Received: 16 August 2021 | Revised: 14 October 2021 | Accepted: 18 October 2021

DOI: 10.1111/eva.13317

ORIGINAL ARTICLE

Evolutionary Applications Open Access WILEY

Genome-wide analysis reveals the genetic stock structure of hoki (*Macruronus novaezelandiae*)

Emily Koot[1] | Chen Wu[2] | Igor Ruza[3] | Elena Hilario[2] | Roy Storey[4] | Richard Wells[5] | David Chagné[1] | Maren Wellenreuther[3,6]

[1]The New Zealand Institute for Plant and Food Research Ltd, Palmerston North, New Zealand

[2]The New Zealand Institute for Plant and Food Research Ltd, Auckland, New Zealand

[3]The New Zealand Institute for Plant and Food Research Ltd, Nelson, New Zealand

[4]The New Zealand Institute for Plant and Food Research Ltd, Te Puke, New Zealand

[5]Deepwater Group, Auckland, New Zealand

[6]School of Biological Sciences, The University of Auckland, Auckland, New Zealand

Correspondence
Maren Wellenreuther, The New Zealand Institute for Plant and Food Research Ltd, Nelson, New Zealand.
Email: maren.wellenreuther@plantandfood.co.nz

Funding information
Seafood Innovations Limited

Abstract

The assessment of the genetic structuring of biodiversity is crucial for management and conservation. This is particularly critical for widely distributed and highly mobile deep-water teleosts, such as hoki (*Macruronus novaezelandiae*). This species is significant to Māori people and supports the largest commercial fishery in New Zealand, but uncertainty about its stock structure presents a challenge for management. Here, we apply a comprehensive genomic analysis to shed light on the demographic structure of this species by (1) assembling the genome, (2) generating a catalogue of genome-wide SNPs to infer the stock structure and (3) identifying regions of the genome under selection. The final genome assembly used short and long reads and is near complete, representing 93.8% of BUSCO genes, and consisting of 566 contigs totalling 501 Mb. Whole-genome re-sequencing of 510 hoki sampled from 14 locations around New Zealand and Australia, at a read depth greater than 10×, produced 227,490 filtered SNPs. Analyses of these SNPs were able to resolve the stock structure of hoki into two genetically and geographically distinct clusters, one including the Australian and the other one all New Zealand locations, indicating genetic exchange between these regions is limited. Location differences within New Zealand samples were much more subtle (global F_{ST} = 0.0006), and while small and significant differences could be detected, they did not conclusively identify additional substructures. Ten putative adaptive SNPs were detected within the New Zealand samples, but these also did not geographically partition the dataset further. Contemporary and historical N_e estimation suggest the current New Zealand population of hoki is large yet declining. Overall, our study provides the first genomic resources for hoki and provides detailed insights into the fine-scale population genetic structure to inform the management of this species.

KEYWORDS
genomics, hoki, *Macruronus*, population, single nucleotide polymorphism, stock structure

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2021 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd

1 | INTRODUCTION

Defining biologically meaningful units for sustaining biodiversity is one of the major goals of population management and conservation biology (Allendorf et al., 2010b; Moritz, 1994). In particular, the detection of genetic structure provides a crucial tool to identify isolated units, and to assess the degree of connectivity among populations (Bernatchez, 2016). Neglecting consideration of population structure may increase risks of overexploitation or mismanagement (Waples, 1998). Marine ecosystems are traditionally considered to be highly connected, and this is typically attributed to the large population sizes of many marine species, coupled with the presence of few barriers to gene flow (Nielsen et al., 2009). In addition, in many marine teleost species, the early life history is characterized by the presence of a planktonic larval stage during which larvae can be transported over long distances by ocean currents (e.g. 100s of km). Based on these characteristics, it is expected that marine species are usually highly connected owing to the combination of large population sizes and high dispersal, which makes it more challenging to characterize population structures accurately.

Small datasets containing neutral loci have been widely used to analyse population structure, gene flow and demographic changes over time. However, the small numbers of markers often lacked the statistical power to detect low rates of genetic differentiation in the high gene flow environments common for marine teleost species (Nielsen et al., 2009). Recent population genomic approaches employing thousands of genome-wide markers hold promise to provide the degree of resolution required for essentially any socioeconomic or ecologically important marine species (Ellegren, 2014). While the field is still evolving, numerous studies on marine species already demonstrate its potential to offer deeper insight into the dynamics of natural populations (Hohenlohe et al., 2010; Larson et al., 2014).

The ability to apply large numbers of DNA markers to conduct dense genome scans not only has greatly enhanced the power to identify genomic regions exhibiting genetic structure (Nielsen et al., 2012), but also has enabled identification of outlier regions associated with adaptation. Adaptive genomic signatures may be associated with local adaptation or reveal traces of cryptic population structure obscured by gene flow across most of the genome (Duranton et al., 2018; Gagnaire et al., 2015). These outlier loci can reveal regions in the genome of genetic differentiation where neutral markers often remain uninformative, and can prove useful to delineate locally adapted stocks and redefine conservation units (Funk et al., 2012). This approach is appealing because selection may be more efficient than genetic drift in opposing the homogenizing effect of migration, in particular when populations have large effective population sizes, which is the case for many important fisheries species. Genome scans work by detecting significant departures from genomic background patterns observed (Ahrens et al., 2018), while Gene-Environment-Associations (GEA) methods work by identifying genetic variants associated with particular environmental factors (Dallaire et al., 2021). However, outlier loci can also arise through a wide variety of evolutionary mechanisms apart from local adaptation (Bierne et al., 2011), in particular in response to varying patterns of recombination (Booker et al., 2020). Such patterns are commonly caused by structural genomic variants (Mérot et al., 2020; Wellenreuther & Bernatchez, 2018; Wellenreuther et al., 2019), and recent work indicates that these can affect more base pairs than SNP variants (Catanach et al., 2019), and are widespread throughout the genome.

Here, we assess the population genomic structure of hoki (*Macruronus novaezelandiae*, Family: Merlucciidae), which supports one of the most valuable deep-water fisheries in New Zealand. Hoki are widely distributed throughout New Zealand and Australian waters, being found in most abundant quantities in depths of 200–800 m (Horn & Sullivan, 1996; Livingston & Schofield, 1996). They have long pelagic larval and juvenile phases, maturing at the age of four, and exhibit extensive migratory behaviours (Horn, 2011). The current stock assessment for hoki in New Zealand is based on an assumed two-stock migration model between the Western and Eastern stocks (Figure 1). These have been defined based on data showing that fish in different geographical locations grow and mature at different rates and have different morphometric characters (Horn & Sullivan, 1996; Livingston & Schofield, 1996; Livingston & Sullivan, 2007). The Western stock encompasses spawning hoki from the West Coast of the South Island of New Zealand. The larvae and juveniles originating from this stock are thought to be then transported to the East Coast nurseries feeding areas on the Western Chatham Rise via the Westland and D'Urville currents (Smith et al., 1996). As young adults, these fish are then thought to migrate from the Chatham Rise to feeding grounds in sub-Antarctic waters, and subsequently moving between these and the spawning grounds on the West Coast of the South Island as mature adults. In contrast, hoki belonging to the Eastern stock are thought to spawn in the Cook Strait (and at other locations east of the South Island), with larvae and juveniles then migrating from these locations to the Chatham Rise nursery and feeding grounds. Juveniles are assumed to recruit to their respective stocks at maturity at ages of 3–8 years (O'Driscoll, 2004), and mature adults are thought to move between the Chatham Rise and the known spawning grounds in Cook Strait and at Pegasus Canyon. In the spawning grounds, hoki typically form large midwater aggregations, consisting almost entirely of the species.

Modern genomic technologies offer a powerful toolset to independently determine stock structure in this species, but has not been used to date. To enable assessment of this species using genomics solutions and given the significance of this species to Māori, Te Ohu Kaimoana and kaitiaki (cultural guardians) provided advice on a process to manage the gathering, storage, access and use of genetic data. We assembled a high-quality de novo genome using a combination of short- and long-read sequencing. To investigate the genomic stock structures, we sampled 12 locations in New Zealand and two locations in Australia (≥30 individuals per location) and performed whole genome sequencing at greater than 10× coverage to generate a powerful genome-wide SNP dataset. This genome-wide

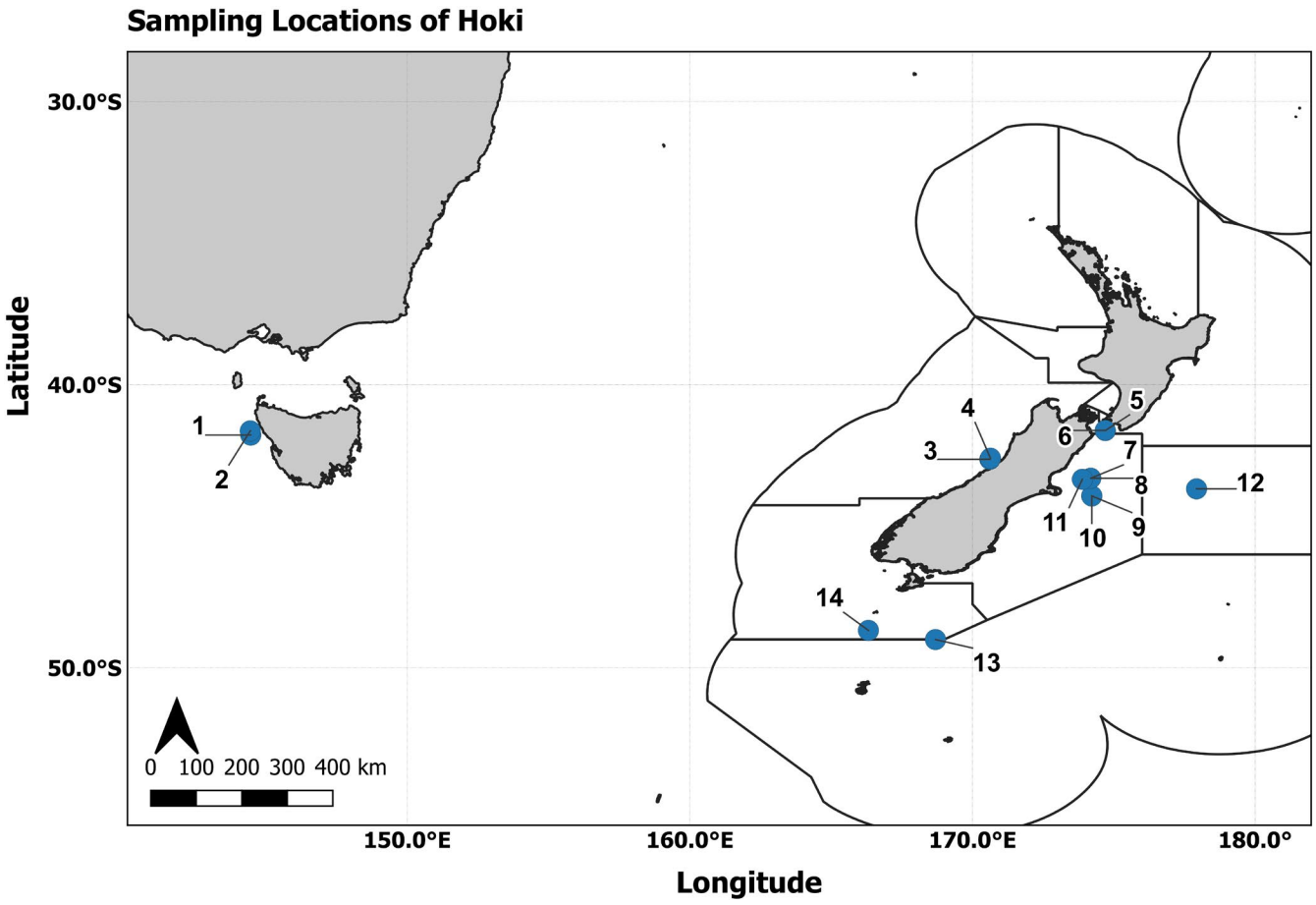

FIGURE 1 Sampling locations of hoki in New Zealand and Australia used for the population genomic part of this study. See Table 1 for site information

dataset was then used to assess the degrees of genetic diversity between all sampling sites to identify independent clusters, and how these are related to other sampling locations. The dataset was further used to test for population genetic differentiation using both neutral SNPs and putative adaptive SNPs. Our results are compared and discussed in the light of previous studies on hoki, and other teleost species.

2 | MATERIALS AND METHODS

2.1 | Indigenous considerations

Gathering genomic data requires strong relationships with indigenous peoples (Hudson et al., 2020). Engagement with Te Ohu Kaimoana started before the operational aspects of the project were initiated and carried on throughout the project. Specifically, a working group was developed to advise on a process to manage the gathering, storage, access and use of samples and genetic data before, during and after the research presented here. An agreement considering Māori tikanga (traditional protocols) was set up for the genomic data lifecycle, and the raw and analysed data were placed in a managed repository based in New Zealand.

2.2 | Sample and metadata collection and DNA extraction

A single adult female hoki from the Cook Strait was sampled in May 2020 (commercial fishing vessel FV Otakou) to generate DNA for the genome assembly. Five tissues were preserved in RNAlater following the manufacturer's instructions: brain, gills, liver, heart and white muscle. High-quality DNA was extracted from preserved liver with a CTAB-based extraction buffer as follows: 100 mg of tissue was homogenized in 1 ml CTAB buffer (2% CTAB [hexadecyltrimethyl ammonium bromide], 2% PVP K40 [polyvinyl pyrrolidinone K40], 2 M NaCl, 25 mM EDTA, 100 mM Tris-HCl pH 8.0) with a sterile plastic pestle, on ice. After adding 25 μl Proteinase K (20 mg/ml) and 10 μl 1 M dithiothreitol, the mixture was incubated at 50°C for 18 h in a thermomixer 30 s on, 300 rpm, 30 min off. The sample was extracted once with an equal volume of chloroform:isoamyl alcohol (24:1 v/v), by gently mixing by hand for 5 min. The phases were separated by centrifugation (5 min at 16,089 *g* at room temperature), and the aqueous phase was transferred to a new tube and precipitated with 2 volumes of 100% ethanol at room temperature. After ~5 min, the DNA was collected by centrifugation (10 min at 13,000 rpm at room temperature). The pellet was dissolved in 400 μl TE buffer

TABLE 1 Description of the hoki sampling

Sample site map ID	Sampling site name	Location	Stage	Region	Latitude	Longitude	Vessel	Time of landing	Number of individual fish sequenced
1	Tasmania A1	Tasmania	Adult	Tasmania	41 47.014 S	144 28.008 E	Tokatu	Jul-20	30
2	Tasmania A2	Tasmania	Adult	Tasmania	41 38.057 S	144 27.019 E	Tokatu	Jul-20	30
3	West Coast 1	Hokitika canyon	Spawning fish	West coast south island	42 38.32 S	170 36.87 E	Thomas Harrison	Jul-20	40
4	West Coast 2	Hokitika canyon	Spawning fish	West coast south island	42 36 S	170 38 E	Thomas Harrison	Jul-20	40
5	Cook Strait 1	Cook Strait St Bank	Spawning fish	Cook Strait	41 37.4 S	174 41.8 E	Otakou	Jul-20	40
6	Cook Strait 2	Cook Strait St Bank	Spawning fish	Cook Strait	41 37.1 S	174 41.7 E	Otakou	Jul-20	40
7	FMA3 J1	West Mernoo	Juvenile	FMA3	43 18.4 S	174 10.9 E	San Discovery	Aug-20	40
8	FMA3 A1	West Mernoo	Adult	FMA3	43 18.4 S	174 10.9 E	San Discovery	Aug-20	40
9	FMA3 J2	West Mernoo	Juvenile	FMA3	43 56.0 S	174 13.2 E	San Discovery	Aug-20	40
10	FMA3 A2	West Mernoo	Adult	FMA3	43 56.0 S	174 13.2 E	San Discovery	Aug-20	40
11	FMA3 A3	Pegasus Canyon	Adult	FMA3	43 20.7 S	173 53.1 E	San Enterprise	Jul-20	30
12	FMA4	North Rise	Adult	FMA4	43 40.986 S	177 55.488 E	San Enterprise	Jun-20	40
13	FMA6 Snare	Snares Shelf	Adult	FMA6	49 0.29 S	168 41.4 E	Amaltal Enterprise	Jul-20	30
14	FMA6 Norgie	Norgie hole	Adult	FMA6	48 40.9 S	166 19.1 E	Amaltal Enterprise	Jul-20	30

(10 mM Tris-HCl, 1 mM EDTA pH 7.5) and 100 µl 5 M NaCl. RNA was digested by adding 8 µl of RNase A (50 mg/ml) and incubating at room temperature for 5 min. Twenty-five microlitres of 10% SDS was added, and the sample was incubated in the thermomixer at 37°C at 300 rpm for 15 min. The sample was extracted with an equal volume of chloroform:isoamyl alcohol (24:1 v/v), mixing gently by hand for 5 min, and the phases were separated as described above. The DNA contained in the aqueous phase was precipitated with 0.7 volumes isopropanol, mixed, and incubated at room temperature for 10 min. The DNA was collected by centrifugation (10 min at 13,000 rpm at room temperature). The pellet was washed with 1 ml 70% ethanol and centrifuged again. The almost invisible pellet was dissolved in 75 µl TE buffer by letting it resuspend overnight at 4°C. The quality of the DNA was assessed by absorbance (A260/280 = 1.91, A260/230 = 2.35) and gel electrophoresis (average size 40 kbp) and quantified with a high sensitivity fluorescent-based method.

For the population genomics analysis, individual fish were sampled from commercial fishing vessels in New Zealand and Australian waters between May and September 2020 (Figure 1). In total, two sets of 100 individual samples were collected per trawl and data were recorded on fish size and sex. An individual larger than 75 cm was considered to be an adult, and smaller individuals were classed as juveniles but sufficiently old to have separated as per the stock model hypothesis. As fish under 45 cm were still deemed to be at risk of being from either presumptive stock, these were not sampled. Some adults were classed as spawning individuals when they were collected during July and August at known spawning sites, and these were as follows: Hokitika Canyon, Kahurangi (both these are off the West Coast of the South Island of New Zealand) and Cook Strait and Pegasus Canyon (the latter off the East Coast of the South Island). The GPS location of trawls was recorded and used as location indicators. One fin clip was collected from each fish and stored in 95% ethanol in a screw-cap tube at room temperature until DNA extraction. A random subset of 30 or 40 fish was used from 14 collection sites for the DNA extraction and sequencing. The collection sites were selected to represent a geographic distribution, including various life stages. A description of the 510 hoki samples can be seen in Table 1. DNA extraction was performed at Neogen GeneSeek, Lincoln, NE, USA, based on the LGC sbeadex™ magnetic bead kit that uses a two-step binding mechanism to provide high-quality DNA for downstream SNP and NGS protocols (www.lgcgroup.com).

2.3 | Genome sequencing and assembly

High molecular weight DNA was sent in July 2020 to the Australian Genome Research Facility (AGRF) for short- and long-read sequencing. The DNA was quality checked by AGRF upon arrival, and passed quality thresholds. The TrueSeq DNA Nano kit (Illumina) was used for library preparation for the Illumina sequencing, and the Illumina NovaSeq 6000 was used for

short-read sequencing (150 bp PE reads). For long-read sequencing, the Oxford Nanopore Technologies (ONT) technology was used. Three MinION FLO-MIN106 flow cells were run in-house and at AGRF, to verify that DNA could be sequenced with this technology, which produced approximately 3 Gb of sequencing data. Then, ONT libraries were prepared at AGRF using the SQK-LSK109 kit and run using a PromethION FLO-PRO002 flow cell, using minKNOW version 4.0.5 and base calling using Guppy 4.0.11, producing 7.57 M reads totalling 19.7 Gb of data (passed reads).

Kmer analysis was performed using jellyfish v2.2.10 with kmer size of 21. The ONT reads generated in-house (MinION) were base-called using Guppy v4.2.2 with parameters '--compress_fastq --input_path --save_path --flowcell FLO-MIN106 –kit SQK-LSK109 -x "cuda:0"' and quality-assessed using pycoqc v2.5.0.21. The long-read assembly was generated using FLYE v2.8.1 with default parameters based on long reads generated from both in-house (MinION) and at AGRF (MinION and PromethION). It was then subjected to two rounds of short-read error-correction using PILON v1.23 with parameters '--genome FLYE.fasta --frags mapped.bam --output --outdir --changes --diploid --fix all --threads 60 --flank 5'. Assembly completeness was assessed using BUSCO v5.1.2 with parameters '-l actinopterygii_odb10 -o test_long -m geno --augustus --augustus_species zebrafish -c 20 --long --out_path busco/' (Simão et al., 2015).

2.4 | Whole genome sequencing, variant calling, and filtering

Illumina paired-end (PE) short reads were generated at Neogen GeneSeek. In total, 510 fish samples were sequenced with a target read depth greater than 10×. A summary of the sequencing data is in Table S1. Variant calling was performed using the hoki genome assembled from the ONT data as a reference. Illumina PE reads were mapped to the reference genome using BWA-mem (v0.7.17), and the resulting sam files were converted to bam files using Samtools v1.9 (Li et al., 2009). SNP calling was performed using Freebayes v1.1.0 (Garrison & Marth, 2012) with the following parameters: -p 2 –C3 –m 10 --min-coverage 15 and --max-coverage 500. Variant filtering was performed using vcftools v0.1.14 (Danecek et al., 2011), applying the following filters: --max-alleles = 2, --max-missing = 0.95, --maf = 0.02, --remove-indels.

2.5 | Outlier scans for stock structure assessment

Outlier SNPs were identified using three methods – tess3 (Caye et al., 2016), pcadapt (Luu et al., 2017) and LEA (Frichot & François, 2015). Tess3 searches for genomic variants under selection by applying matrix factorization algorithms to allele frequencies; pcadapt runs genome-wide selection scans based on principal component analyses; and LEA calculates F_{ST} statistics from ancestral allele frequencies that are estimated using the packages snmf function. Tess3 was implemented using the tess3r package in R v3.6.1 applying the following parameters: method = "projected.ls", ploidy = 2, max.iteration = 5000, rep = 10, keep = "best", tolerance = 1e−05. The R package pcadapt v4.3.3 was implemented using default settings and method = "mahalanobis". The following parameters were used to run LEA's snmf function: genotype, K = 1:14, entropy = T, ploidy = 2, repetitions = 10, tolerance = 1e−05. The three analyses were conducted on two different datasets – the first contained all filtered SNPs and all hoki samples (510 samples); the second contained all filtered SNPs but only New Zealand samples (450 samples). A Q-value of 0.05 was used as a threshold of statistical significance for all three analyses. Q-values were calculated using the qvalue R package (Dabney et al., 2010). Variants that met this threshold for all three analyses were considered as outliers that were putatively under divergent selection, whilst any variant that was not considered under selection by any of the analyses was considered to be putatively neutral.

2.6 | Stock structure and size

Genetic diversity and population structure were investigated using six datasets: (1) all SNPs and all hoki samples, (2) all SNPs and New Zealand-only samples, (3) neutral SNPs and all hoki samples, (4) neutral SNPs and New Zealand-only samples, (5) adaptive SNPs and all hoki samples, and (6) adaptive SNPs and New Zealand-only samples. Nucleotide diversity (π), observed (H_o) and expected (H_e) heterozygosity and Tajima's D statistics were calculated for each sampling site and for each of the six datasets using vcftools v0.1.14. Weir and Cockerham's (1984) pairwise F_{ST} distances and accompanying p-values were also estimated for each dataset and each sampling site using R package StAMPP (Pembleton et al., 2013), applying: nboots = 1000, per cent = 95. Population structure was investigated using the adegenet R packages find.clusters and DAPC functions (Jombart & Ahmed, 2011). Find.clusters was used to initially explore the optimal K value (number of ancestral clusters) for each dataset before the DAPC analysis was run using this optimized K value, in addition to an optimized number of principal components (PCs). The snmf function in the R package LEA was used to explore the ancestral admixture (Frichot & François, 2015). A range of K values were explored (K = 1:14), in addition to the following parameter settings: entropy = T, ploidy = 2, repetitions = 10, tolerance = 0.00001. For the DAPC and LEA analyses, individual analyses were run for each of the six datasets. Contemporary effective population size (N_e) was estimated for New Zealand samples using the linkage disequilibrium (LD) method in NeEstimator v2 (Do et al., 2014), whilst SNeP v1.1 was used to estimate historical N_e (Barbato et al., 2015). The complete SNP dataset for New Zealand samples was thinned by 10,000 sites in vcftools before contemporary and historical N_e values were estimated.

3 | RESULTS

3.1 | Reference genome assembly

In total, 22.7 Gb of ONT long-read sequencing data was generated. PromethION and MinION sequencing resulted in 19.69 and 3.01 Gb of passed reads with N50 greater than 4.9 and 5.2 kb, respectively. Assembly using the ONT data and using the FLYE software resulted in a total of 501,488,236 bp assembled into 566 scaffolds (Table 2). The N50 of scaffolds was greater than 11 Mb and 15 scaffolds accounted for more than half the total assembly (L50). The largest scaffolds were approximately 26 Mb in length. BUSCO analysis after polishing using Illumina short reads and using PILON resulted in 93.8% complete BUSCOs (0.9% duplicated) out of 3640 zebrafish conserved genes.

3.2 | Individual hoki samples sequencing, variant calling and filtering

A total of 6148 Gb of Illumina sequencing data were generated for all 510 sampled fish (Table S1), corresponding to a mean of 12.05 Gb of data per individual (median 11.68 Gb), which corresponds to a mean read depth of ~24× based on the assembled 501 Mb hoki genome. Only four individuals had less than 9 Gb of total data and the individual with the lowest yield had greater than 5 Gb of data, which is still greater than the target 10× coverage. The mapping rate to the reference genome ranged between 81.5% and 94.9%, with a median of 93%, and only seven individuals had mapping rates lower than 90%. In total, 3,906,729 raw variants were detected of which 227,490 were retained after filtering.

3.3 | Outlier scans

Tess3 identified 8268 SNPs under selection in the dataset containing all samples, and 2862 in the dataset consisting of only New Zealand samples (Table 3; Figure S1). Pcadapt identified 3502 SNPs and 3720 in each dataset, respectively. LEA identified 9778 SNPs and 5396 in each dataset, respectively. 2797 SNPs were identified as outliers across all three analyses for the complete dataset, whilst only 10 SNPs were identified across all analyses for the New Zealand-only dataset (A list of outlier SNPs is available upon request). This subsequently resulted in the two neutral datasets having 216,572 SNPs (all hoki) and 216,974 SNPs (New Zealand only) each.

TABLE 2 Assembly metrics for the hoki reference genome

Number of scaffolds	566
Total size of scaffolds	501,488,236 bp
Longest scaffold	25,972,667 bp
Mean scaffold size	886,022 bp
N50 scaffold length	11,052,189 bp
L50 scaffold count	15
Complete BUSCOs (single)	92.90%
Complete BUSCOs (duplicated)	0.90%
Fragmented BUSCOs	1%
Missing BUSCOs	5.20%

3.4 | Population structure analyses

3.4.1 | All SNPs and all hoki samples

Mean nucleotide diversity across New Zealand sampling sites ranged from 0.163 to 0.169 (Table 4). Mean nucleotide diversity for the two Australian sites was slightly higher (0.189 and 0.203). Mean observed heterozygosity for New Zealand sites ranged from 0.173 to 0.188 and was greater than expected heterozygosity in all instances. Mean observed heterozygosity for Australian sites was higher (0.234 and 0.260) and also higher than expected heterozygosity. Mean Tajima's *D* was negative across all New Zealand sites and one Australian site (Tasmania A2); however, it was positive for the other (Tasmania A1) site. Pairwise F_{ST} was examined between all sampling sites, F_{ST} estimates were very low among New Zealand sites, ranging from 0.00027 ($p < 0.001$) to 0.00164 ($p < 0.001$) (mean = 0.00067), with pairwise F_{ST} between the two Australian sites being higher (0.0025, $p < 0.001$) (Table 5). Pairwise F_{ST} distance between all New Zealand and all Australian sites was 0.02139 ($p < 0.001$), and when the F_{ST} genetic distances were examined as a heatmap, clear genetic differentiation between New Zealand and Australian samples was visible (Figure 2a).

Kmeans clustering analysis (using find.clusters) of all hoki samples and all SNPs identified two clusters ($K = 2$) within the dataset (Figure S2a,b). When further examined via DAPC analysis, the two clusters clearly identified two geographically and genetically distinct groups – one containing all New Zealand hoki samples and one containing all Australian hoki samples (Figure 2b). This was further supported by the LEA ancestral admixture analysis, which also identified $K = 2$ as the optimal number of clusters (Figure S2c), and also identified the New Zealand and Australian samples as belonging to two separate clusters (Figure 2c).

The complete SNP dataset was also examined with only New Zealand samples included to determine whether there was any additional structure among the New Zealand sampling sites. The kmeans clustering analysis determined there was one cluster ($K = 1$) within the dataset (Figure S3a,b), with the DAPC scatterplot displaying overlap between all sampling sites (Figure S4). FMA6 (FMA denotes Fisheries Management Areas) Snares and FMA3 J1 did, however, display a low rate of genetic divergence from the other sites when DA 1 and DA 2 were examined. However, it should be noted here that a replicate trawl from FMA3 did not show the same pattern as FMA3 J1. The LEA ancestral admixture analysis also determined that the optimal number of ancestral populations for the New Zealand samples was $K = 1$ (Figure S3c).

TABLE 3 Outlier SNPs

New Zealand and Australia	Threshold	<1e−04	<0.001	<0.01	<0.025	<0.05	<0.1
tess3	*p*-Value	4555	7227	13,203	18,772	25,985	37,936
	q-Value	2045	3221	5364	6651	8268	10,502
	Local FDR	1467	2230	3662	4516	5323	6401
pcadapt	*p*-Value	2191	3264	5855	8283	11,373	17,065
	q-Value	1286	1698	2404	2929	3502	4328
	Local FDR	1043	1275	1722	1975	2304	2733
LEA	*p*-Value	5687	8274	14,656	19,908	26,346	36,523
	q-Value	3506	4616	6708	8094	9778	12,321
	Local FDR	2727	3508	4739	5565	6383	7498
						2797	
New Zealand only	**Threshold**	**<1e−04**	**<0.001**	**<0.01**	**<0.025**	**<0.05**	**<0.1**
tess3	*p*-Value	909	3696	11,153	17,031	24,175	35,554
	q-Value	4	11	125	921	2862	6068
	Local FDR	0	9	60	436	1411	3299
pcadapt	*p*-Value	1419	3173	8000	11,546	15,578	21,778
	q-Value	399	718	1496	2369	3720	5985
	Local FDR	284	459	921	1313	1947	3219
LEA	*p*-Value	2130	5060	12,489	18,347	24,782	35,077
	q-Value	140	554	2058	3559	5396	8421
	Local FDR	101	346	1132	1998	3030	4589
						10	

Note: Results of tess3, pcadapt and LEA for all sampling sites (top table) and New Zealand sites (bottom table). SNPs meeting the 0.05 *q*-value threshold are highlighted in grey, with the total number of intercepting SNPs noted at the bottom of this column.

3.4.2 | Neutral SNPs for all hoki and New Zealand only

Mean nucleotide diversity across neutral SNPs and all hoki ranged from 0.166 to 0.171 within New Zealand sites, and slightly higher for the two Australian sites (0.181–0.195) (Table S2). Mean observed heterozygosity for the New Zealand sites ranged from 0.176 to 0.191, and from 0.225 to 0.249 for the two Australian sites. Observed heterozygosity was higher than expected heterozygosity in all instances. Mean Tajima's *D* was negative for all sites and ranged from −0.424 to −0.307 for the New Zealand sites, and from −0.232 to −0.095 for the two Australian sites. Pairwise F_{ST} distance between all New Zealand and all Australian sites was 0.01026 ($p < 0.001$). Pairwise F_{ST} distance among New Zealand sites ranged from 0.00026 ($p < 0.001$) to 0.00163 ($p < 0.001$) (mean = 0.00067), and pairwise F_{ST} for the two Australian sites was 0.0024 ($p < 0.001$) (Table 5). Mean nucleotide diversity, heterozygosity and Tajima's *D* results for the New Zealand-only neutral SNPs were the same as those for the all-hoki neutral dataset (Table 4); however, F_{ST} distance among New Zealand sites was slightly lower, ranging from 0.00026 ($p < 0.001$) to 0.00158 ($p < 0.001$) (mean = 0.00062) (Table 6; Figure 3a).

Kmeans clustering, DAPC and LEA analysis results for both the neutral datasets (all hoki and New Zealand only) were very similar to the results of the complete dataset, with two clusters being identified in the dataset containing all sites (a New Zealand cluster and an Australian cluster), and one cluster in the dataset containing only New Zealand sites (Figure 3b,c; Figures S5 and S6). No clustering between juvenile and adult data was present.

3.4.3 | Adaptive SNPs for all hoki and New Zealand only

For the dataset containing adaptive SNPs for all hoki, mean nucleotide diversity for New Zealand sites ranged from 0.124 to 0.134, and from 0.419 to 0.442 for Australian sites (Table S2). Mean observed heterozygosity for New Zealand sites ranged from 0.126 to 0.141 and from 0.570 to 0.638 for the two Australian sites. Mean Tajima's *D* ranged from −0.522 to −0.401 for New Zealand sites, and from 1.341 to 1.495 for the two Australian sites. The pairwise F_{ST} estimates among New Zealand sites ranged from 0.0000 ($p = 0.83$) to 0.00237 ($p < 0.001$) (mean = 0.00079), and pairwise F_{ST} between the two Australian sites was 0.0027 ($p < 0.001$) (Table 5). Pairwise F_{ST} distance between all New Zealand and all Australian sites was much higher than for the complete SNP dataset ($F_{ST} = 0.32393$, $p < 0.001$), and when the F_{ST} genetic distances were examined as a heatmap, the genetic differentiation between New Zealand and Australian samples remained very clear. In addition, there appeared

TABLE 4 Diversity statistics for three of the datasets: all SNPs and all sampling sites; neutral SNPs and New Zealand sites only; and adaptive SNPs and New Zealand sites only

	π (mean)	H_o (mean)	H_e (mean)	Tajima's *D* (mean)
All SNPs, all sampling sites				
Cook Strait 1	0.169	0.188[a]	0.167	−0.342
Cook Strait 2	0.164	0.176[a]	0.162	−0.386
FMA3 A1	0.167	0.182[a]	0.165	−0.359
FMA3 A2	0.165	0.178[a]	0.163	−0.379
FMA3 A3	0.165	0.178[a]	0.162	−0.442
FMA3 J1	0.169	0.188[a]	0.167	−0.331
FMA3 J2	0.164	0.176[a]	0.162	−0.385
FMA4	0.163	0.173[a]	0.161	−0.397
FMA6 Norgie	0.164	0.175[a]	0.161	−0.449
FMA6 Snares	0.168	0.185[a]	0.165	−0.405
Tasmania A1	0.203	0.260[a]	0.199	0.012
Tasmania A2	0.189	0.234[a]	0.186	−0.140
West Coast 1	0.166	0.180[a]	0.164	−0.375
West Coast 2	0.166	0.179[a]	0.163	−0.377
Neutral SNPs, New Zealand only				
Cook Strait 1	0.171	0.191[a]	0.169	−0.317
Cook Strait 2	0.167	0.179[a]	0.165	−0.360
FMA3 A1	0.170	0.185[a]	0.167	−0.334
FMA3 A2	0.167	0.181[a]	0.165	−0.353
FMA3 A3	0.168	0.180[a]	0.165	−0.417
FMA3 J1	0.171	0.191[a]	0.169	−0.307
FMA3 J2	0.167	0.179[a]	0.165	−0.359
FMA4	0.166	0.176[a]	0.164	−0.371
FMA6 Norgie	0.167	0.178[a]	0.164	−0.424
FMA6 Snares	0.171	0.188[a]	0.168	−0.382
West Coast 1	0.168	0.183[a]	0.166	−0.348
West Coast 2	0.168	0.182[a]	0.166	−0.350
Adaptive SNPs, New Zealand only				
Cook Strait 1	0.155	0.150	0.159	−0.281
Cook Strait 2	0.092	0.098[a]	0.091	−0.653
FMA3 A1	0.192	0.200[a]	0.189	−0.060
FMA3 A2	0.107	0.115[a]	0.106	−0.565
FMA3 A3	0.146	0.130	0.143	−0.411
FMA3 J1	0.175	0.180[a]	0.173	−0.161
FMA3 J2	0.114	0.110	0.112	−0.524
FMA4	0.100	0.105[a]	0.099	−0.608
FMA6 Norgie	0.131	0.110	0.129	−0.497
FMA6 Snares	0.138	0.143[a]	0.135	−0.460
West Coast 1	0.166	0.178[a]	0.164	−0.213
West Coast 2	0.115	0.123[a]	0.114	−0.516

Note: H_e, expected heterozygosity; H_o, observed heterozygosity; π, nucleotide diversity.

[a]Observed heterozygosity is higher than expected heterozygosity.

TABLE 5 Pairwise F_{ST} for neutral (top matrix) and adaptive (bottom matrix) SNPs for all sampling sites

	Cook Strait 1	Cook Strait 2	FMA3 A1	FMA3 A2	FMA3 A3	FMA3 J1	FMA3 J2	FMA4	FMA6 Norgie	FMA6 Snares	Tasmania A1	Tasmania A2	West Coast 1	West Coast 2
Cook Strait 1		0.0004***	0.0005***	0.0004***	0.0007***	0.0009***	0.0005***	0.0006***	0.0008***	0.0013***	0.0117***	0.0085***	0.0005***	0.0004***
Cook Strait 2	0.0005		0.0004***	0.0004***	0.0005***	0.00010***	0.0003***	0.0004***	0.0007***	0.0013***	0.0131***	0.0094***	0.0003***	0.0003***
FMA3 A1	0.0006	0.0000		0.0004***	0.0007***	0.0010***	0.0004***	0.0005***	0.0007***	0.0012***	0.0124***	0.0089***	0.0003***	0.0003***
FMA3 A2	0.0004	0.0000	0.0000		0.0004***	0.0010***	0.0003***	0.0004***	0.0006***	0.0011***	0.0129***	0.0093***	0.0005***	0.0004***
FMA3 A3	0.0013*	0.0008	0.0012*	0.0014*		0.0010***	0.0004***	0.0005***	0.0004***	0.0009***	0.0128***	0.0093***	0.0006***	0.0005***
FMA3 J1	0.0012**	0.0012**	0.0013***	0.0012**	0.0016**		0.0010***	0.0010***	0.0012***	0.0016***	0.0122***	0.0091***	0.0010***	0.0009***
FMA J2	0.0015***	0.0008*	0.0001	0.0004	0.0022***	0.0016***		0.0003***	0.0005***	0.0012***	0.0131***	0.0095***	0.0004***	0.0004***
FMA4	0.0010**	0.0000	0.0002	0.0000	0.0016**	0.0012**	0.0004		0.0006***	0.0013***	0.0134***	0.0096***	0.0004***	0.0004***
FMA6 Norgie	0.0009	0.0004	0.0008	0.0000	0.0003	0.0015**	0.0009	0.0000		0.0009***	0.0130***	0.0093***	0.0008***	0.0006***
FMA6 Snares	0.0012*	0.0015**	0.0010	0.0009	0.0020***	0.0013**	0.0012*	0.0017***	0.0008		0.0124***	0.0094***	0.0012***	0.0011***
Tasmania A1	0.2395***	0.2511***	0.2452***	0.2487***	0.2300***	0.2394***	0.2490***	0.2510***	0.2334***	0.2269***		0.0024***	0.0128***	0.0128***
Tasmania A2	0.2152***	0.2272***	0.2213***	0.2242***	0.2075***	0.2155***	0.2252***	0.2266***	0.2110***	0.2039***	0.0027**		0.0093***	0.0093***
West Coast 1	0.0002	0.0000	0.0000	0.0002	0.0016**	0.0014***	0.0010**	0.0005	0.0001	0.0024***	0.2460***	0.2219***		0.0003***
West Coast 2	0.0004	0.0002	0.0002	0.0000	0.0013**	0.0012***	0.0008*	0.0000	0.0000	0.0007	0.24567***	0.2215***	0.0000	

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

to be some genetic differentiation occurring within the New Zealand samples, with FMA3 A3, FMA6 Snares and FMA6 Norgie forming their own subcluster within New Zealand; however, as there was poor support for many of these pairwise F_{ST} values, this should be interpreted cautiously (Figure S7).

For the dataset containing adaptive SNPs for only the New Zealand sites, mean nucleotide diversity ranged from 0.092 to 0.192 (Table 4). Mean observed heterozygosity ranged from 0.098 to 0.20, with observed heterozygosity being higher than expected heterozygosity for all sites except FMA3 J2, FMA6 Norgie, FMA3 A3 and Cook Strait 1. Mean Tajima's *D* was negative cross all sites and ranged from −0.653 to −0.028. Pairwise F_{ST} distance among New Zealand sites ranged from 0.0000 ($p = 0.92$) to 0.02358 ($p = 0.16$) (mean = 0.00517) (Table 6). When F_{ST} distances were plotted as a heatmap, some substructure among populations was visible, with FMA3 A1 and FMA3 J1 forming a cluster and appearing to be genetically separate from the remaining sites; however, again, due to poor support for many of these F_{ST} values, this should be interpreted cautiously (Figure 4).

Kmeans clustering, DAPC and LEA analysis results for both the adaptive SNP datasets (all hoki and New Zealand only) were very similar to the results of the complete and neutral datasets, with two clusters being identified in the dataset containing all sites (a New Zealand cluster and an Australian cluster), and one cluster in the dataset containing only New Zealand sites (Figure 4; Figures S8 and S9).

3.4.4 | Contemporary and historical N_e for New Zealand

After thinning the complete SNP dataset of New Zealand samples by 10,000 sites, 2731 SNPs remained for the NeEstimator and SneP analyses. NeEstimator estimated a contemporary LD N_e of 133,463.4 (95% CI: 38,997; Infinite) (MAF = 0.05). SNeP estimated historical N_e to be 228,681 999 generations ago (2997–7992 years ago), declining to an N_e of 42,816 189 generations ago (567–1512) years ago) (Table S3, Figure S10).

4 | DISCUSSION

Knowledge of the stock structure of important fisheries species is a key component that informs management plans. Marine fish populations typically consist of large numbers of individuals with high dispersal potential, which can result in high gene flow and weak population structure (Bernatchez et al., 2017; Papa et al., 2020). Genomic approaches provide an excellent tool to gain insights into the fine-scale population structure and can also shed light on the environmental drivers that impose selection on the genome (Bernatchez, 2016). Here we assembled a de novo reference genome and used genome-wide SNP data to evaluate the stock structure of the deep-water teleost species hoki (*M. novaezelandiae*). Results

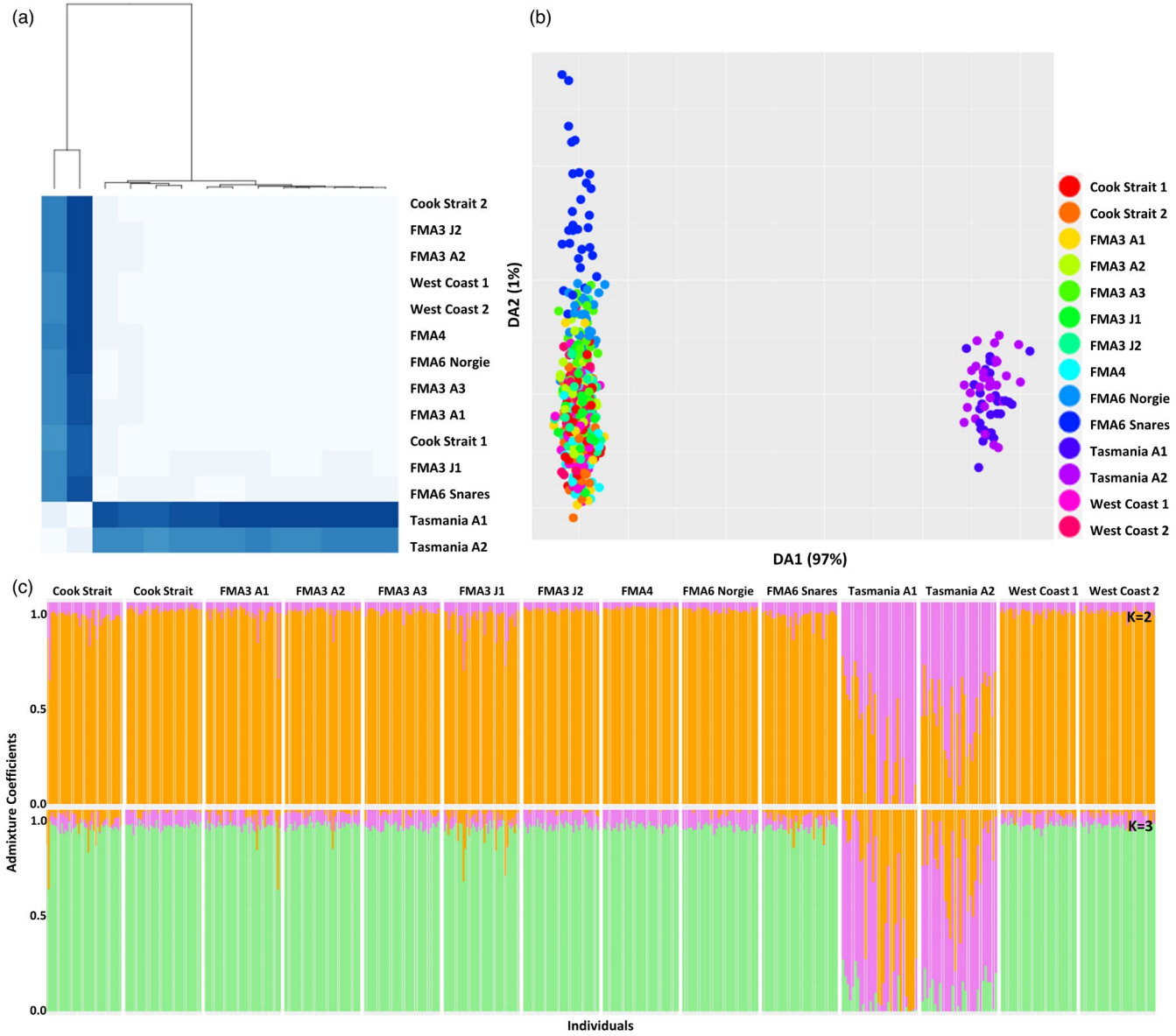

FIGURE 2 Population genomic analysis outputs for the hoki dataset consisting of all SNPs and all sampling sites: (a) pairwise F_{ST} heatmap with hierarchical clustering dendrogram—darker blues indicate higher pairwise F_{ST} values and lighter blues indicate lower pairwise F_{ST} values, (b) DAPC scatterplot of DA1 (97%) and DA2 (1%) with points coloured by sampling sites, and (c) LEA ancestral admixture plots for $K = 2$ and $K = 3$

showed two clear clusters, one grouping both Australian sites in Tasmania, and another one grouping all New Zealand sites. Within New Zealand, genetic differentiation between sampling locations was weak, indicating reproductive mixing between locations and no clear differentiation between different life history stages (adults vs. juveniles) within New Zealand waters.

The high gene flow and low genetic differentiation between populations characteristic of many teleost species (reviewed in Benestan, 2019; Nielsen et al., 2009) were detected here also for the New Zealand hoki sampling locations. Taking into account both neutral and adaptive SNPs, some evidence for increased genetic differentiation was identified for a small number of subclustering sampling locations based on pairwise F_{ST} values. When comparing the dataset including all samples (Australia and New Zealand) with the dataset for only the New Zealand samples, it became evident that the sampling locations within each subcluster were different (FMA3 A3, FMA6 Snares and FMA6 for the whole dataset, and FMA3 A1 and FMA3 J1 for the New Zealand dataset, Figure 4, Figure S4), suggesting the clusters were spurious. Furthermore, all three clustering approaches (DAPC, LEA and Kmeans clustering) using either the neutral or the adaptive dataset did not support F_{ST}-based subclusters, and detected only the two clusters differentiating the Australian and New Zealand sampling locations. Together, this evidence indicates that the F_{ST} differentiation within New Zealand waters is overall weak and does not carry a strong geographic signal.

TABLE 6 Pairwise F_{ST} for neutral (top matrix) and adaptive (bottom matrix) SNPs for New Zealand sampling sites

	Cook Strait 1	Cook Strait 2	FMA3 A1	FMA3 A2	FMA3 A3	FMA3 J1	FMA3 J2	FMA4	FMA6 Norgie	FMA6 Snares	West Coast 1	West Coast 2
Cook Strait 1		0.0004***	0.0005***	0.0004***	0.0006***	0.0009***	0.0005***	0.0005***	0.0007***	0.0012***	0.0005***	0.0003***
Cook Strait 2	0.0104*		0.0004***	0.0004***	0.0005***	0.0008***	0.0003***	0.0004***	0.0006***	0.0012***	0.0003***	0.0003***
FMA3 A1	0.0000	0.0176***		0.0004***	0.0006***	0.0009***	0.0003***	0.0004***	0.0006***	0.0012***	0.0003***	0.0003***
FMA3 A2	0.0007	0.0000	0.0178*		0.0004***	0.0009***	0.0003***	0.0004***	0.0005***	0.0011***	0.0005***	0.0004***
FMA3 A3	0.0000	0.0041	0.0028	0.0001		0.0009***	0.0004***	0.0004***	0.0004***	0.0009***	0.0006***	0.0005***
FMA3 J1	0.0057	0.0220	0.0063	0.0145	0.0000		0.0009***	0.0008***	0.0011***	0.0016***	0.0008***	0.0008***
FMA J2	0.0090	0.0053	0.0167	0.0067	0.0000	0.0036		0.0003***	0.0005***	0.0011***	0.0004***	0.0003***
FMA4	0.0076*	0.0000	0.0194***	0.0000	0.0003	0.0236	0.0045		0.0005***	0.0012***	0.0003***	0.0003***
FMA6 Norgie	0.0000	0.0000	0.0007	0.0000	0.0000	0.0003	0.0027	0.0056		0.0009***	0.0007***	0.0006***
FMA6 Snares	0.0051	0.0094	0.0080	0.0019	0.0025	0.0099	0.0136*	0.0001	0.0067		0.0012***	0.0010***
West Coast 1	0.0000	0.0057	0.0000	0.0009	0.0000	0.0000	0.0027	0.0032	0.0000	0.0000		0.0003***
West Coast 2	0.0045	0.0045	0.0115*	−0.0002	0.0005	0.0215	0.0074	0.0000	0.0097	0.0000	0.0039	

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

While panmixia of New Zealand hoki stocks was inferred, outlier scans including the two Australian sites revealed a large number of loci under putative selection. In total, 2797 SNPs could be identified as outliers that were in common across all three analysis methods. This high number of SNPs under putative selection indicates that the wider Australasian hoki population has a high potential for future genetic adaptation in response to changing environmental conditions. In the context of a rapidly changing world, this high standing adaptive potential may provide this species with a resilience buffer to combat future changes. In contrast to the large number of outlier loci in the overall dataset, when only New Zealand locations were analysed, the number of shared outlier loci between sampling locations fell to only 10 SNPs. This finding could indicate that large-scale environmental selection selects for different locally adapted genome associations in each cluster; however, further analysis is required to better understand this small number of shared outlier SNPs. Our finding of multiple outlier loci, particularly in the complete dataset, is not in line with previous views on the adaptive ability of marine teleosts (Waples, 1998). In the past, adaptive divergence was widely believed to be rare in marine fishes owing to fewer barriers to gene flow in the marine environment, the often-long dispersal phase of larvae, and also migratory adults; together, these factors were thought to hinder local adaptation (Kawecki & Ebert, 2004). Our results, and those of other studies in recent years (see for a review Bernatchez, 2016), strongly indicate that this view is no longer warranted and that many marine fishes exhibit signatures of local adaptation. Indeed, recent work suggests that the often-large population sizes of marine fishes may be fuelling selection, particularly when selection is stable and consistent over large areas, and the selection differentials are large (Hemmer-Hansen et al., 2007, 2014). This appears to apply strongly to species that are characterized by large populations, meaning that they are generally little affected by random genetic drift, and probably respond to even relatively weak selection, as locally beneficial alleles have a good chance of sweeping through the population.

Our population size estimates of N_e indicated a general decline in the population size of hoki, from a historical N_e of 228,681, 999 generations ago, to a N_e of 42,816, approximately 189 generations ago. These estimates have to be taken with caution, however, as estimation of N_e values is always plagued by high uncertainty owing to the need to sample a high number of individuals to produce accurate estimates (Blower et al., 2019). In fact, simulations indicate that around 1% of the total number of individuals might have to be sampled to ensure sufficiently precise estimates of $N_{e,}$ and our sample size is well below this recommend threshold. Indeed for species with large populations, this would mean that several thousands to millions of individuals would have to be sampled (Marandel et al., 2019). Promising recent developments that rely on the reconstruction of close relatives, such as the close kin mark–recapture (CKMR) approach (Bravington et al., 2016), have demonstrated an improved ability to calculate N_e with smaller confidence intervals (Ruzzante et al., 2019). However, application of this method to species with large population sizes has been little explored to date and is likely to be cost prohibitive in most cases.

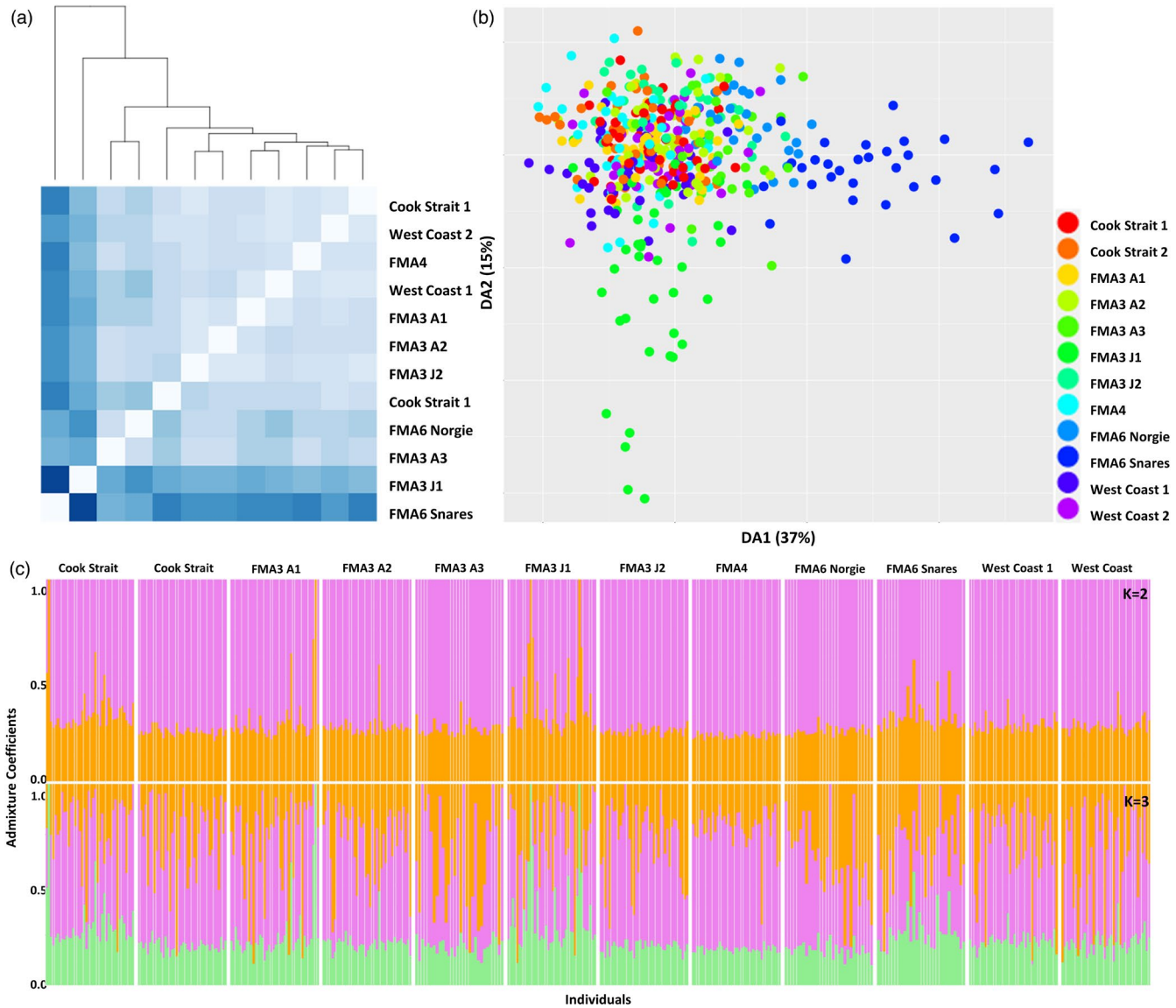

FIGURE 3 Population genomic analysis outputs for the hoki dataset consisting of neutral SNPs and New Zealand sampling sites only: (a) pairwise F_{ST} heatmap with hierarchical clustering dendrogram—darker blues indicate higher pairwise F_{ST} values and lighter blues indicate lower pairwise F_{ST} values, (b) DAPC scatterplot of DA1 (37%) and DA2 (15%) with points coloured by sampling site, and (c) LEA ancestral admixture plots for $K = 2$ and $K = 3$

In New Zealand, hoki forms the largest fishery by volume of catch (Papa et al., 2020). Our population genomics findings about hoki stock structure should be interpreted alongside other biological data on the species. Data on spawning patterns, morphometrics and growth rates in part corroborate the currently used two-stock model of hoki (Hicks et al., 2003; Livingston & Schofield, 1996; Livingston & Sullivan, 2007). However, our findings suggest that the Eastern and Western spawning grounds may undergo sufficient mixing to prevent genetic divergence to the degree where genetic clustering emerges. Furthermore, the finding of overall genetic panmixia and the low rate of pairwise location differentiation together indicate that growth differences could be due to phenotypic plasticity rather than genetically derived stock differences. Despite genetic panmixia in New Zealand, we detected some evidence for selection on some genomic regions, indicating that the overall hoki stock may hold some locally adapted genomic regions that may convey a fitness advantage. In the light of the prediction that changing climatic conditions can negatively affect marine productivity (Lindegren et al., 2018), it will be important for fisheries management to monitor the adaptive potential of this fishery. Future steps may focus on management approaches that seek to maintain locally adaptive variants to avoid depletion of biodiversity that could potentially lead to population decline (Reiss et al., 2009). The demonstration of sustainable management practices is particularly relevant for hoki, because since 2001 it has received Marine Stewardship Council (MSC) Fisheries Certification, which is based on criteria and audit processes that are internationally recognized as the world's highest global scientific standards for Ecosystem Based Fisheries Management (Fisheries,

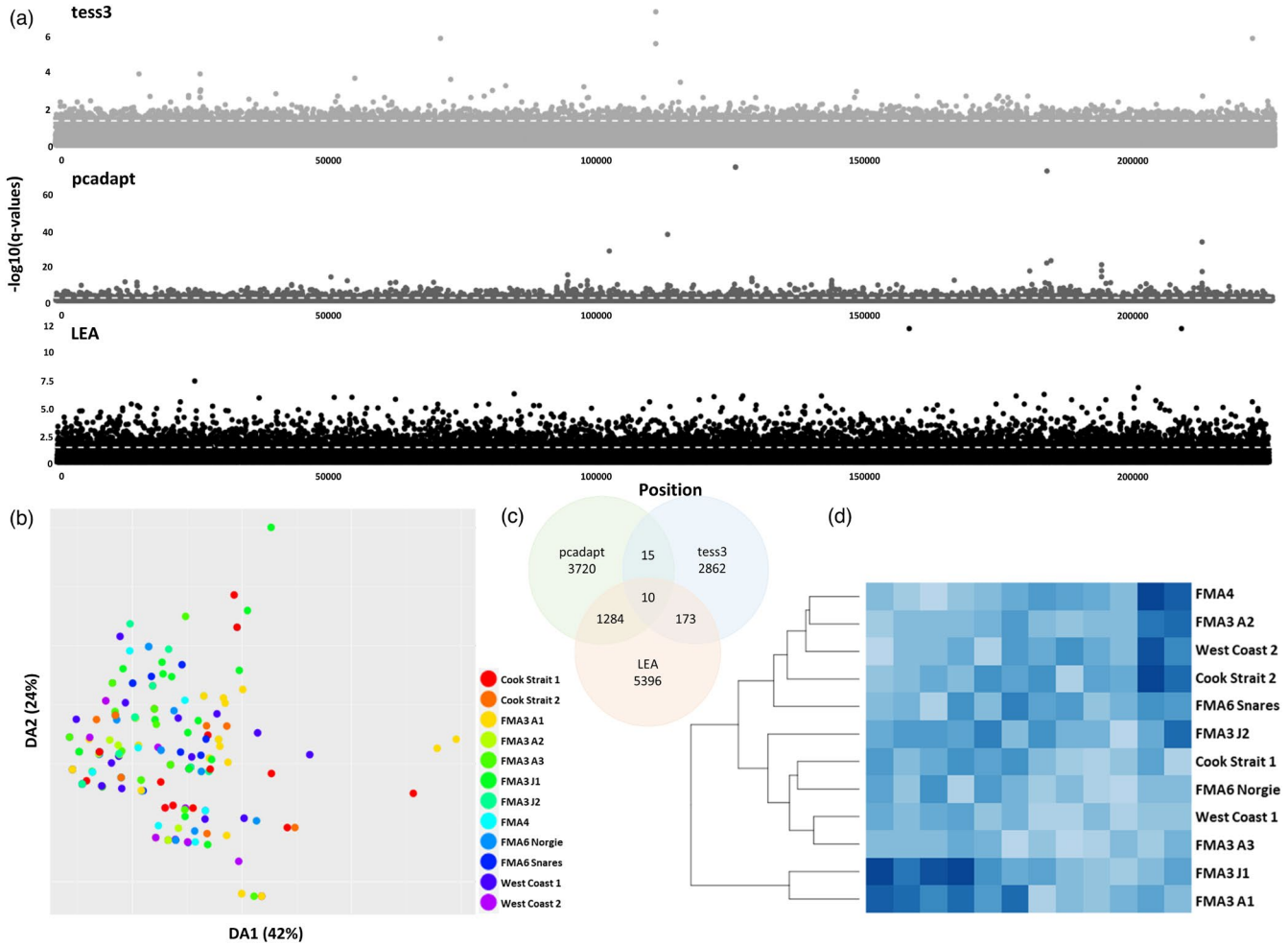

FIGURE 4 Panel (a) shows the results from Tess3, pcadapt and LEA *q*-value Manhattan plots of SNPs under selection for the hoki dataset consisting of New Zealand sampling sites only. White dashed lines corresponds to the Q-value threshold of 0.05. (c) the accompanying Venn diagram of putative SNPs under selection for the Tess3, pcadapt and LEA analyses, (b) DAPC scatterplot of DA1 (42%) and DA2 (24%) with points coloured by sampling site for dataset consisting of New Zealand SNPs under selection, (d) pairwise F_{ST} heatmap with hierarchical clustering dendrogram–darker blues indicate higher pairwise F_{ST} values and lighter blues indicate lower pairwise F_{ST} values for dataset consisting of New Zealand SNPs under selection

2019). More detail of temporal and spatial partitioning of the neutral and adaptive variation in this species is needed. Specifically, a genome assembly scaffolded to chromosome scale would enable linkage between variants to be better accounted for, which would facilitate the application of Gene-Environment Association (GEA) analyses. Nevertheless, the genome assembly presented here is highly contiguous and complete based on the size of the scaffolds obtained, the number of BUSCOs retrieved and the percentage of reads mapping back to the reference. This indicates the SNP set used here is representative of the genome-wide variation within the species. Additionally, extending the genomic analyses to include structural genomic variation, such as inversions, fusions and copy number variants (CNVs), appears an important next step, as this would allow more holistic capture of the full extent of segregating genomic variation in this species. This will be needed to improve both demography analyses and analyses to identify locally selected variants (Wellenreuther et al., 2019). Knowledge of geographically selected and divergent variants could be further used for the development of a genetic tool applicable to monitoring populations in time and space (Dahle et al., 2018; Hemmer-Hansen et al., 2019). With the data from this study, it is already possible to use identified divergent SNPs between the New Zealand and Australian clusters to inform hoki seafood traceability, as these could be used to link the catch to its geographic origin (A list of these SNPs is available upon request).

In conclusion, genetic variation at the genome-wide level is invaluable to identify fish stock structure in fisheries management, and the recent increase in the accessibility and resolution of population genetic data has facilitated the detection of previously unidentified structures, as well as signatures for natural selection in wild populations (Bernatchez, 2016). We highlight the importance of using large numbers of markers distributed across the genome to fully characterize the genetic diversity of marine species. In our current study, this allowed us to separate the Australian and New

Zealand clusters, as well as some more subtle differences within New Zealand hoki that otherwise could have been overlooked, and allowed us to scan the genome to detect regions under selection. This study contributes to increasing the genetic knowledge of this important fisheries species, and results can be used to improve our understanding of population dynamics and stock structure. Insights into the species demography is particularly challenging in high gene flow environments, such as many marine fishes, where small genetic differences across most of the genome can mask genetic divergence of strong functional significance. Thus, our study also serves as an example of the increased power offered by population genomics for conservation and management (Allendorf et al., 2010a; Hunter et al., 2018).

ACKNOWLEDGEMENTS

We like to thank all the vessel owners, managers and crew as well as Ministry for Primary Industries Fisheries Observers who helped to collect hoki samples. Funding was provided by a Seafood Innovations Limited grant to MW, with co-funding received from the Deepwater Group. The authors thank Ms Te Taiawatea Moko-Mead and Mr Kim Drummond (Te Ohu Kaimoana) for facilitating the engagement with Māori guardians/kaitiaki. The authors would also like to thank Sheridan Waitai (Ngāti Kuri Trustee) and Doug Jones (CE Ngai Tamanuhiri) for their cultural guidance with this project. The genomic data are stored in a managed repository to protect *Māori* rights and quota holders. We would also like to thank Chris Kirk for assisting with the DNA extraction, Pieter van As and Jeremy Walker from Neogen for providing sequencing advice and Rebecca Kirk and Mark Jarvis for help with contracting.

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

Permission from representatives of the Indigenous Peoples (Māori) was obtained for specimens used in this study. Further studies using this material, raw sequencing data and final genome assembly will require consent from the Māori iwi (tribe) who exercises guardianship for this material according to Aotearoa New Zealand's Treaty of Waitangi and the international Nagoya protocol on the rights of indigenous peoples. Raw and analysed data are available through the Genomics Aotearoa data repository at https://www.genomics-aotearoa.org.nz/data. Access to these data will require permission from Te Ohu Kaimoana (ika@teohu.maori.nz).

ORCID

Emily Koot https://orcid.org/0000-0002-1956-3415

REFERENCES

Ahrens, C. W., Rymer, P. D., Stow, A., Bragg, J., Dillon, S., Umbers, K. D. L., & Dudaniec, R. Y. (2018). The search for loci under selection: Trends, biases and progress. *Molecular Ecology*, *27*(6), 1342–1356. https://doi.org/10.1111/mec.14549

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010a). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, *11*(10), 697–709. https://doi.org/10.1038/nrg2844

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010b). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, *11*, 697–709. https://doi.org/10.1038/nrg2844

Barbato, M., Orozco-terWengel, P., Tapio, M., & Bruford, M. W. (2015). SNeP: A tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in Genetics*, *6*, 109. https://doi.org/10.3389/fgene.2015.00109

Benestan, L. (2019). Population genomics applied to fishery management and conservation. In M. F. Oleksiak & O. P. Rajora (Eds.), *Population genomics: Marine organisms* (pp. 399–421). Springer.

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

Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M. I., Beacham, T. D., Maes, G. E., Martinsohn, J. T., Miller, K. M., Naish, K. A., Ovenden, J. R., Primmer, C. R., Young Suk, H. O., Therkildsen, N. O., & Withler, R. E. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends in Ecology & Evolution*, *9*, 665–680. https://doi.org/10.1016/j.tree.2017.06.010

Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: Why genome scans may fail to map local adaptation genes. *Molecular Ecology*, *20*(10), 2044–2072. https://doi.org/10.1111/j.1365-294X.2011.05080.x

Blower, D. C., Riginos, C., & Ovenden, J. R. (2019). neogen: A tool to predict genetic effective population size (Ne) for species with generational overlap and to assist empirical Ne study design. *Molecular Ecology Resources*, *19*(1), 260–271.

Booker, T. R., Yeaman, S., & Whitlock, M. C. (2020). Variation in recombination rate affects detection of outliers in genome scans under neutrality. *Molecular Ecology*, *29*(22), 4274–4279. https://doi.org/10.1111/mec.15501

Bravington, M. V., Skaug, H. J., & Anderson, E. C. (2016). Close-kin mark-recapture. *Statistical Science*, *31*(2), 259–274. https://doi.org/10.1214/16-STS552

Catanach, A., Deng, C., Charles, D., Bernatchez, L., & Wellenreuther, M. (2019). The genomic pool of standing structural variation outnumbers single nucleotide polymorphism by more than three-fold in the marine teleost *Chrysophrys auratus*. *Molecular Ecology*, *28*(6), 1210–1223.

Caye, K., Deist, T. M., Martins, H., Michel, O., & François, O. (2016). TESS3: Fast inference of spatial population structure and genome scans for selection. *Molecular Ecology Resources*, *16*(2), 540–548. https://doi.org/10.1111/1755-0998.12471

Dabney, A., Storey, J. D., & Warnes, G. (2010). *qvalue: Q-value estimation for false discovery rate control*. R package version 1(0).

Dahle, G., Quintela, M., Johansen, T., Westgaard, J.-I., Besnier, F., Aglen, A., Jørstad, K. E., & Glover, K. A. (2018). Analysis of coastal cod (*Gadus morhua* L.) sampled on spawning sites reveals a genetic gradient throughout Norway's coastline. *BMC Genetics*, *19*(1), 1–17. https://doi.org/10.1186/s12863-018-0625-8

Dallaire, X., Normandeau, É., Mainguy, J., Tremblay, J.-É., Bernatchez, L., & Moore, J.-S. (2021). Genomic data support management of anadromous Arctic Char fisheries in Nunavik by highlighting neutral and putatively adaptive genetic variation. *Evolutionary Applications*, *14*, 1880–1897.

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., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: Re-implementation of software for the

estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, *14*(1), 209–214.

Duranton, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F., & Gagnaire, P.-A. (2018). The origin and remolding of genomic islands of differentiation in the European sea bass. *Nature Communications*, *9*(1), 2518. https://doi.org/10.1038/s41467-018-04963-6

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, *29*(1), 51–63. https://doi.org/10.1016/j.tree.2013.09.008

Fisheries, N. Z. (2019). *Review of sustainability measures for hoki (HOK1) for 2019/20*. Fisheries New Zealand.

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925–929. https://doi.org/10.1111/2041-210X.12382

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, *27*(9), 489–496. https://doi.org/10.1016/j.tree.2012.05.012

Gagnaire, P. A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F., Arnaud-Haond, S., & Bierne, N. (2015). Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary Applications*, *8*(8), 769–786. https://doi.org/10.1111/eva.12288

Garrison, E., & Marth, G. (2012). *Haplotype-based variant detection from short-read sequencing*. arXiv:12073907 [q-bio].

Hemmer-Hansen, J., Hüssy, K., Baktoft, H., Huwer, B., Bekkevold, D., Haslob, H., Herrmann, J.-P., Hinrichsen, H.-H., Krumme, U., Mosegaard, H., Nielsen, E. E., Reusch, T. B. H., Storr-Paulsen, M., Velasco, A., Dewitz, B., Dierking, J., & Eero, M. (2019). Genetic analyses reveal complex dynamics within a marine fish management area. *Evolutionary Applications*, *12*(4), 830–844. https://doi.org/10.1111/eva.12760

Hemmer-Hansen, J., Nielsen, E. E., Frydenberg, J., & Loeschcke, V. (2007). Adaptive divergence in a high gene flow environment: Hsc70 variation in the European flounder (*Platichthys flesus* L.). *Heredity*, *99*(6), 592–600. https://doi.org/10.1038/sj.hdy.6801055

Hemmer-Hansen, J., Therkildsen, N. O., & Pujolar, J. M. (2014). Population genomics of marine fishes: Next-generation prospects and challenges. *The Biological Bulletin*, *227*(2), 117–132. https://doi.org/10.1086/BBLv227n2p117

Hicks, A. C., Smith, P. J., Horn, P. L., & Gilbert, D. (2003). Differences in otolith measurements and gill raker counts between the two major spawning stocks of hoki (*Macruronus novaezelandiae*) in New Zealand. *New Zealand Fisheries Assessment Report*, *7*, 23.

Hohenlohe, P. A., Phillips, P. C., & Cresko, W. A. (2010). Using population genomics to detect selection in natural populations: Key concepts and methodological considerations. *International Journal of Plant Sciences*, *171*(9), 1059–1071.

Horn, P. (2011). Natal fidelity: A literature review in relation to the management of the New Zealand hoki (*Macruronus novaezelandiae*) stocks. *New Zealand Fisheries Assessment Report*, *34*, 18.

Horn, P. L., & Sullivan, K. (1996). Validated aging methodology using otoliths, and growth parameters for hoki (*Macruronus novaezelandiae*) in New Zealand waters. *New Zealand Journal of Marine and Freshwater Research*, *30*(2), 161–174.

Hudson, M., Garrison, N. A., Sterling, R., Caron, N. R., Fox, K., Yracheta, J., Anderson, J., Wilcox, P., Arbour, L., & Brown, A. (2020). Rights, interests and expectations: Indigenous perspectives on unrestricted access to genomic data. *Nature Reviews Genetics*, *21*, 1–8.

Hunter, M. E., Hoban, S. M., Bruford, M. W., Segelbacher, G., & Bernatchez, L. (2018). Next-generation conservation genetics and biodiversity monitoring. *Evolutionary Applications*, *11*(7), 1029–1034. https://doi.org/10.1111/eva.12661

Jombart, T., & Ahmed, I. (2011). 2011 adegenet 1.3-1: New tools for the analysis of genome-wide SNA data. *Bioinformatics*, *27*, 3070–3071.

Kawecki, T., & Ebert, J. D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*(12), 1225–1241. https://doi.org/10.1111/j.1461-0248.2004.00684.x

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.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Subgroup, G. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lindegren, M., Van Deurs, M., MacKenzie, B. R., Worsoe Clausen, L., Christensen, A., & Rindorf, A. (2018). Productivity and recovery of forage fish under climate change and fishing: North Sea sandeel as a case study. *Fisheries Oceanography*, *27*(3), 212–221. https://doi.org/10.1111/fog.12246

Livingston, M., & Schofield, K. (1996). Stock discrimination of hoki (*Macruronus novaezelandiae*, Merlucciidae) in New Zealand waters using morphometrics. *New Zealand Journal of Marine and Freshwater Research*, *30*(2), 197–208.

Livingston, M., & Sullivan, K. (2007). Successes and challenges in the Hoki fishery of New Zealand. In T. R. McClanahan & J. C. Castilla (Eds.), *Fisheries Management*. (pp. 253–273). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470996072.ch12:253-273

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.

Marandel, F., Lorance, P., Berthelé, O., Trenkel, V. M., Waples, R. S., & Lamy, J.-B. (2019). Estimating effective population size of large marine populations, is it feasible? *Fish and Fisheries*, *20*(1), 189–198. https://doi.org/10.1111/faf.12338

Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends in Ecology & Evolution*, *35*(7), 561–572. https://doi.org/10.1016/j.tree.2020.03.002

Moritz, C. (1994). Defining evolutionary significant units for conservation. *Trends in Ecology & Evolution*, *9*, 373–375.

Nielsen, E. E., Cariani, A., Aoidh, E. M., Maes, G. E., Milano, I., Ogden, R., Taylor, M., Hemmer-Hansen, J., Babbucci, M., Bargelloni, L., Bekkevold, D., Diopere, E., Grenfell, L., Helyar, S., Limborg, M. T., Martinsohn, J. T., McEwing, R., Panitz, F., Patarnello, T., … Carvalho, G. R. (2012). Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. *Nature Communications*, *3*, 851. https://doi.org/10.1038/ncomms1845

Nielsen, E. E., Hemmer-Hansen, J., Larsen, P. F., & Bekkevold, D. (2009). Population genomics of marine fishes: Identifying adaptive variation in space and time. *Molecular Ecology*, *18*(15), 3128–3150. https://doi.org/10.1111/j.1365-294X.2009.04272.x

O'Driscoll, R. L. (2004). Estimating uncertainty associated with acoustic surveys of spawning hoki (*Macruronus novaezelandiae*) in Cook Strait, New Zealand. *ICES Journal of Marine Science*, *61*(1), 84–97. https://doi.org/10.1016/j.icesjms.2003.09.003

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*, *48*, 1–31. https://doi.org/10.1080/03014223.2020.1788612

Pembleton, L. W., Cogan, N. O., & Forster, J. W. (2013). St AMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, *13*(5), 946–952.

Reiss, H., Hoarau, G., Dickey-Collas, M., & Wolff, W. J. (2009). Genetic population structure of marine fish: Mismatch between biological and fisheries management units. *Fish and Fisheries*, *10*(4), 361–395. https://doi.org/10.1111/j.1467-2979.2008.00324.x

Ruzzante, D. E., McCracken, G. R., Førland, B., MacMillan, J., Notte, D., Buhariwalla, C., Mills Flemming, J., & Skaug, H. (2019). Validation of close-kin mark–recapture (CKMR) methods for estimating population abundance. *Methods in Ecology and Evolution*, *10*(9), 1445–1453. https://doi.org/10.1111/2041-210X.13243

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Smith, P., McVeagh, S., & Ede, A. (1996). Genetically isolated stocks of orange roughy (*Hoplostethus atlanticus*), but not of hoki (*Macruronus novaezelandiae*), in the Tasman Sea and southwest Pacific Ocean around New Zealand. *Marine Biology*, *125*(4), 783–793. https://doi.org/10.1007/BF00349261

Waples, R. S. (1998). Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, *89*(5), 438–450. https://doi.org/10.1093/jhered/89.5.438

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370.

Wellenreuther, M., & Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends in Ecology & Evolution*, *33*(6), 427–440. https://doi.org/10.1016/j.tree.2018.04.002

Wellenreuther, M., Mérot, C., Berdan, E., & Bernatchez, L. (2019). Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification. *Molecular Ecology*, *28*(6), 1203–1209. https://doi.org/10.1111/mec.15066

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Koot, E., Wu, C., Ruza, I., Hilario, E., Storey, R., Wells, R., Chagné, D., & Wellenreuther, M. (2021). Genome-wide analysis reveals the genetic stock structure of hoki (*Macruronus novaezelandiae*). *Evolutionary Applications*, 14, 2848–2863. https://doi.org/10.1111/eva.13317