currents are limiting the dispersal of snapper at these genetic disjunctions. The heterogeneous coastal environment is also likely driving evolutionary change. A genome scan identified four significantly divergent genomic regions between genetic clusters. A diverse pattern of genetic variation in these regions implies that different evolutionary processes drive local adaptation in these clusters. Identification of candidate genes in these regions also provides a tentative connection to which traits may be under selection. Our results provide novel insights into New Zealand's coastal environment influences evolutionary processes, and valuable information for effective management of the snapper fisheries.

Title page

Title : Whole genome sequencing reveals distinct population structure in a high gene flow marine species, Australasian snapper (Chrysophrys auratus)

Running title : Genetic structure in a high gene flow fish

Authors :

Tom Oosting (Corresponding author) , School of Biological Sciences, Victoria University of Wellington, Wellington, New Zealand tom.oosting@gmail.com

Bastiaan Star , Centre for Ecological and Evolutionary Synthesis (CEES), University of Oslo, Oslo, Norway bastiaan.star@ibv.uio.no

Maren Wellenreuther , The New Zealand Institute for Plant & Food Research Ltd, Nelson, New Zealand & School of Biological Sciences, University of Auckland, Auckland, New Zealand, maren.wellenreuther@plantandfood.co.nz

Peter A. Ritchie , School of Biological Sciences, Victoria University of Wellington, Wellington, New Zealand, peter.ritchie@vuw.ac.nz

Abstract (243/250 words)

The identification of genetically distinct populations is central to the management and conservation of wild populations. Whole-genome-sequencing allows for high-resolution assessment of genetic structure, demographic connectivity and the impacts of selection acting on different parts of the genome. Here, we utilise population genomics to investigate the genetic structure of the Australasian snapper or Tāmure (*Chrysophrys auratus*), an ecologically, economically, and culturally important (taonga) marine fish. We analysed over four million high-quality SNPs obtained by whole-genome sequencing from 382 individuals collected across its New Zealand range. We identified two genetic clusters (an eastern and western cluster) with genetic disjunctions around on either side of the North Island of New Zealand. These genetic clusters do not match the current fisheries management areas. Pairwise-F_{ST} and ADMIXTURE analyses showed the presence of directional gene flow occurring at both genetic disjunctions from the East to the West cluster. We hypothesize that major ocean currents are limiting the dispersal of snapper at these genetic disjunctions. The heterogeneous coastal environment is also likely driving evolutionary change. A genome scan identified four significantly divergent genomic regions between genetic clusters. A diverse pattern of genetic variation in these regions implies that different evolutionary processes drive local adaptation in these clusters. Identification of candidate genes in these regions also provides a tentative connection to which traits may be under selection. Our results provide novel insights into New Zealand's coastal environment influences evolutionary processes, and valuable information for effective management of the snapper fisheries.

Keywords: Population Genomics, Whole-Genome-Sequencing, Selection, Gene Flow, Fisheries Science

Wordcount main text: 6059 (8000 max)

Introduction

The assessment of biodiversity, including the genetic structure observed within species, is one of the primary goals of population management (Moritz, 1994). Specifically, the identification of reproductively discrete units allows management to assess risks and threats, and implement measures to protect the genetic diversity present within the species (Palsbøll et al., 2007). Genetic analysis has also proven invaluable for the identification of biologically significant units (Allendorf, 2017; Allendorf et al., 2010). However, genetic structure is often difficult to identify in marine fish due to low levels of genetic differentiation, a feature that

is facilitated by high dispersal potential, large population sizes, and limited barriers to gene flow (Cano et al., 2008; Ward, 2000). Such low levels of genetic differentiation imply that population genetic analyses (< 100 loci) have often lacked the statistical power to detect fine-scale structure that may exist across a species range (Morin et al., 2009). Today, the ability to identify thousands of single nucleotide polymorphisms (SNPs) provides improved statical power compared to traditional genetic markers, such as microsatellite and mitochondrial DNA (Luikart et al., 2003; Luikart et al., 2018). Besides increased statistical power to identify fine-scale genetic structure, genotyping thousands of SNPs across the genome also allows for the identification of outlier loci (Allendorf et al., 2010). Such outliers are a strong indication of selection acting on specific regions of the genome and provide putative evidence for the adaption of populations to their local environment (Hohenlohe et al., 2020). The identification and conservation of adaptive variation is key for the long-term viability of natural populations, as it provides resilience to environmental perturbations such as climate change (Sgrò et al., 2011).

Various studies have already shown how population genomics can significantly improve the ability to detect genetic structure in marine species. Benestan et al. (2015) used 10,156 SNPs to identify previously undetected genetic structure in American lobster (*Homarus americanus*). Assignment tests also showed that individuals could be assigned with high confidence to their respective genetic cluster, providing a powerful tool monitoring of the species. In Chinook salmon (*Oncorhynchus tshawytscha*), 10,944 SNPs were used to improve the understanding of their genetic structure in Alaskan waters (Larson et al., 2014). The identification of 733 outlier loci showed three genomics under putative selection, which could be used to develop SNP arrays for low coast genetic monitoring. Additional examples of population genomic analyses be able to identity previously unknown or unresolved fine-scale genetic structure in marine species include commercial squid (*Doryteuthis opalescens*), European scallops (*Pecten maximus*), European anchovies (*Engraulis encrasicolus*), (Cheng et al., 2021; Le Moan et al., 2016; Vendrami et al., 2019). These studies show how genomics can provide a important role in the management of marine species, and why we should integrate these tools into the standard approaches to data collection for fisheries management purposes (Bernatchez et al., 2017).

Australasian snapper or tāmure (*Chrysophrys auratus*) is an important commercial and recreational inshore fisheries in New Zealand (Parsons et al., 2014), with an total allowable annual catch of 6,400 tonnes (Fisheries New Zealand, 2018). There are thought to be eight distinct stocks distributed over four quota management areas (SNA; see left panel Figure 1) (Parsons et al., 2014; Walsh et al., 2011; Walsh et al., 2012; M. R. Walsh et al., 2006). Three are recognized in SNA1 based on differences in year-class strength and growth rates (Northland, Hauraki Gulf and Bay of Plenty) (Walsh et al., 2011). Two stocks in SNA2 on either side of the Mahia Peninsula (Figure 1), showing differences in year class strength, age structure and growth rates (Walsh et al., 2012). Snapper north of the Mahia Peninsula showed similar year-class strength to the Bay of Plenty stock in SNA1 (Walsh et al., 2012), suggesting demographic connectivity. Additionally, genetic differences between snapper on either side of the Mahia Peninsula have also been detected using microsatellite and allozyme data (Bernal-Ramírez et al., 2003; Smith et al., 1978). Snapper in SNA7 likely consist of two stocks (Marlborough Sounds and Tasman Bay) (Fisheries New Zealand, 2018). Finally, SNA8 is thought to consist of a single stock. Snapper form this area are also genetically distinct from SNA1 (Bernal-Ramírez et al., 2003; Smith et al., 1978), suggesting the presence of a genetic disjunction around Cape Reinga. The identification of these genetic disjunctions at Cape Reinga and Mahia Peninsula shows the presence of an East and West cluster (Figure 1). This genetic structure was detected using allozyme and microsatellite data, while mitochondrial Dloop sequences showed no apparent genetic structure (Bernal-Ramírez et al., 2003; Smith et al., 1978). This lack of contemporary MT structure is consistent with recent whole MT genome analyses, which showed that patterns of MT variation reflect long-term and broad scale evolutionary processes instead (Oosting et al., 2023). Allozyme and microsatellite data also showed that Hawke's Bay was genetically closets to the West Coast (Bernal-Ramírez et al., 2003; Smith et al., 1978), despite these sites being located on either side of the North Island (Figure 1). One of the allozyme loci was genetically linked to the esterase locus (Smith et al., 1978), of which allele frequencies correlated with fluctuations in ocean temperature over time (Smith, 1979), suggesting natural selection could be influencing the genetic structure. However, similar genetic structure was observed in the selectively neutral microsatellite

3

loci suggesting historic gene flow or other demographic factors like ocean current are restricting gene flow between areas (Bernal-Ramírez et al., 2003). A discrepancy in genetic structure was observed in the Tasman Bay population. Allozyme data suggested that the Tasman Bay population is closely related to the West Coast (Smith et al., 1978), while microsatellite data showed Tasman Bay as a distinct group (Bernal-Ramírez et al., 2003). Similar environmental conditions were used to explain the genetic similarity between Tasman Bay and West Coast (Smith et al., 1978). Bernal-Ramírez et al. (2003) hypothesized that microsatellite loci have higher sensitivity to recent changes in demographic connectivity, such as a population size reduction due to overfishing. This hypothesis is consistent with the loss of genetic variation in the Tasman Bay population which is thought to be caused by intense fishing effort (Hauser et al., 2002). An important next step is to assess whether higher resolution genomic methods can resolve the genetic structure, potentially to the level that of non-genetic methods. If relevant SNPs are identified, the development of SNP arrays could provide a more cost-effective methods for monitoring compared to current methods (Bernatchez et al., 2017; Montanari et al., 2023).

Here, we analyse 382 whole-genome sequences from snapper samples collected from 12 locations, spanning most of the geographic range of this species around New Zealand, to test the genetic stock structure that have been proposed by lower resolution genetic markers. This approach could be used as a framework for how genomic markers can be used for stock structure analyses and to better match the management stocks with natural reproductive units. First, we assessed levels of genetic diversity among 12 sample locations and investigate the presence of adaptive (outlier) SNPs. Second, we test for population genetic differentiation using neutral SNPs and outlier SNPs. Our results are compared and discussed in the context the previous genetic studies of snapper, and we argue how this approach is of value for fisheries and conservation management.

Materials and Methods

1.

Sampling & DNA extraction

Fin clips were collected for 382 individuals between 2017 and 2022 from 12 sampling sites (Figure 1 & Table S1). Samples were obtained through commercial and recreational fishing sources, and no fish were killed specifically for this study. Samples were preserved in DESS (20% DMSO, 0.5 M EDTA pH 8, saturated NaCl) and stored at -20 °C (Oosting et al., 2020). DNA was isolated using a high-salt extraction protocol adapted from Aljanabi and Martinez (1997) (see Appendix 1 for protocol). Purified DNA was stored in 10 mM Tris-HCl pH 8 and 0.1 mM EDTA, and stored at -20 °C. The DNA quantity was determined using Qubit broad range kit ($>$20 ng/µ) (Invitrogen$^{\copyright}$), following manufactures instructions. Sample quality was assessed using Implen$^{\copyright}$ NP80 spectrophotometer (260/280 $<$ 2.0 and/or 260/230 $>$ 2.0) and gel-electrophoresis (selecting for samples with high-molecular-weight DNA, $>$20kbp fragments).

Whole-genome-sequencing

Library preparation and whole-genome-sequencing for the 370 samples was performed by the Australian Genome Research Facility (AGRF). Library preparation was done using the Nextera flex protocol with unique dual indexing. Sequencing was performed using the Illumina NovaSeq 6000 S4, generating 150 bp paired-end reads with an average sequencing depth of ˜11 x per individual. Reads from an additional 12 individuals were included from a previous study (unpublished). Samples were sequenced using the Illumina HiSeq 4000, generating 100bp paired-end reads with an average sequencing depth of ˜30 x per individual. Reads for these 12 individuals were generated over two separate lanes.

Read alignment & genotyping

Read quality was assessed using FastQC v0.11.7 (Andrews, 2010), and compiled with MultiQC v1.7 (Ewels et al., 2016). Reads were aligned to both the nuclear and mitochondrial snapper reference genome using PALEOMIX v1.2.13.3 using the bam pipeline (Ashton, 2013; Catanach et al., 2019; Schubert et al., 2014).

First, Adapter removal v2.2.3 was used for trimming of Illumina forward and reverse adapters, allowing a 33% mismatch rate (–mm 3) and a minimum read length of 25 bp. N's and low-quality bases were trimmed of the 5' and 3' ends. Burrows-Wheeler Aligner (BWA) v0.7.15 (bwa-mem algorithm) was used to align reads (Li & Durbin, 2009), using a minimum mapping quality of 25. MarkDuplicates, Picard tools v2.18.20 was used to remove PCR duplicates (http://broadinstitute.github.io/picard/), and local realignment was done using GATK v4.0.8.1 (Van der Auwera & O'Connor, 2020). Soft-clipped reads were removed using SAMtools v1.9 (Danecek et al., 2021) and bam files were normalized for depth of coverage, by randomly subsampling individuals with significantly more data (n = 48) down to the mean depth using SAMtools view (Danecek et al., 2021). Genotypes were called using BCFtools (Li, 2011). First, BCFtools functions mpileup and call were used to genotype and write multiallelic sites to VCF per linkage group. Missing annotations were added using BCFtools +fill-tags, and VCF for each linkage group were merged using BCFtools concat.

SNP filtering

First, we extracted all biallelic SNPs (–remove-indels –max-alleles 2 –min-alleles 2), and setting all genotypes with an allelic depth below 3 (–minDP 3) to missing using VCFtools v0.1.16 (Danecek et al., 2011). Subsequently, low quality SNPs were filtered using VCFtools (–minQ 600 –max-missing 0.95 –maf 0.01 –min-meanDP 8 –max-meanDP 25). A custom script adapted from Pinsky et al. (2021) was used to remove heterozygote SNPs subjected to allelic imbalance. In short, the script performs a two-sided binomial test to assess whether the reference and alternative alleles occur in equal proportions at heterozygous sites. SNPs significantly deviating from equal proportions were excluded from the data set. The retained high-quality SNPs were used to identify outlier and neutral loci.

First, outliers were identified using OutFlank (Whitlock & Lotterhos, 2015) and pcadapt (Luu et al., 2017). OutFlank was run using a minimum heterozygosity of 0.1 (Hmin 0.1), and LeftTrimFraction and RightTrim-Fraction fraction of 0.05 according to the author's recommendations (Whitlock & Lotterhos, 2015). Loci with a false discovery threshold below 1% (q < 0.01) were considered outliers. For pcadapt, we first ran the program with 10 principal components (K = 4) and used a screeplot to determine the number of informative principal components using Cattell's rule (K = 1) (Figure S1). Outliers were identified for K=1 informative principal components, using a minimum allele frequency of 0.05 (min.maf 0.05). Again, loci with a false discovery threshold below 1% (q < 0.01) we considered outliers. Overlapping outliers between OutFlank and pcadapt were retained, and subsequently thinned using non-overlapping sliding-window of 50,000 bp to obtain independently segregating sites. For each sliding-window, we retained the locus with highest F_{ST} estimated by OutFlank.

A neutral SNP dataset was obtained by first removing all outlier loci (OutFlank and pcadapt). Retained SNPs were then filtered for deviations from Hardy-Weinberg equilibrium (–hwe 0.05), and thinned (–thin 5000) using VCFtools v0.1.16 (Danecek et al., 2011). Finally, linked loci were removed using Plink (–indep-pairwise 50 5 0.2) (Purcell et al., 2007).

Population structure

First, PCAs were performed on the high-quality, neutral and outlier loci using SNPRelate (Zheng et al., 2012), and visualised using a custom script from Clucas et al. (2019). Second, ADMIXTURE was run for K populations between 1 and 4 (Alexander et al., 2009) using 10,000 iterations. The number of K populations that best explain the structure in the data set was evaluated using cross validation selecting K with the lowest error. To investigate the levels of gene flow between the genetic clusters we used the output from AD-MIXTURE, and method described by Robinet et al. (2020), https://github.com/tonyrobinet/introgression. The analysis estimated the level of ancestry for each individual to one of the K ancestral populations. Global and pairwise levels of population differentiation were estimated by calculating F_{ST} from all SNPs using SNPRelate (Zheng et al., 2012). F_{ST} values for each SNP were estimated using the snpgdsFST function, estimating weighted F_{ST} values using the Weir & Cockerham, 1984 method (Weir & Cockerham, 1984), results were visualised using the R-package pheatmap (Kolde & Kolde, 2015). A mantel test was used to test

for isolation by distance (IBD) between sample locations using the R-package ade4 (Thioulouse et al., 2018). A distance matrix was obtained by calculating the least-cost distance using the R-package gdistance (van Etten, 2017). Locations were determined by taking the mean longitudes and latitudes of all samples for each sample location. In locations where GPS data was not available, an arbitrary point representing the area was chosen (i.e. Gisborne and Hawkes Bay).

Genome scan To test for evidence of selection we performed a genome scan using a 5,000bp non-overlapping window. Estimates for nucleotide diversity (π), relative population divergence (F_{ST}), and absolute population divergence (dxy) were obtained using pixy v1.2.7.beta1 (Korunes & Samuk). AllSites VCF files were generated for each linkage group using BCFtools mpileup and call (-m) (Li & Durbin, 2009). Genotypes with a sequencing depth below three (DP<3), and genotype quality below 20 (GQ<20) were set to missing using BCFtools filter. Subsequently, BCFtools view was used to remove indels and multiallelic sites (STRLEN(REF)!=1 | STRLEN(ALT) >=2), and biallelic sites were filtered for quality below 600 and mean coverage between 8 and 25X (QUAL<600 | AVG(FMT/DP)<8 | AVG(FMT/DP)>25). Resulting AllSites VCF files were tab indexed using HTSlib v1.9 (tabix) and run in pixy (Bonfield et al., 2021). Estimates for Tajima's D (D_T & ΔD_T) were estimated using VCFtools v0.1.16 (–TajimaD) (Danecek et al., 2011), using all high-quality biallelic SNPs as described in the SNP filtering paragraph. Manhattan plots were generated in R (R Core Team, 2013). Regions showing significant levels of relative population divergence ($F_{ST} > 0.15$) were further investigated. Genes located in regions of interest were extracted from the genome annotation produced by Catanach et al. (2019).

Genetic diversity Levels of heterozygosity was estimated for each individual, sample location, and genetic cluster using all three SNP datasets (i.e. high-quality, neutral and outlier SNPs). An ANOVA to test for significant differences in heterozygosity between sample locations and genetic clusters using R (R Core Team, 2013). An hierarchical analysis of molecular variance (AMOVA) was performed using the R-package poppr (Kamvar et al., 2014). The hierarchical model was applied to compare the variation between identified genetic clusters, fisheries quota management areas and sample locations. Significance levels were tested using randtest from R-package adegenet, using 999 permutations.

Results

1.

Sequencing and SNP filtering

In total, 14.6 billion reads were generated over 382 individuals, with an average of 38.2 million reads per individual. Sequencing depth ranged between 7.17-60.1x, with an average of 10.8x per individual. After read alignment, bam files of 48 individuals with significantly more reads compared to the average were randomly subsampled to remove bias between diversity estimates (Figure S2). BCFtools genotyped 11.4 million SNPs, and 4,185,623 SNPs were retained after removing indels and low-quality SNPs. High-quality SNPs showed a missing rate of 0.8% and were used for outlier selection. OutFlank identified 3,843 outlier loci using a false discovery rate of 1% ($q = 0.01$). Pcadapt identified 6,008 outlier loci using a false discovery rate of 1% ($q = 0.01$). Combined, outlier analyses identified 6,448 unique outliers, of which 3,403 were overlapping. The final outlier dataset consisted of 278 independently segregating loci after thinning. The neutral dataset consisted of 69,700 independently segregating loci after removing all 6,448 outlier loci, non-independently segregating sites (thinning & loci in Linkage Disequilibrium), and sites deviating from Hardy-Weinberg. Missing rate among neutral SNPs was 0.9%.

Population structure

PCAs from all data three sets were consistent and showed the presence of two genetic clusters (Figure 2A & Figure S3). The "East" cluster (including Northland, Hauraki, Bay of Plenty ands East Cape) and "West" cluster (including Gisborne, Hawke's Bay, Marlborough Sounds, Tasman Bay, Karamea Bight, Kapiti Coast, and West Coast) show genetic disjunctions around Tauroa Peninsula and East Cape. Only the first principal

component (PC1) of the PCA was informative for inferring genetic structure and explained a small proportion of the total amount of variation present in the neutral data (0.59%). ADMIXTURE also supported the presence of two genetic clusters with gene flow (K = 2) (Figure 2B and Figure S4). Both analyses showed movement of individuals and presence of gene flow between genetic clusters, which was primarily observed in West Coast, Cape Reinga, Hawke's Bay and Gisborne. Further analyses showed differential gene flow was occurring from the East to the West cluster as migrant or admixed individuals were only identified in the West cluster (Figure 2C). The higher number of private alleles present in the East cluster compared to the West cluster also supports that gene flow is stronger from East to West (1,846 & 1,351, respectively among all high-quality SNPs). Global genetic differentiation (F_{ST}) was 0.0021, while genetic differentiation (F_{ST}) between the East and West cluster was 0.0031. Pairwise F_{ST} estimates between sample locations ranged between 0.00003 and 0.00461 (Figure 2D). Within genetic clusters, sample locations showed low levels of genetic differentiation ($F_{ST} = 0.00003 – 0.00086$). Mixing of individuals and presence of differential gene flow resulted moderate levels of pairwise genetic differences in Gisborne, Hawke's Bay and Cape Reinga. A significant correlation ($r^2 = 0.6917$, $p = 4.99^{-9}$) was observed between genetic and geographic distance (Figure S5), showing the presence of IBD. A significant mantel test also suggested the presence of IBD ($p = 0.001$). We did observe low genetic differentiation between the geographically separated sample locations Hawke's Bay and West Coast ($F_{ST} = 0.005$).

Genome scan

The genome scan showed heterogeneous levels of differentiation between genetic clusters across the genome (Figure 3). Linkage groups 1, 7, 8, and 10 had regions of high relative genetic differentiation ($F_{ST} > 0.15$) (Figure 4). The region of interest on LG1 spanned 90 Kbp and showed a small increase in absolute genetic divergence (d_{xy}) and contained a coding region for gypsy retrotransposon integrase-like protein 1 (gin1). The of high relative divergence (F_{ST}) on LG7 spanned 305 Kbp but did not show an increase in absolute divergence (d_{xy}). This region did show difference in Tajima's D with the West cluster having elevated levels suggesting a lack of rare alleles. The region contained the coding for microtubule-associated serine/threonine-protein kinase 2 (mast2). The region of high relative genetic divergence on LG8 spanned 30 Kbp and showed a small increase in absolute genetic divergence compared to the surrounding region. The region contained the coding region for cyclic amp-dependent transcription factor atf-2 (atf2). Finally, regions of interest on LG10 spanned 35 Kbp and showed the highest level of relative genetic divergence between the East and West cluster. Absolute genetic divergence (d_{xy}) was also high within that region, and the West cluster showed high nucleotide diversity and lack of rare alleles (Tajima's D > 2). The region contained multiple genes, but hexokinase-2 (hk2) was identified to be relevant for our discussion.

Genetic diversity

Analyses of molecular variance (AMOVA) showed that the majority of variation is observed within individuals (Table 2), 99 and 84 percent for neutral and outlier loci respectively. Only 0.28% of neutral variation was observed between genetic clusters ($p = 0.194$), in contrast to 12.27% of variation in outlier loci ($p = 0.001$). Variation between quota management areas (QMAs) within the same genetic cluster was 0.01 and 0.52% for neutral and outlier loci respectively. Similar, variation between sample locations within the QMAs was 0.01 and 0.73% for neutral and outlier loci respectively. Levels of heterozygosity varied between datasets (Table 1). Mean heterozygosity was lowest among high-quality loci (0.215), with neutral and outlier loci having a mean heterozygosity of 0.301 and 0.372 respectively. Heterozygosity also varied between sample locations, rangin between 0.213-0.218, 0.298-0.304, and 0.359-0.393 for high-quality, neutral, and outlier loci respectively. Significant differences in heterozygosity were observed between 12 sample location comparisons (Figure S6A). Here, Northland and East Cape had significant higher levels of heterozygosity than Gisborne, Karamea Bight, Kapiti Coast, Marldborough Sounds, and Tasman Bay (among high-quality and neutral loci). These significant differences in heterozygosity were not observed among outlier loci (Figure S6B). Heterozygosity was significnatly different between genetic clusters (Figure 5), with heterozygocity among high-quality and neutral loci being higher in the East cluster (9.8^{-10} and 8.0^{-9} respectievly), while hetrozygosity among outlier

loci was significnatly higher in the West cluster ($p = 4.1^{-5}$).

Discussion

The goal of this study was to assess the genetic structure and genetic variation of snapper across their entire distribution around New Zealand. We identified the presence of two genetic clusters with genetic disjunctions around Tauroa Peninsula and the East Cape, and directional gene flow occurring from the East to the West cluster. Levels of heterozygosity among high-quality and neutral loci were higher in the East Cluster, while heterozygosity among outlier loci was higher in the West Cluster. A genome scan showed genetically differentiated regions between the two clusters. Below we will discuss how these results have improved our understanding of the genetic structure in snapper and the potential implications for fisheries management.

Genetic structure

Our results provide unequivocal support for the presence of two genetic clusters (Figure 2), with genetic disjunctions around Tauroa Peninsula and East Cape (Figure 6), that were suggested by microsatellite and alloenzyme data (Bernal-Ramírez et al., 2003; Smith et al., 1978). These genetic disjunctions also correspond with the boundaries of proposed bioregions (Shears et al., 2008), showing the genetic structure of the snapper matches that of general pattens of biodiversity. For example, genetic disjunctions in these areas were also identified for tarakihi (*Cheilodactylus macropterus*), barracouta (*Thyrisites atun*), jack mackerel (*Trachurus declivis*) (Gauldie & Johnston, 1980; Papa et al., 2020), waratah anemone (*Actinia tenebrosa*) (Veale & Lavery, 2012), and New Zealand Estuarine Clam (*Austrovenus stutchburyi*) (Ross et al., 2012). Restricted gene flow at the genetic disjunctions resulted in low genetic differentiation in snapper between the East and West cluster ($F_{ST}= 0.0031$). This level of genetic divergence is similar to those reported for snapper (*C. auratus*) along the South and West Coast of Australia ($F_{ST} = 0.0041$) (Bertram et al., 2022) and other species of fish, such as Atlantic cod (*Gadus morhua; $F_{ST} = 0.0037$*), blue whiting (*Micromesistius australis; $F_{ST} = 0.006$*), and herring (*Clupea harengus; $F_{ST} = 0.012$*) (Guo et al., 2016; Knutsen et al., 2011; McKeown et al., 2017).

Ocean currents likely play a dominant role in the direction of movement around these genetic disjunctions. At Cape Reinga, the Tasman Flow hits the North Island of New Zealand and splits up into the East and West Auckland current (Figure 6) (Bull et al., 2018; Papa et al., 2020). Around the East Cape, the East Coast current originates from the East Auckland current and flows around the East Cape southward (Figure 6). Mixing of individuals from both clusters was observed at Cape Reinga and Gisborne (Figure 2A), and resulted in intermediate pairwise F_{ST} estimates (Figure 2D). While dispersal of individuals across the genetic disjunctions was observed in both directions, individuals that showed intermediate levels of ancestry to both genetic clusters were only present south of the genetic disjunctions (Figure 2B&C), suggesting that gene flow is primarily occurring from the East to the West clusters. Snapper around Gisborne also have similar life history traits to those in the Bay of Plenty (Parsons et al., 2014; Walsh et al., 2012), suggesting that Gisborne is demographically connected to the Bay of Plenty. Similar levels of demographic connectivity have not been reported between sample locations on either side of Tauroa Peninsula, and are through to constitute of two distinct stocks (Parsons et al., 2014; C. Walsh et al., 2006). The area between Cape Reinga and Tauroa peninsula was also identified as the most prominent genetic break for marine species around the North Island of New Zealand (Arranz et al., 2021).

While SNPs were unable to identify such fine-scale genetic structure, structural variation in snapper may provide additional resolution and statistical power to detect genetic differences between different stocks. In other species, structural variants have provided much more detailed population differentiation compared to SNPs (Dorant et al., 2020). The amount of structural variation in snapper is known to outweigh the number of SNPs threefold (Catanach et al., 2019). Structural variation also plays an important function in adaptation to local environments (Hoban et al., 2016), both copy number variation (Dorant et al., 2020; Perry et al.,

2007), and inversions (Berg et al., 2016; Star et al., 2017). Future studies utilizing structural variation will hopefully continue to improve the fine-scale structure of snapper.

Genetic variation

The genetic clusters also showed significant differences in levels of diversity. Genetic variation was primarily observed within individuals, showing 99% and 84% for neutral and outlier loci respectively (Table 2). Such levels of within species diversity are typical for a marine species with high potential for dispersal are large effective population sizes (Cano et al., 2008; Palumbi, 2003; Ward, 2000). Less than 1% of genetic variation in neutral and outlier loci was observed between sample locations or quota management areas (QMAs), which shows high levels of genetic connectivity between fisheries management areas (Fisheries New Zealand, 2018; Parsons et al., 2014). However, 12.27% of variation among outlier loci was observed between genetic clusters, suggesting environmental differences are driving local adaptation (discussed below).

Heterozygosity estimates provided new insights into the diversity contained within each of the genetic clusters. Heterozygosity was significantly higher in the East cluster for high-quality and neutral loci (Figure 5). Similarly, significant differences in heterozygosity were observed among neutral loci between sample locations from respective genetic clusters (Figure S6A). This shows that variation is primarily observed between clusters rather than individual sample locations which closely resemble the recognized stocks by fisheries management (Parsons et al., 2014). Higher levels of genetic variation in the East cluster is consistent with the (historic) larger stock sizes in fisheries management area SNA1 which makes up the majority of the East genetic cluster (Fisheries New Zealand, 2018). In contrast, heterozygosity among outlier loci was significantly higher for the West cluster (Figure 5). We hypothesize that higher diversity among outlier loci is because the West cluster is spread over a more heterogenous environment. Moreover, the West cluster has a larger latitudinal range, with more prominent environmental gradients compared to the East cluster (Figure S7). The genomic regions attributing to the higher levels of variation in outlier loci in the West cluster also correlate with the regions that show high levels of genetic differentiation between the two clusters (Figure 3).

Heterozygosity is a key indicator of overall diversity and often used to assess changes in diversity over time or between a populations/species (Chapman et al., 2009; Hauser et al., 2002). However, the use of different genetic markers and variation in sequencing depth or SNP filtering parameters can make difficult to compare results between studies. Within this study alone, sequencing depth and the level of SNP filtering resulted in range of heterozygosity estimates (Table 1, Figure S9). For accurate comparisons, sequencing depths should be normalized, or high enough that fluctuations in depth no longer significantly affect heterozygosity levels among individuals. Neutral loci would be the most appropriate genetic marker for population or species comparisons as they are not influenced by selection, and thus best reflect the demographic processes (e.g. population size, migration) (Charlesworth et al., 2003). Previous studies on snapper utilized microsatellites to assess levels of genetic variation (Ashton, 2013; Hauser et al., 2002; Le Port et al., 2017). Heterozygosity estimates reported in these studies ranged between 0.521 and 0.938. In contrast, heterozygosity estimates in this study ranged between 0.215 and 0.372 (Table 1). High levels of heterozygosity among microsatellite loci are common because they are often selected based on the number of alleles. This form of ascertainment bias implies estimates of genetic variation are inflated (Brandström & Ellegren, 2008; Väli et al., 2008), and incompatible with heterozygosity estimates based on SNP data. For example, Hauser et al. (2002) reported a reduction in heterozygosity in Tasman Bay snapper between 1950 and 1989, suggesting loss of variation due to overexploitation. However, we did not detect consistent differences in heterozygosity between Tasman Bay and other sample locations that would suggest this population has suffered from loss of genetic variation due to exploitation (Figure S6). This could imply that either the Tasman Bay population has since recovered from exploitation, or our data was unable to detect the reduction in heterozygosity. It is possible that microsatellites could also be more sensitive to demographic changes because of the high number of low frequency alleles compared to SNPs. While this implies that microsatellites are better at detecting fluctuations in heterozygosity is thus reduction in diversity, it's important to question whether such changes are relevant for the overall health of a population. Recent studies using genomic data have suggested

9

that populations are able to maintain genetic variation throughout short periods of heavy exploitation (Therkildsen et al., 2010). Most notably, Pinsky et al. (2021) showed there was no loss of genetic variation in the heavily exploited North Atlantic cod. Like Atlantic cod, the period of exploitation for snapper was relatively short. Such observations are encouraging as it may imply that the species will have no immediate long-term loss of diversity due to fishing.

Putative evidence for adaptation

The heterogeneous environment of New Zealand's coastal ecosystems is likely leading to local adaptions in snapper. We found that 12.27% of the observed variation among outlier loci was identified between genetic clusters (Table 2), and the genome scan showed regions of high genetic differentiation (F_{ST}) between genetic clusters (Figure 3). Genomic regions on linkage groups 1, 7, 8 and 10 showing strong relative genetic differentiation ($F_{ST} > 0.15$) were of particular interest (Figure 4). Other summary statistics were utilized to investigate which evolutionary processes could be underlying to the observed genetic differentiation. Studies have emphasized the distinction between relative (F_{ST}) and absolute (d_{xy}) measures of genetic differentiation and the implications regarding the evolutionary processes that are driving genetic differentiation (Cruickshank & Hahn, 2014; Delmore et al., 2018; Nachman & Payseur, 2012). Regions of interest were selected for high relative divergence (F_{ST}). However, it is unclear whether this is due to 1) a reduction in gene flow between the two clusters, 2) or local selection in one of the clusters is affecting the within-population diversity. Both mechanisms result in an increase in F_{ST}, but only the first mechanism is associated with the process of speciation (Cruickshank & Hahn, 2014). Absolute genetic divergence (d_{xy}) which only looks at fixed differences between clusters can be used to distinguish between these processes. If gene flow is reduced due to selection, we would expect the number of fixed differences between the two clusters to increase.

Overall, absolute genetic divergence (d_{xy}) was not elevated in regions of high relative divergence (F_{ST}) (Figure 3), suggesting that gene flow is unrestricted in these regions. Regions of interest on LG1 and LG10 did show high relative and absolute genetic divergence (Figure 4). The region of high absolute genetic divergence on LG1 is restricted to a single 5 Kbp window containing the coding regions for gypsy retrotransposon integrase-like protein 1 (gin1). Transposons are associated with high mutation rates (Wicker et al., 2016), and likely to induce population specific mutations inflating absolute divergence (d_{xy}). Absolute divergence in the region of interest on LG10 is more distinct, showing increased d_{xy} across multiple windows. Following the logic of Cruickshank and Hahn (2014), this region shows support for restricted gene flow due selection which is associated with process of "speciation with migration". However, it is unlikely that the two genetic clusters are currently diverging into fully diverged populations due to the isolated nature of this distinct pattern on divergence.

We hypothesize that the regions of interest are associated with local adaptation in one of the genetic clusters (Figure 4). In this scenario, local selection effects with-population variation causing an increase in relative divergence (F_{ST}) (Cruickshank & Hahn, 2014). Evidence for local selection was strongest in regions of interest on LG7 and LG10, suggesting the presence of selection in the West cluster. In both regions, elevated Tajima's D (> 2) in the West cluster indicates the lack of rare alleles and higher levels of heterozygosity than expected if selectively neutral. This observation is consistent with mean level heterozygosity being higher in the West cluster compared to the East cluster (Figure 5). It is possible that multiple alleles are maintained in the West Cluster because it is distributed over a large geographical area with environmental gradients in ocean temperature, oxygen concentration, pH along a latitudinal axis (Figure 6 & Figure S7). Ocean temperature has already been associated with the observed genetic differences between the East and West Clusters (Smith, 1979; Smith et al., 1978). An allozyme locus linked to the *est-4* gene was shown to have the same genetic structure as reported in this study. The coding regions for *est-4* was not identified within the regions of interest in this study. However, coding regions for genes in the regions of interest did show a tentative connection between adaptation to ocean temperature, based on the assumption that ocean temperature effects the metabolic rate and growth rates of fish (Boscolo-Galazzo et al., 2018; Johansen & Jones, 2011). On LG7, *mast2* is linked to the expression of epidermal growth factor (Ahn et al., 1991). On LG8, *atf2* is a transcription factor that is activated by stress-activated protein kinases like *p38* and involved

in a wide-range of cellular functions including *(Zarubin & Han, 2005)* . In *drosophila* , the homolog for *atf2* (*dATF-2*) is critical for regulation of fat metabolism, and in mice inhabitation of the p38-Atf-2 pathway was shown to reduce white adipose tissue (Maekawa et al., 2010). In humans, *atf2* is thought to may be critical for glucose metabolism and tumorigenesis (Zhao et al., 2019). On LG10, *hk2* codes for a rate limiting enzyme during the first step of glycolysis (Li et al., 2020), and has been shown to play an important role in cell growth (Wellen et al., 2010). The connection between the identified genes within the regions of genetic divergence and their association with local adaptation is highly speculative and requires extensive research that goes beyond the scope of the manuscript. Genotype by environment analyses will be conducted to investigate which environmental factors may be attributed to observed patterns of genetic divergence between the two clusters.

Conclusion

This study utilised whole-genome-sequencing data to investigate the genetic structure and demographic connectivity of Australasian snapper in New Zealand to inform management decisions. Our results provided conclusive evidence for the presence of two distinct genetic clusters with genetic disjunctions around the Tauroa Peninsula and East Cape. Estimates of genetic differentiation and gene flow suggest dispersal primarily occurs from the East to the West cluster, with ocean currents likely playing a key role in their ability to disperse across the genetic disjunctions. Genomic regions of high relative genetic divergence provide tentative evidence for adaptation to local environments. Fisheries management agencies should utilize the genetic structure to refine management areas to key fisheries species (Bernatchez et al., 2017). Currently, quota management areas do not match with the genetic disjunctions observed in snapper and other fisheries species (Figure 6). Ideally, genomics would be able to identify genetic structure that corresponds to the demographically independent stocks based on year-class strength and growth rates (Parsons et al., 2014; Walsh, 2003; Walsh et al., 2011). Combined with non-genetic methods, genomics can significantly contribute to the management of the snapper fisheries.

Author contributions

TO was responsible for the overall design of the study, sample collection and DNA extraction of the samples, performing all analyses, and the lead over the writing this manuscript. BS contributed by helping with the overall design of this study and helped write the manuscript. MW was responsible for funding acquisition, supervision, and provided feedback on analyses and helped with the writing of the manuscript. PR was responsible for funding acquisition, supervision, project administration, contributed by helping with the overall design of this study, and helped write the manuscript.

Data accessibility and Benefit-Sharing

Illumina short-read data and associated metadata are deposited on the Genomics Aotearoa repository (https://repo.data.nesi.org.nz/), Repository ID: AGDR00033 (Te Aika et al., 2023). All scripts used in

the analyses are openly available on GitHub at https://github.com/tomoosting (to be made public at time of publication).

Conflict of interests

Nothing to declare.

References

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

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome research* ,*19* (9), 1655-1664.

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

Allendorf, F. W. (2017). Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology* ,*26* (2), 420-430.

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

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

Arranz, V., Fewster, R. M., & Lavery, S. D. (2021). Geographic concordance of genetic barriers in New Zealand coastal marine species.*Aquatic Conservation: Marine and Freshwater Ecosystems* ,*31* (12), 3607-3625.

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

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

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

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

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

Bertram, A., Fairclough, D., Sandoval-Castillo, J., Brauer, C., Fowler, A., Wellenreuther, M., & Beheregaray, L. B. (2022). Fisheries genomics of snapper (Chrysophrys auratus) along the west Australian coast.*Evolutionary Applications* , *15* (7), 1099-1114. https://doi.org/https://doi.org/10.1111/eva.13439

Bonfield, J. K., Marshall, J., Danecek, P., Li, H., Ohan, V., Whitwham, A., Keane, T., & Davies, R. M. (2021). HTSlib: C library for reading/writing high-throughput sequencing data. *GigaScience* ,*10* (2). https://doi.org/10.1093/gigascience/giab007

Boscolo-Galazzo, F., Crichton, K. A., Barker, S., & Pearson, P. N. (2018). Temperature dependency of metabolic rates in the upper ocean: A positive feedback to global climate change? *Global and Planetary Change* , *170* , 201-212.

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

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

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

Chapman, J., Nakagawa, S., Coltman, D., Slate, J., & Sheldon, B. (2009). A quantitative review of heterozygosity–fitness correlations in animal populations. *Molecular Ecology* , *18* (13), 2746-2765.

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

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

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

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* , *23* (13), 3133-3157. https://doi.org/10.1111/mec.12796

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

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., & Davies, R. M. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* ,*10* (2), giab008.

Delmore, K. E., Lugo Ramos, J. S., Van Doren, B. M., Lundberg, M., Bensch, S., Irwin, D. E., & Liedvogel, M. (2018). Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. *Evolution Letters* ,*2* (2), 76-87.

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

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

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

Gauldie, R. W., & Johnston, A. (1980). The geographical distribution of phosphoglucomutase and glucose phosphate isomerase alleles of some New Zealand fishes. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* , *66* (2), 171-183.

Guo, B., Li, Z., & Merilä, J. (2016). Population genomic evidence for adaptive differentiation in the Baltic Sea herring. *Molecular Ecology* , *25* (12), 2833-2852.

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

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

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

Johansen, J., & Jones, G. (2011). Increasing ocean temperature reduces the metabolic performance and swimming ability of coral reef damselfishes. *Global Change Biology* , *17* (9), 2971-2979.

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

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

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

Korunes, K. L., & Samuk, K. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data.*Molecular ecology resources* , *n/a* (n/a). https://doi.org/https://doi.org/10.1111/1755-0998.13326

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

Le Moan, A., Gagnaire, P. A., & Bonhomme, F. (2016). Parallel genetic divergence among coastal–marine ecotype pairs of European anchovy explained by differential introgression after secondary contact.*Molecular Ecology* , *25* (13), 3187-3202.

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

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

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* , *25* (14), 1754-1760.

Li, W. C., Huang, C. H., Hsieh, Y. T., Chen, T. Y., Cheng, L. H., Chen, C. Y., Liu, C. J., Chen, H. M., Huang, C. L., Lo, J. F., & Chang, K. W. (2020). Regulatory Role of Hexokinase 2 in Modulating Head and Neck Tumorigenesis [Original Research]. *Front Oncol* ,*10* (176), 176. https://doi.org/10.3389/fonc.2020.00176

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

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

Luu, K., Bazin, E., & Blum, M. G. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular ecology resources* , *17* (1), 67-77.

Maekawa, T., Jin, W., & Ishii, S. (2010). The Role of ATF-2 Family Transcription Factors in Adipocyte Differentiation: Antiobesity Effects of p38 Inhibitors. *Molecular and Cellular Biology* , *30* (3), 613-625. https://doi.org/10.1128/MCB.00685-09

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

Montanari, S., Deng, C., Koot, E., Bassil, N. V., Zurn, J. D., Morrison-Whittle, P., Worthington, M. L., Aryal, R., Ashrafi, H., Pradelles, J., Wellenreuther, M., & Chagné, D. (2023). A multiplexed plant–animal SNP array for selective breeding and species conservation applications. *G3 Genes|Genomes|Genetics* . https://doi.org/10.1093/g3journal/jkad170

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

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

Nachman, M. W., & Payseur, B. A. (2012). Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philosophical Transactions of the Royal Society B: Biological Sciences* , *367* (1587), 409-421.

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

Oosting, T., Martínez-García, L., Ferrari, G., Verry, A. J. F., Scarsbrook, L., Rawlence, N. J., Wellenreuther, M., Star, B., & Ritchie, P. A. (2023). Mitochondrial genomes reveal mid-Pleistocene population divergence, and post-glacial expansion, in Australasian snapper (Chrysophrys auratus). *Heredity* , *130* (1), 30-39. https://doi.org/10.1038/s41437-022-00579-1

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

Palumbi, S. R. (2003). Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications* ,*13* (1), S146-S158. https://doi.org/https://doi.org/10.1890/1051-0761(2003)013[0146:PGDCAT]2.0.CO;2

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

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

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

Pinsky, M. L., Eikeset, A. M., Helmerson, C., Bradbury, I. R., Bentzen, P., Morris, C., Gondek-Wyrozemska, A. T., Baalsrud, H. T., Brieuc, M. S. O., Kjesbu, O. S., Godiksen, J. A., Barth, J. M. I., Matschiner, M.,

Stenseth, N. C., Jakobsen, K. S., Jentoft, S., & Star, B. (2021). Genomic stability through time despite decades of exploitation in cod on both sides of the Atlantic. *Proceedings of the National Academy of Sciences* , *118* (15), e2025453118. https://doi.org/10.1073/pnas.2025453118

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., & Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* , *81* (3), 559-575.

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

Robinet, T., Roussel, V., Cheze, K., & Gagnaire, P. A. (2020). Spatial gradients of introgressed ancestry reveal cryptic connectivity patterns in a high gene flow marine fish. *Molecular Ecology* ,*29* (20), 3857-3871.

Ross, P. M., Hogg, I. D., Pilditch, C. A., Lundquist, C. J., & Wilkins, R. J. (2012). Population Genetic Structure of the New Zealand Estuarine Clam Austrovenus stutchburyi (Bivalvia: Veneridae) Reveals Population Subdivision and Partial Congruence with Biogeographic Boundaries.*Estuaries and Coasts* , *35* (1), 143-154. https://doi.org/10.1007/s12237-011-9429-z

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

Sgrò, C. M., Lowe, A. J., & Hoffmann, A. A. (2011). Building evolutionary resilience for conserving biodiversity under climate change. *Evolutionary Applications* , *4* (2), 326-337.

Shears, N. T., Smith, F., Babcock, R. C., Duffy, C. A., & Villouta, E. (2008). Evaluation of biogeographic classification schemes for conservation planning: Application to New Zealand's coastal marine environment. *Conservation biology* , *22* (2), 467-481.

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

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

Star, B., Boessenkool, S., Gondek, A. T., Nikulina, E. A., Hufthammer, A. K., Pampoulie, C., Knutsen, H., Andre, C., Nistelberger, H. M., Dierking, J., Petereit, C., Heinrich, D., Jakobsen, K. S., Stenseth, N. C., Jentoft, S., & Barrett, J. H. (2017). Ancient DNA reveals the Arctic origin of Viking Age cod from Haithabu, Germany. *Proc Natl Acad Sci U S A* , *114* (34), 9152-9157. https://doi.org/10.1073/pnas.1710186114

Te Aika, B., Liggins, L., Rye, C., Perkins, E. O., Huh, J., Brauning, R., Godfery, T., & Black, M. A. (2023). Aotearoa genomic data repository: An ahuru mowai for taonga species sequencing data. *Mol Ecol Resour* , *n/a* (n/a). https://doi.org/10.1111/1755-0998.13866

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

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

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

Van der Auwera, G. A., & O'Connor, B. D. (2020). *Genomics in the cloud: using Docker, GATK, and WDL in Terra* . O'Reilly Media.

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

Veale, A., & Lavery, S. (2012). The population genetic structure of the waratah anemone (Actinia tenebrosa) around New Zealand. *New Zealand Journal of Marine and Freshwater Research* , *46* (4), 523-536.

Vendrami, D. L. J., De Noia, M., Telesca, L., Handal, W., Charrier, G., Boudry, P., Eberhart-Phillips, L., & Hoffman, J. I. (2019). RAD sequencing sheds new light on the genetic structure and local adaptation of European scallops and resolves their demographic histories.*Scientific Reports* , *9* (1), 7455. https://doi.org/10.1038/s41598-019-43939-4

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

Walsh, C., McKenzie, J., & Armiger, H. (2006). Spatial and temporal patterns in snapper length and age composition and movement, west coast North Island, New Zealand. *New Zealand Fisheries Assessment Report* , 6.

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

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

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

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

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

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

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

Wicker, T., Yu, Y., Haberer, G., Mayer, K. F., Marri, P. R., Rounsley, S., Chen, M., Zuccolo, A., Panaud, O., Wing, R. A., & Roffler, S. (2016). DNA transposon activity is associated with increased mutation rates in genes of rice and other grasses. *Nat Commun* , *7* , 12790. https://doi.org/10.1038/ncomms12790

Zarubin, T., & Han, J. (2005). Activation and signaling of the p38 MAP kinase pathway. *Cell Research* , *15* (1), 11-18. https://doi.org/10.1038/sj.cr.7290257

Zhao, S., Feng, J., Li, C., Gao, H., Lv, P., Li, J., Liu, Q., He, Y., Wang, H., Gong, L., Li, D., & Zhang, Y. (2019). Phosphoproteome profiling revealed abnormally phosphorylated AMPK and ATF2 involved in glucose metabolism and tumorigenesis of GH-PAs. *Journal of Endocrinological Investigation* , *42* (2), 137-148. https://doi.org/10.1007/s40618-018-0890-4

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

Table 1: Sample collection location and estimates of genetic diversity

Population
Northland
Hauraki Gulf
Bay of Plenty
East Cape
Gisborne
Hawkes Bay
Marlborough Sounds
Tasman Bay
Karamea Bight
Kapiti Coast
West Coast
Cape Reinga
All

Notes: Lath = Latitude, Long = Longitude, N_{ind} = number of individuals, H_{hq} = observed heterozygosity based on the hig

Tables

Table 2: AMOVA among 12 populations using neutral and outlier loci

Between GCs
Between QMA within GCs
Between POP within QMA
Between samples within POP
Within samples
Total

Notes: GCs = Genetic clusters, QMA = Quota management areas, POP = sample locations, d.f. = degrees of freedom, ɸ =

Figure legends

Figure 1: Map showing sample distribution. The blue contour line shows 50m line where snapper primary occur, and red shows 200m indicating the maximum depth where snapper is found. Left panel shows the five different management areas in which snapper are currently managed. Top right shows an illustration of snapper (Courtesy of Plant and Food Research).

Figure 2: Genetic structure of snapper in New Zealand. A) PCA showing the genetic structure based on neutral loci. Individuals were grouped by sample location. similar PCAs based using high-quality and outlier loci are shown in Figure S3) Admixture plot for K=2. Admixture proportions correspond with PCA results, with evidence of populations mixing on either side of the proposed genetic disjunctions at Cape Reinga and Mahia peninsula. Blue shows ancestry to the West cluster, red shows ancestry to the East cluster. C) Analyses of ancestry proportions to the East cluster shows the presence of directional gene flow from the East to the West cluster. D) Heatmap showing pairwise weighted FST estimates for sampling locations.

Figure 3: Manhattan plot showing summary statistics for 5,000 base pair non-overlapping windows. Nucleotide diversity (π), F_{ST}, d_{xy}, and delta *Tajima's* $D(\Delta D_T)$. Dashed black line shows the $F_{ST} = 0.15$ boundary, and $F_{ST} > 0.15$ are indicated in red. Regions of interest on Linkage groups 1, 7, 8 and 10 are indicated with vertical black lines and shown in Figure 4.

Posted on 3 Oct 2023 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.169634755.50451607/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Figure 4: Manhattan plots showing the four genomics regions showing high levels of genetic differentiation between the East and West genetic cluster. Nucleotide diversity (π),F_{ST} , d_{xy} and delta *Tajima's D* (ΔD_T). Grey areas indicate regions of high relative genetic divergence ($F_{ST} > 0.15$). Solid black horizontal lines indicated the median. Dashed black horizontal lines indicate the 0.95 quantile (π,F_{ST} , and d_{xy}) or 0.025 and 0.975 quantiles (D_T and ΔD_T). Locations of genes within regions of interest are shown at the top of each plot in black and grey. Relevant genes are indicted by name. X-axis shows position on linkage group in Mbp.

Figure 5: Box plots showing levels of heterozygosity among individuals between the East and West genetic clusters for all high-quality loci, neutral loci, and outlier loci. Mean levels of heterozygosity are depicted in the middle of each boxplot. Levels of significant between genetic clusters for each SNP dataset are shown at the top of each panel.

Figure 6: Spatial distribution of the East and West genetic cluster. Arrows show the direction of gene flow from the East to the West genetic cluster. Blue and red show the 50m bathymetry line for each of the genetic clusters, and light blue and light red indicate the 200m bathymetry line for each of the genetic lusters. Dashed lines show boundaries between fisheries management areas (SNA1, SNA2, SNA7, and SNA8). Solid arrows show the direction of gene flow, dashed lined with triple arrowheads indicate relevant ocean currents, WAUC = West Auckland Current, EAUC = East Auckland Current, ECC = East Coast Current.

Supplementary table – legends

Table S1 Information Samples sites

Supplementary figure – legends

Figure S1: Scree plot used to determine the number of informative principal components (K=1) for the identification of outliers using pcadapt.

Figure S2: Coverage of each individual A) before and B) after normalisation of bam files. Each point is an individual, and colours indicate sample locations.

Figure S3: PCA generated using three different SNP datasets. High-Quality, Neutral, and Outlier loci. PCAs show a consistent grouping of sample location which cluster into two genetic clusters, East and West.

Figure S4: Cross validation scores for ADMIXTURE suggesting K=2 is the most appropriate number of clusters. Lowest cross validation error indicating most appropriate number of K is shown in red (K=2).

Figure S5: Genetic distance (weighted FST) between populations plotted against geographic distance. show the linear regression between points. Colours indicate if comparisons were made between sample locations from the same cluster, or between different clusters. The black line shows the regression line between all points.

Figure S6: Boxplots showing levels of heterozygosity for A) neutral and B) outlier loci per sample location.

Figure S7: plots showing environmental gradient for mean ocean temperature at max depth, pH, and Dissolved oxygen concentration at max depth. Figures show heterogeneous environmental gradients that pose selective pressures on snapper populations. Data has been obtained from the Bio-Oracle database and plotted in R.

Figure S8: Boxplots showing levels of heterozygosity among neutral loci before normalisation of bam files.

19
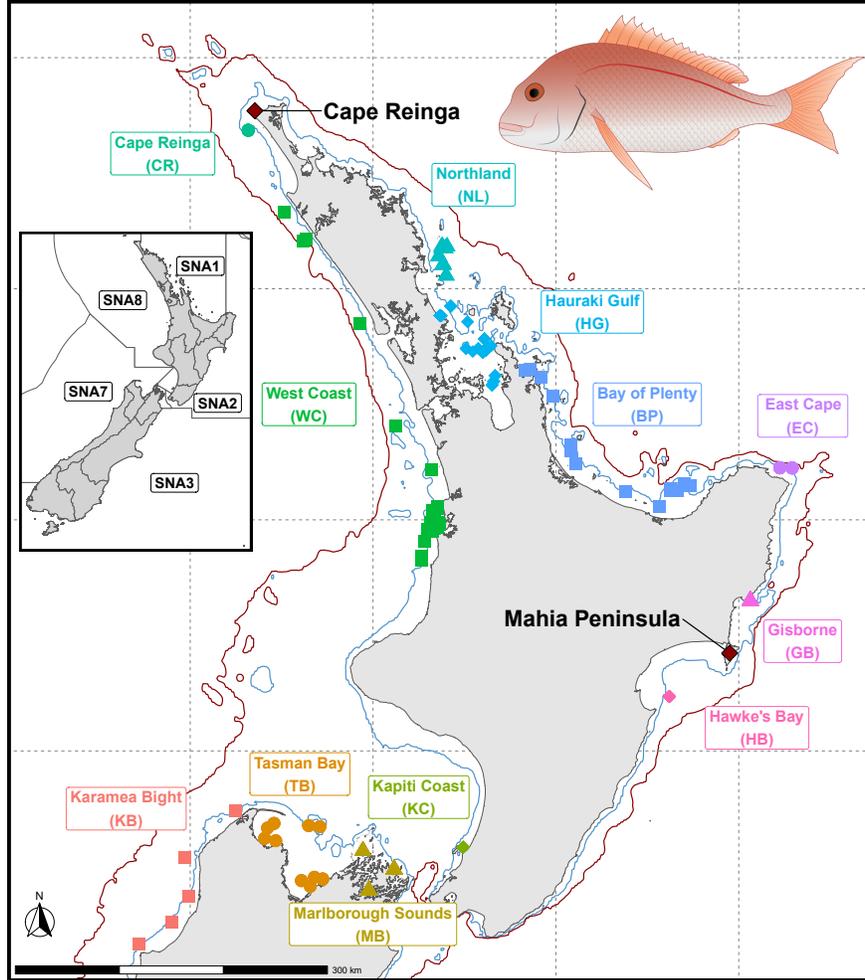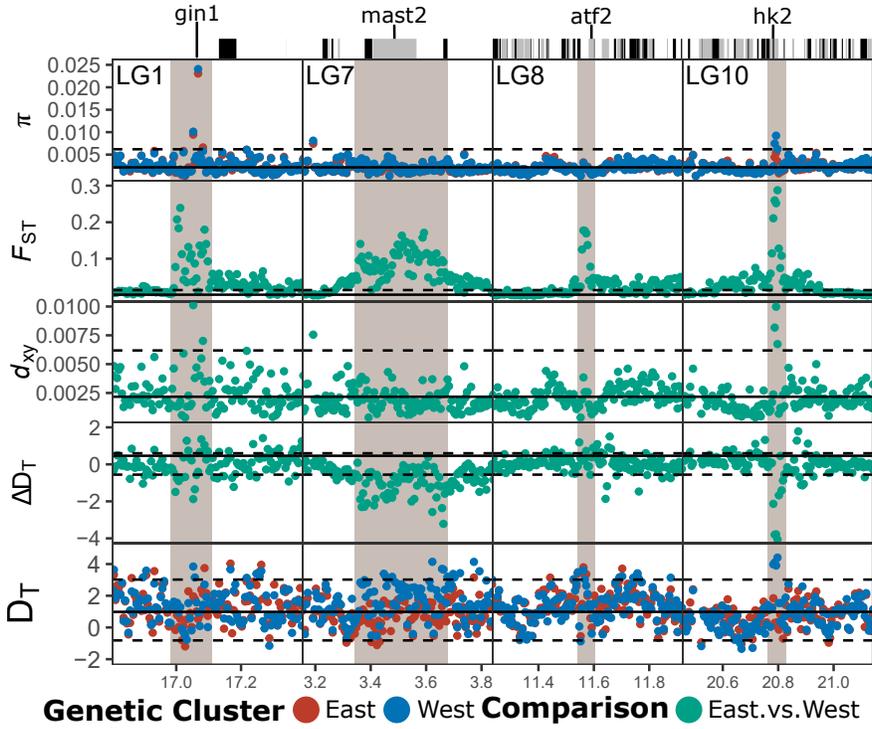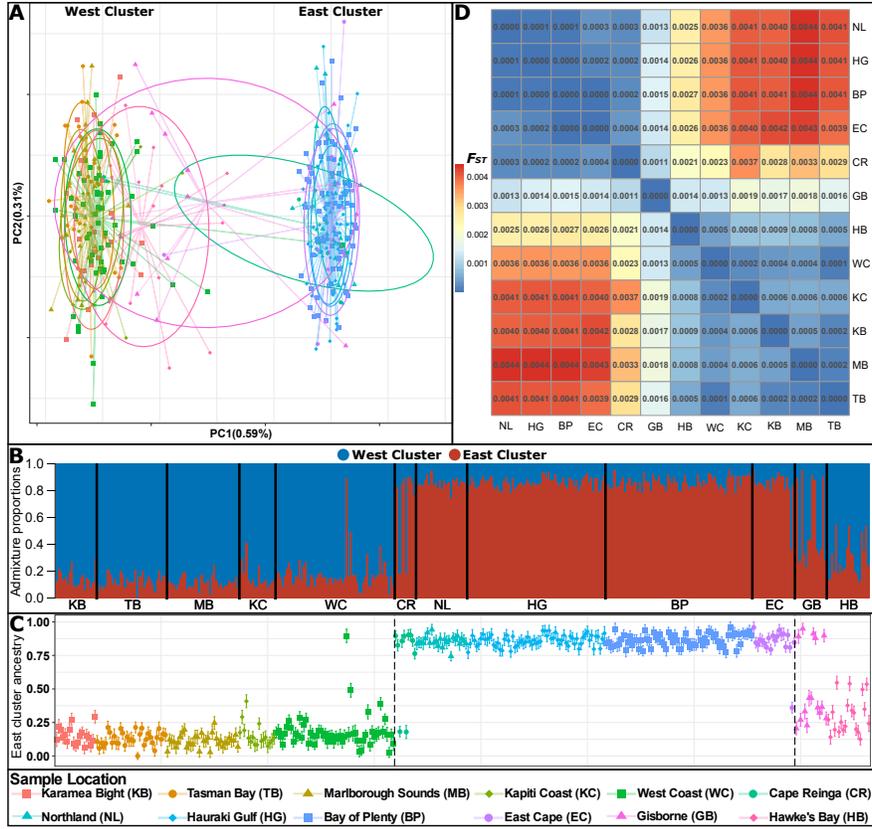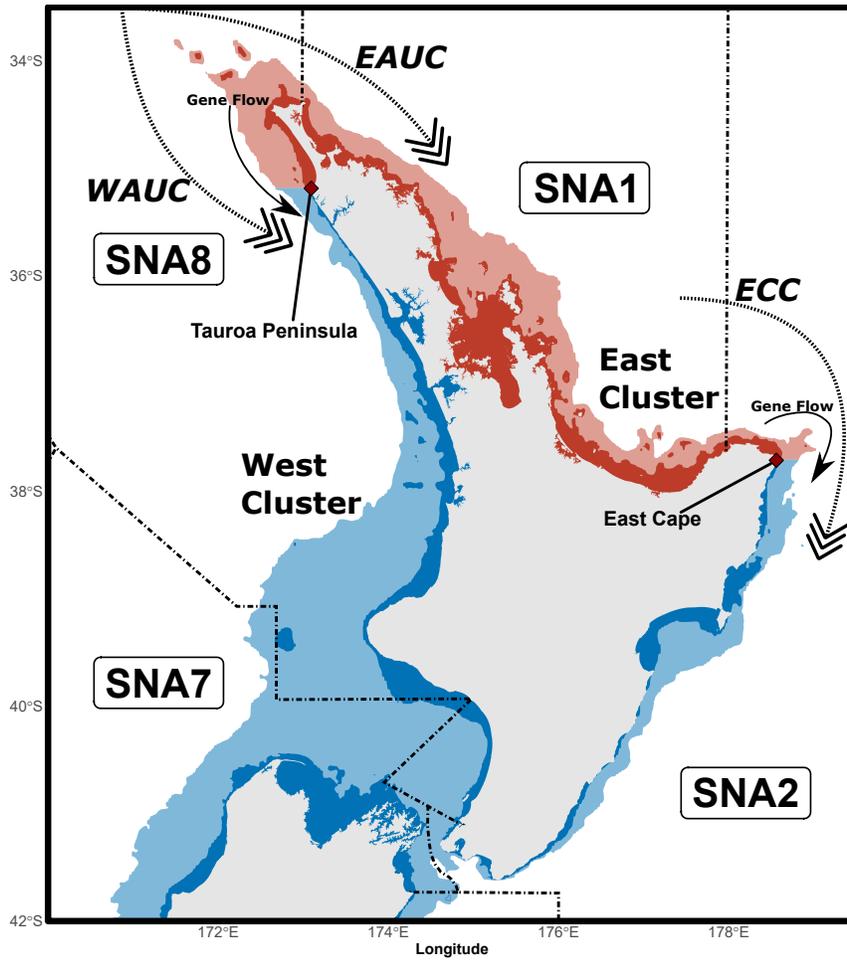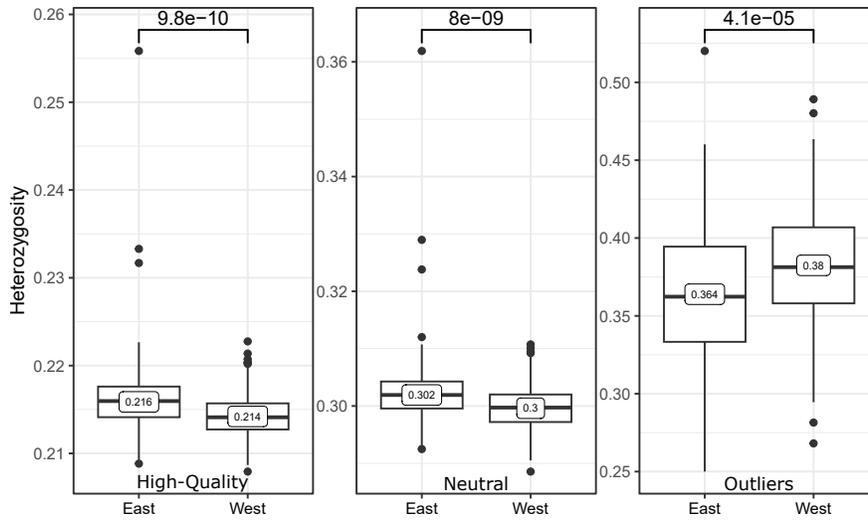
Genetic Cluster ● East ● West Comparison ● East.West