G3, 2022, jkac015
https://doi.org/10.1093/g3journal/jkac015
Advance Access Publication Date: 20 January 2022
Genomic Prediction

Genomic prediction of growth in a commercially, recreationally, and culturally important marine resource, the Australian snapper (*Chrysophrys auratus*)

Jonathan Sandoval-Castillo [1], Luciano B. Beheregaray,[1] Maren Wellenreuther [2,3,*]

[1]Molecular Ecology Laboratory, College of Science and Engineering, Flinders University, Bedford Park, SA 5042, Australia,
[2]School of Biological Sciences, The New Zealand Institute for Plant and Food Research Limited, Nelson 7010, New Zealand,
[3]Seafood Production Group, The School of Biological Sciences, University of Auckland, Auckland 1010, New Zealand

*Corresponding author: Molecular Ecology Laboratory, College of Science and Engineering, Flinders University, Bedford Park, SA 5042, Australia.
Email: maren.wellenreuther@plantandfood.co.nz

Abstract

Growth is one of the most important traits of an organism. For exploited species, this trait has ecological and evolutionary consequences as well as economical and conservation significance. Rapid changes in growth rate associated with anthropogenic stressors have been reported for several marine fishes, but little is known about the genetic basis of growth traits in teleosts. We used reduced genome representation data and genome-wide association approaches to identify growth-related genetic variation in the commercially, recreationally, and culturally important Australian snapper (*Chrysophrys auratus*, Sparidae). Based on 17,490 high-quality single-nucleotide polymorphisms and 363 individuals representing extreme growth phenotypes from 15,000 fish of the same age and reared under identical conditions in a sea pen, we identified 100 unique candidates that were annotated to 51 proteins. We documented a complex polygenic nature of growth in the species that included several loci with small effects and a few loci with larger effects. Overall heritability was high (75.7%), reflected in the high accuracy of the genomic prediction for the phenotype (small vs large). Although the single-nucleotide polymorphisms were distributed across the genome, most candidates (60%) clustered on chromosome 16, which also explains the largest proportion of heritability (16.4%). This study demonstrates that reduced genome representation single-nucleotide polymorphisms and the right bioinformatic tools provide a cost-efficient approach to identify growth-related loci and to describe genomic architectures of complex quantitative traits. Our results help to inform captive aquaculture breeding programs and are of relevance to monitor growth-related evolutionary shifts in wild populations in response to anthropogenic pressures.

Keywords: genome-wide association studies; fisheries genomics; aquaculture; teleost; ddRAD; reduced genome representation; ecological genomics; genomic prediction; GenPred; shared data resource

Introduction

Body size is considered one of the most important organismal traits because it influences several biological characteristics, from survivorship to fecundity (Peters 1986). It has significant ecological and evolutionary consequences, not just for the individual but for its population and community (Lorenzen 2016; Dijoux and Boukal 2021). Moreover, the somatic growth patterns of exploited fishes have a considerable role in the sustainability and economic viability of their fisheries stocks and aquaculture programs (Gjedrem et al. 2012; Murua et al. 2017; Barneche et al. 2018). These characteristics make growth and related traits major targets for selective breeding and human-induced evolutionary studies (Sharpe and Hendry 2009; Gjedrem et al. 2012; Denechaud et al. 2020; Huang et al. 2021). Teleost fish typically show indeterminate growth, but both the growth rate and growth-related plasticity are determined by the genetic makeup and the prevalent environmental conditions (De-Santis and Jerry 2007; Boltaña et al. 2017; Tung and Levin 2020). However, despite the general importance of growth, the genetic bases of growth traits in marine species are still not well understood.

Anthropogenic activities are rapidly transforming environments, resulting in changes to the selection landscape that organisms to which are exposed, and consequently, phenotypic changes in response to altered selection pressures (Hendry et al. 2017; Laufkötter et al. 2020; Silvy et al. 2020). Fish growth rates can respond rapidly to external selective pressures by phenotypic plasticity or by changes in genotypic frequencies in the population (Bowles et al. 2020; Denechaud et al. 2020; Oke et al. 2020; Pinsky et al. 2021). Fish growth is also a quantitative trait that is easy to measure, and as such, can be used to assess physiological and demographic responses to human-mediated environmental changes (Denechaud et al. 2020). In addition, pressures from capture fisheries have produced one of the fastest rates of phenotypic change observed in wild populations (Oke et al. 2020), with

Received: November 05, 2021. **Accepted:** January 07, 2022

effects not only to targeted species but also to associated communities (Andersen et al. 2016; Dijoux and Boukal 2021). However, size-selective harvesting and climatic effects could be species and system specific (Denechaud et al. 2020), and despite increasing concerns, the evolutionary consequences have so far been investigated in only a limited number of species. Thus, the extent to which changes to growth rate in response to human activities are evolutionary or plastic, and the degree to which are they reversible, remain debated (e.g. Pinsky et al. 2021).

The idea of rapid evolutionary responses due to fisheries pressures and aquaculture practices is supported by both theoretical and empirical evidence. In theory, high rates of harvesting of a population will favor earlier sexual maturity and slower growth (Uusi-Heikkilä et al. 2015; Monk et al. 2021). It is also expected that selective breeding programs increase the frequency of genetic variants underlying improved growth through the successive selective breeding of only the fastest growing individuals in each cohort. Experiments in the laboratory, where either large or small individuals are removed for several generations, have shown drastic changes in allele frequencies, loss of genetic diversity, and increased linkage disequilibrium in specific parts of the genome (Le Rouzic et al. 2020; Valenza-Troubat et al. 2021a). Some genes and genomic regions associated with growth have been identified in a few marine fish species (Wang et al. 2015; Robledo et al. 2016; Wu et al. 2019; Zhou et al. 2019; Yang et al. 2020; Valenza-Troubat et al. 2021c). However, because of the apparently polygenic nature and high phenotypic plasticity associated with the trait, the genetic basis of growth is not well understood, and its study is a persistent challenge (Wellenreuther and Hansson 2016; Gong et al. 2021; Valenza-Troubat et al. 2021a). The combination of genotyping based on reduced genome representation data sets and genome-wide association studies (GWAS) potentially provides a cost-effective and convenient approach for accurate localization and identification of growth-related genomic regions in fishes.

The Australian snapper (*Chrysophrys auratus*) is a large (~1 m) and long-lived (>50 years) marine teleost (Fig. 1) that supports important recreational and commercial fisheries in Australia, and New Zealand (Parsons et al. 2014; Cartwright et al. 2021). The species is also of major cultural importance, having been caught by New Zealand's indigenous Māori people for hundreds of years, and is considered taonga (a treasure) by them. The species stocks have large biomass, but since the early 2000s most have shown declines, with some losing over 75% of their biomass (Cartwright et al. 2021). These reductions in snapper productivity have negative economic and ecological consequences that could be amplified by global environmental change (Ogier et al. 2020; Parsons et al. 2020). A strategy for mitigating some of these negative effects, while enhancing wild snapper stock productivity, is the development of captive production. Although there is not yet commercial aquaculture of the species, research programs intending to develop Australian snapper into a profitable and sustainable aquaculture species exist in New Zealand and Australia (Catanach et al. 2019; Ashton et al. 2019a). The goal of our work is to contribute to both the long-term New Zealand efforts to develop this aquaculture activity, and the understanding of the genetic bases of growth-related traits in marine fish more generally. We took advantage of several resources available from the New Zealand snapper aquaculture program, including a chromosome-scale assembly (Catanach et al. 2019) and fish from the New Zealand snapper breeding program representing extreme growth phenotypes following a grow-out phase in a sea pen. Using these resources, we then applied reduced genome representation to conduct GWAS analyses to identify growth-related genetic variation in a cost-efficient way. In addition, we implemented a Bayesian mixture model to describe the genetic architecture of growth and to test the accuracy of size phenotype prediction based on single-nucleotide polymorphism (SNP) genotypes. This information will help to improve the selection of high-quality broodstock and the monitoring of human-induced evolution in wild stocks.

Fig. 1. Juveniles of Australian snapper (*Chrysophrys auratus*) reared in captivity at The New Zealand Institute for Plant & Food Research Limited finfish facility.

Methods

Fish rearing and grow-out conditions in the sea pen

In 2016, a snapper breeding program was started in Nelson (New Zealand) with wild-caught animals at The New Zealand Institute for Plant and Food Research Limited (PFR). Uncontrolled mass spawning of a wild-caught broodstock at the PFR finfish facility at the Nelson Research Centre produced 2 batches of eggs that hatched on 2015 January 31 and February 7, respectively. Mass spawning is the typical method for most bream species, where group interactions are often necessary to induce spawning behavior, with all individuals being able to mate freely with other individuals in the broodstock tank population. These F_1 snapper probably contained a diverse mix of related individuals including full-siblings, half-siblings, and unrelated individuals. Approximately 80,000 larvae were reared in 5,000-l tanks held at 21°C and supplied with a filtered, UV-sterilized supply of incoming water. First standard grading of animals occurred on 2015 May 1, whereupon approximately 54,000 fish were retained for on-growing. This reduction in fish numbers was also caused by natural mortality events common in larval rearing, particularly during the first few weeks, due to the lack of a fully functional immune system (Bricknell and Dalmo 2005), digestive system (Holt 2011), osmotic system, and compromised host microbe interaction (Vadstein et al. 2013) in young fish larvae. Routine grading was repeated every 2–6 weeks thereafter to ensure that fish have sufficient space and to maintain animal welfare. Smaller animals were pooled separately from larger individuals but were retained at the same ration and water temperatures. Overstocking was prevented by grading fish into additional 5,000-l tanks as required. The water supply of the tanks was retained at temperatures above environmental (ambient) temperature, set to a minimum of 16°C. When fish were around 180 days old, they were graded following standard husbandry procedures to remove fish

with deformities and lesions. Following the grading, approximately 17,000 individuals of mixed sizes were moved to a 50,000-l, another 10,000 large individuals were retained in two 5,000-l tanks, and another 12,000 small individuals were separated into 2 additional 5,000-l tanks, totaling to 39,000 fish that were kept for further on growing in the facility. These fish movements into new tank space are routine procedures in fish farming aimed to provide fish with the necessary space for optimal grow out, where the separation of different size classes aims to avoid intercohort cannibalism and food competition. Following the movement into new tanks, all fish were retained on equivalent rations and at temperatures above ambient. Stocking densities were held below 25 kg/cm^3 in all instances. Final grading of the cohort to again removed fish with deformities and lesions was performed in the week of 23 November 2015. Fish suitable for transport to the sea pen (i.e. fish in good condition, uninjured, and of normal conformation) were graded to a minimum of 100-mm fork length and hand counted, yielding 21,891 individuals. Transport to PFR's sea pen in Beatrix Bay (Pelorus, Marlborough) took place on 2015 December 16, when fish were either 319 or 312 days old.

On 2017 April 26–27, the snapper sea pen trial was ended following 17.5 months of on-growing. The mean size of snapper at harvest was 238-mm fork length, where sizes ranged between 139- and 301-mm fork lengths (Cook and Black 2017). Of these snappers, 200 of the smallest and 200 of the largest were selected by eye for this study, roughly falling within the 5% extremes of the size distribution (representing 5% of the smallest and 5% of the largest fish). We have deliberately opted for such a sampling design of extreme phenotypes because it improves the power of QTL studies without incrementing their false discovery rate. This also reduces the cost of genome scale variant calling, which could be prohibitively expensive (e.g. Li et al. 2019). This is indeed a standard approach in QTL study design, and is akin to the rationale used for the selection of recombinant inbreed lines for QTL mapping studies (for an exhaustive review of study designs for fish QTL studies in the genomics era see: Ashton et al. 2017). In this study, the extreme growth phenotypes were selected during harvesting of a sea pen snapper cohort that were on-grown in this environment for well over a year. This deliberately capitalizes on the power that comes with the application of an extreme study design, as applied in several investigations seeking to identify the genomic basis of phenotypic traits in plants (e.g. Yang et al. 2015) and fish (Houston et al. 2008; Wang et al. 2008; al. 2015).

DNA extraction and library preparation

Total genomic DNA was extracted from fin tissue samples collected from the 5% biggest (L) and the 5% smallest (S) individuals (200 for each size group), following the protocol described by Ashton et al. (2019). The concentration, purity, and integrity of the extractions were assessed using Qubit (Life Technologies), NanoDrop (Thermo Scientific), and 2% agarose electrophoresis gels, respectively. The 188 samples with the best quality for each size group, plus 8 replicates, were used for the preparation of 364 libraries using the reduced genome representation technique of ddRADseq (double-digest restriction site-associated DNA sequencing). We followed the protocol described by Peterson et al. (2012) with a few modifications as described by Sandoval-Castillo et al. (2018). For each sample, approximately 300 ng of DNA was digested with the restriction enzymes *SbfI* and *MseI*, and then ligated with forward and reverse adaptors, with forward adaptors including 1 of 96 individual barcodes designed in-house. Libraries were size selected for 250–800-bp fragments with a Pippin Prep (Sage Science), and then amplified using PCR. Individual libraries were mixed in equimolar concentration in 4 pools of 96 samples, and each pool was sequenced in an Illumina HiSeq 4000 lane (paired-end 150 bp) at Novogene Co.

Sequence quality checking and filtering

Sequence quality was checked using FASTQC 0.11 (Andrews 2011). Raw reads were then demultiplexed into individual samples using "process_radtags" from STACKS 1.5 (Catchen et al. 2013). Barcodes, rag-tags, bad quality base pairs (slide window 5, $Q < 25$), and adapters were trimmed using TRIMMOMATIC 3.9 (Bolger et al. 2014). Reads with more than 5% "Ns," average quality lower than 20, and length shorter than 70 bp were also removed. SNPs were called using the GATK pipeline, following recommendations by Van der Auwera et al. (2013). We used BOWTIE 2.3 (Langmead and Salzberg 2012) to align filtered reads to the chromosome-level Australian snapper genome (Ashton et al. 2019b). We improved the alignments by locally realigning indels, then called SNPs using BCFtools 1.19 (Danecek et al. 2021). To reduce the number of false variants produced as inherent artifacts of the ddRAD sequencing approach (e.g. paralogous, sequencing errors, allele drop), several filtering steps were applied to the raw, called SNPs (Supplementary Table 1) using VCFTOOLS 0.1.16 (Danecek et al. 2011). First, we removed SNPs called in <80% of the samples, with a minimum allele frequency <3%, or with allele balance <20% or >80% on heterozygote genotypes. After removing indels, and to avoid sequencing or misalignment errors, we selected SNPs with ≥20% depth/quality ratio and ≥30 mapping quality (Li 2014). Additionally, to reduce paralogous loci and allelic dropout, we eliminated SNPs with too high or too low coverage (>mean depth plus twice the SD, or ≤10× average depth per individual). Loci that did not conform to Hardy Weinberg Equilibrium in samples of both size groups, missing in more than 5% of samples of either of the size groups, or that showed discrepancy between 2 or more replicates, were also removed. For some analyses, we created an unlinked data set by pruning SNPs using a window size of 100 kb, a variance inflation factor of 5 and an R^2 of 0.5 in Plink 1.9 (Chang et al. 2015).

As recommended for GWAS analyses, we removed individual that deviate ±3SD from the average heterozygosity (Marees et al. 2018). Because inbreeding and relatedness can affect GWAS analyses, we calculated the individual inbreeding coefficient (F_{IS}) in Plink 1.9 and pairwise relatedness in VCFTOOLS using the unlinked SNP data set (11,588). We use the algorithm implemented in Plink that is based on the relative observed and expected homozygosity under Hardy–Weinberg equilibrium. We used the formula $F = (O\#hom - E\#hom)/(1 - E\#hom)$. Where O#hom and E#hom are the observed and expected number of homozygous genotypes in the sample, respectively. We then compared the average F_{IS} and relatedness of each size group using a nonparametric Mann–Whitney *U* test. Age, sex, and population structure are factors that can potentially influence GWAS analyses, but given all fish used in this study were from the same population and age class and were immature at the time of the terminal sampling, these factors were considered not to bias the GWAS analyses.

With the aim of identifying SNPs involved in growth-related traits, we compared the allele frequencies between size groups using 3 different methods, each of which implement a different approach to control for relatedness and possible inbreeding. Firstly, based on identical-by-state pairwise genetic distances, we used complete linkage agglomerative clustering to create multidimensional scaling plots (MDS). Then, to control for family group substructure, we used MDS components as covariables in a logistic regression implemented in Plink and calculated

asymptotic P-values to test for significance. This single-locus approach tests for odds of the minor allele frequency of each SNP being associated with either phenotype (in this case large or small fish), when adjusting for potential confounders. Secondly, a pairwise kinship matrix was used as a covariable in a linear polygenic mixed model. The significance of the model was calculated using the FAmily-based Score Test for Association, all implemented in the R package GenABEL (Aulchenko et al. 2007). Thirdly, the same pairwise kinship matrix was used as a covariable in a mixed linear model, but the significance was tested using a combination of a sequence kernel association test and a likelihood-ratio (LR) test, comparing a model with association against a model of nonassociation; these approaches were implemented using the R package rainbowR (Hamazaki and Iwata 2020). Like GenABEL's model, rainbowR tests for a significant contribution of a polygenic component in the total genetic variance between size groups, when considering a random component and multiple covariance component effects. The main difference is implementation of LR in rainbowR, which is expected to better control for false positives (Hamazaki and Iwata 2020).

These analyses were based on single-SNP effects, but considering that gene function could be more commonly influenced by the small effects of several SNPs combined than the large effect of 1 SNP, haplotype-based approaches can improve the detection power of causal variance (Hamazaki and Iwata 2020; Sinclair-Waters et al. 2020). We therefore implemented 2 different methods to calculate haplotype blocks in rainbowR. One method considers pairs of variants within 200 kb of each other that show high coefficients of linkage disequilibrium ($D_{\text{lowerCI90}} > 0.7$ and $D_{\text{higherCI90}} > 0.98$), as implemented in Plink (Chang et al. 2015). The other method uses a sliding window of 15 to cluster the closest 15 SNPs, implemented directly in rainbowR (Hamazaki and Iwata 2020). For each GWAS analysis, we calculated a significant P-values threshold following Brzyski et al. (2017). For this, we clustered the number of tests (SNPs or Haplotypes) by performing a LD clumping with a threshold $r^2 > 0.1$ at maximum distance of 1 Mb. Then, the best P-values per cluster was selected and the maximum P-value significant at 5% false discovery rate was used as threshold. In addition to investigating the genetic architecture of growth in Australian snapper, we examined the proportion of size variance explained by each SNP and chromosome using the package bayesR (Moser et al. 2015) with default priors. This software estimates the genetic variance explained by each SNP and describes the genetic architecture of complex traits using Markov chain Monte Carlo and Bayesian mixture models. We calculated SNP-based heritability (h^2_{SNP}) as the proportion of phenotypic variance explained by all SNPs and considered this an approximation of the genome narrow-sense heritability (h^2_g). Accuracy of genomic prediction was evaluated using 10 runs of bayesR, in each of which 80% of samples were randomly selected as training data and the remaining 20% was used as validation data. Since we used size as a categorical variable, the accuracy was measured in the validation sets as the beta coefficient of a logistic regression (predicted phenotype as a function of real phenotype) divided by the square root of heritability (h^2_{SNP}).

To explore the function of the candidate SNPs, we clustered SNPs in sliding windows of 10 kb and extracted the sequence of each cluster from the reference genome. These sequences were then aligned to a Teleostei protein database obtained from UniProt (2021), using blastx 2.2.28 (Camacho et al. 2009) with a minimum alignment length of 30 amino acids, a similarity score of 50%, and an e-value threshold of 1×10^{-06}. Enrichment of genetic functions on the putative candidate genes was explored using the gene ontology (GO) terms provided by UniProt and the R package topGO 2.26 (Alexa et al. 2016). Pathways and protein interaction analyses were performed using the web servers String 11 (Szklarczyk et al. 2021) and reactome 3.7 (Jassal et al. 2020).

Results

Approximately 3.6 billion raw reads were obtained from the 4 Illumina lanes. After demultiplexing, trimming, and filtering, we retained over 3.5 million pairs of reads per sample, from which >83% were successfully aligned to the snapper reference genome. From these, 1,273,017 variable sites were identified. After strict filtering, the data set was reduced to 17,490 high-quality SNPs, which were evenly distributed among the 24 chromosomes of the reference genome (Supplementary Table 2). After removing replicates and 12 samples (5 large and 8 small fish) owing to large amounts of missing data or extreme heterozygosity, 363 individuals (183 large and 180 small) were retained, with an average of 0.8% missing data.

The mean inbreeding coefficient was slightly higher in the small fish group (large $=-0.1225 \pm 0.1990$; small$=-0.0486 \pm 0.1805$, $P < 0.01$). However, all individuals in both groups had inbreeding coefficients lower than 0.15, with the average per group lower than 0. As such, we considered the extent of inbreeding to be low in both groups. Average relatedness was not significantly different within compared with between the 2 size groups (large $= 0.0089 \pm 0.1320$; small $= 0.0098 \pm 0.1170$; large vs small $= -0.0071 \pm 0.1353$; all $P > 0.1$).

The 3 single-SNP-based methods identified 31 candidate loci under selection between large and small fish. Of these 9, 11, and 12 were selected by Plink, GenABEL, and rainbowR, respectively (Fig. 2). None of these was identified by 2 or more tests; however, these 31 candidates were in 10 chromosomes and 1 scaffold, and most of them (18 loci; 58%) were in physical proximity to 1 region of chromosome 16 (Fig. 2). The haplotype-based methods identified 77 SNPs in 7 unique blocks as candidates, with 3 of these blocks identified by the 2 approaches (Fig. 3). The 7 blocks were localized on 5 chromosomes and 1 scaffold, but more (3 blocks; 43%) were found on chromosome 16 than on any other chromosome (Fig. 3). All these results suggest a key role of chromosome 16 on growth in the Australian snapper. This was concordant with the bayesR results, which found that the SNPs on chromosome 16 explained 16.4% of the size variance in our samples (Fig. 4 and Supplementary Table 2). Interestingly, this analysis suggested that a significant amount of growth variance in the Australian snapper can be explained by large effects of a relatively small number of SNPs (60 SNPs; Supplementary Table 3). However, to explain as much as 76% of the phenotypic variation, small effects of many loci need to be considered (2,102 SNPs; Supplementary Table 3). Based on the 10 repetitions of crossvalidation, the phenotype prediction accuracy was considerably high (74.4%), but expected given that the set of SNPs can explain most of the phenotypic variation in our data.

From both single-SNP and haplotype block approaches, 100 unique candidates were identified. We extracted 65 contigs of 10 kb that included all unique candidate SNPs. Of these contigs, 57 were annotated to 51 proteins, where 16 SNPs were localized in exon regions. The proteins were assigned to 153 GO terms, but none of the functions or pathways were significantly enriched ($P = 0.18$; results not shown). Despite this, most of these genes were

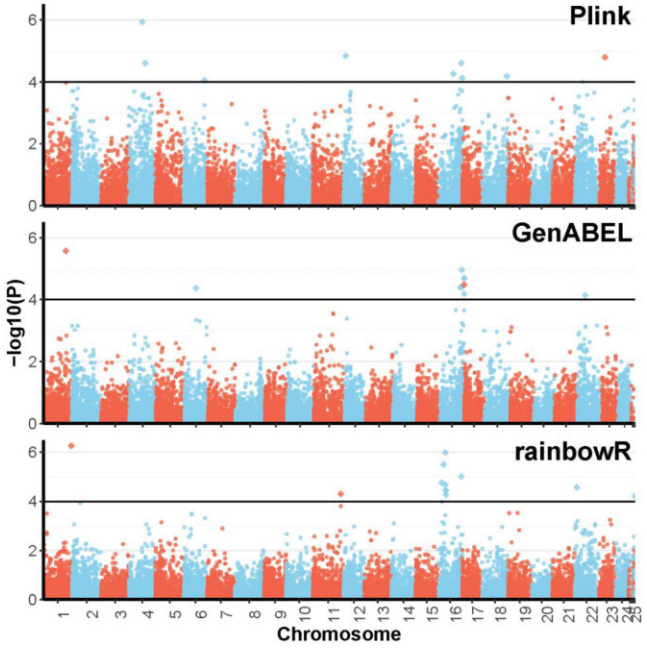

Fig. 2. Manhattan plot of results from the 3 single-SNP-based genome-wide association analyses of growth for 363 Australian snapper (*Chrysophrys auratus*) using 17,490 SNPs. SNPs are plotted according to their chromosomal position against their −log 10 (*P*-value). Significant SNPs are the diamonds over the horizontal line (−log 10(*P*)>4).

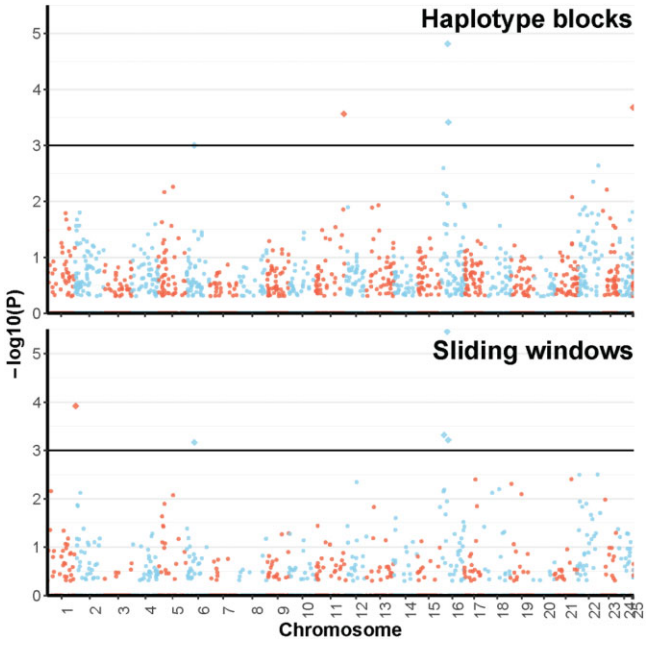

Fig. 3. Manhattan plot of results from the 2 haplotype-based genome-wide association analyses of growth on 363 Australian snapper (*Chrysophrys auratus*) using 17,490 SNPs in 1,337 blocks and 8,201 SNPs in 2,959 windows. Haplotypes are plotted according to their chromosomal position against their −log 10 (*P*-value). Significant haplotypes are the diamonds over the horizontal line (−log 10(*P*)>3).

involved in the metabolism of proteins, cell organization, and tissue development, including 5 directly involved in growth (Supplementary Table 4).

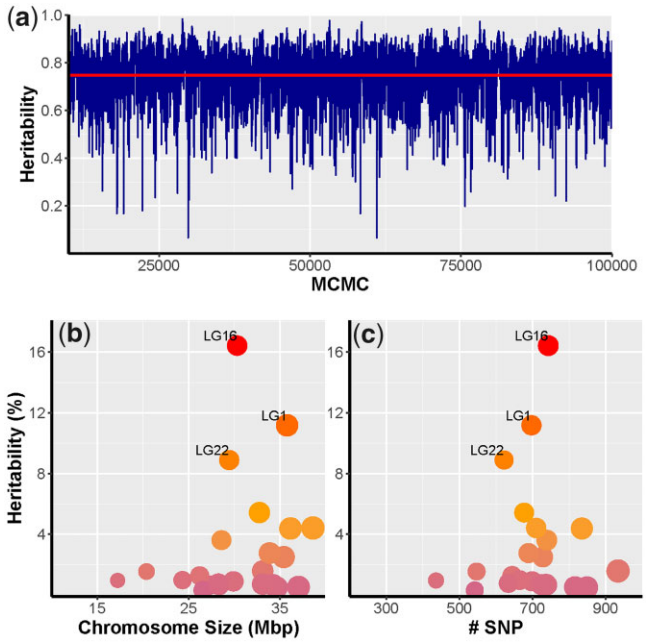

Fig. 4. SNP-based heritability (h^2_{SNP}) of growth in Australian snapper (*Chrysophrys auratus*) estimated from Bayesian mixture models using 17,490 SNPs in 24 chromosomes. a) Convergence of MCMC sampling with average heritability ($h^2_{SNP} = 0.757$) shown by the red horizontal line. Proportion of heritability or genetic variance of growth explained by each chromosome as a function of b) chromosome size in megabase pairs and c) number of SNPs per chromosome.

Discussion

In this study, we successfully identified 100 growth-related SNPs, of which 60 were located on chromosome 16. We also provide estimates of parameters related to the genomic architecture of growth, including heritability, and the number and effect sizes of SNPs contributing to size variation. These results demonstrate that the reduced genome representation SNPs, combined with appropriate bioinformatic tools, provide a cost-efficient approach to identify growth-related loci and to broadly describe complex trait genomic architectures. Our results provide crucial information to aquaculture breeding programs about the targets of selection to enhance growth in this and related species. Knowledge about the genomic basis of growth also improves our general understanding of the impacts of anthropogenic stressors, such as climate change and extractive fisheries, on growth trait evolution in wild fisheries stocks.

Genome architecture of a complex growth trait

Several studies have identified genes with large effects underlying simple Mendelian traits, but because of the statistical limitations of single locus GWAS approaches, they have been less successful with polygenetic quantitative traits like growth or size (Wellenreuther and Hansson 2016). Perhaps the most famous example of this is human height, where even with large samples sizes (~250,000 people), <20% of variation can be explained using genome-wide markers (Yengo et al. 2018; Sohail et al. 2019). Akin to many other nonhuman studies, our study relied on a relatively small sample size, which reduced the power to detect loci of medium or small effect. However, by transforming the quantitative trait to a qualitative trait (big vs small) by comparing extreme phenotypes representing the 5% extremes of the phenotypic variation, and by using both haplotype-based and Bayesian mixture

model approaches, we not only identified several growth-related loci, but also confirmed the polygenetic nature of the trait in the Australian snapper. These growth-related genes were detected in F_1 obtained from a wild broodstock thus indicates that these growth genes are segregating in wild snapper populations. Wild snapper populations in New Zealand show high genetic diversity (Papa et al. 2021), which is common for many marine teleost species with large population sizes (Nielsen et al. 2009; Koot et al. 2021). The snapper cohort used here was obtained following mass spawning of wild broodstock. While our on-growing of snapper resulted in natural mortality and only retained healthy individuals (e.g. grading removed deformities and fish with lesions) of the starting 80,000 strong cohort, our study population contained high genome-wide diversity (observed heterozygosity for the 363 graded snapper was 0.307), providing sufficient diversity for our GWAS. As such, our study provides a clear example of how switching the focus away from single significant loci, to using information from all genotyped SNPs, can improve the statistical power to detect polygenic components of important phenotypic variants, including some regions with small but significant effects. Our results suggest that a relatively small number of SNPs (<20K) is enough to explain most of the phenotypic variation related to growth. The overall heritability was generally higher than those reported in other studies (but see Valenza-Troubat et al. 2021a), however, the large heritability estimate is likely confounded upwards due to the well-known inflation of allele substitution effects and heritability estimates when using selective genotyping approaches (Falconer and Mackay 1996). Moreover, the 100 candidate SNPs explain similar phenotypic variation (4.4%) to those in other fish studies (0.1–19.0%; Wu et al. 2019; Gong et al. 2021), including a quantitative trait loci analysis of the same species (Ashton et al. 2019b). As expected in complex traits like growth, individual candidate effects were small and multiple candidates were required to explain most of the phenotypic variation (Wellenreuther and Hansson 2016; Sinclair-Waters et al. 2020). However, only 4 candidates in the last 100 kb of chromosome 16 could explain up to 3% of the total phenotypic variation. In addition, if we consider all SNPs in this chromosome, the explanatory power is elevated to over 16%. The variation explained by the markers on chromosome 16 is very high and could suggest physical linkage of genetic variation associated with growth. This is consistent with empirical examples of evolution of multiple linked variations that together modify the function of a gene or a complex of genes (Koshikawa et al. 2015; Kingman et al. 2021). This type of evolution promotes the formation of tightly linked haplotype blocks, allowing for the selection and inheritance of multiple sites with effect over several aspects of a complex trait, such as growth.

Haplotype-based evolution of polygenic regulation of growth could explain the lack of overlapping candidates from different analyses, such lack of concordance is expected when compare different GWAS algorithms (Manunza et al. 2014; Dong et al. 2016). The statistical approach of each method will capture loci with a particular degree of effect, and therefore only a fraction of the complex polygenic architecture. Such architecture would be more likely to include a mixture of loci with large, medium, small, pleiotropic, and synergic effects (Wellenreuther and Hansson 2016; Ashraf et al. 2020; Sinclair-Waters et al. 2020). Although the selected candidate SNPs differed between methods, all methods selected at least 1 candidate SNP within the last 5% of chromosome 16's length. This chromosome region is broadly consistent with putative growth-related quantitative trait loci peaks identified for 3 growth traits in 1-year-old Australian snapper reared in a land-based finfish facility (Fig. 5 in Ashton et al. 2019b), and several candidate genes on chromosome 16 were also detected in a study investigating domestication selection following selective breeding for improved growth in this species (Baesjou and Wellenreuther 2021). In the study by Ashton et al. (2019b), chromosome-level additive variance was not calculated; however, their results also suggested that chromosome 16 tended to explain most of the phenotypic variation. Here, neither chromosome length nor SNPs per chromosome was significantly correlated with the proportion of variation explained (Fig. 4). This suggests that the high relative contribution of chromosome 16 is not just an artifact of our reduced genome representation approach, but an important part of the genetic architecture of growth in Australian snapper. The results from this study being in part consistent with those by Ashton et al. (2019) also shows that the genomic basis of growth is comparable across different rearing environments (land-based finfish facilities vs a sea pen). Chromosome 6 showed the second largest number of candidates, but the SNPs in this chromosome explained <4% of the phenotypic variation. In contrast, chromosomes 1 and 22 each explain over 8% of the phenotypic variation, despite chromosome 1 having a similar number of candidates and chromosome 22 having only a few candidates compared with chromosome 6. This is additional evidence for the very complex genomic architecture of growth, which is composed by a mixture of several loci with small effects detected by the haplotype approaches on chromosome 6, and a few loci with larger effects detected by single-SNP approaches, especially on chromosome 22. Thus, it is important to combine conceptually different approaches to extract information from all markers, to gain a better understanding of the genetic components of phenotypic variation (de Maturana et al. 2014; Akond et al. 2021; Gong et al. 2021).

Candidate genes

The high-quality genome assembly made it possible to identify 51 genes surrounding the 100 SNP candidates (Supplementary Table 4). Although most of these SNPs were in introns, several were in the proximity of, or inside, exons of the relevant candidate gene. Nevertheless, we are cautious about inferred associations. Previous studies have suggested several candidate genes and pathways for teleost growth-related traits (De-Santis and Jerry 2007; Johnston et al. 2011; Baesjou and Wellenreuther 2021). Of these, the somatotropic axis is perhaps the most important, since it plays a central role in the regulation of metabolic and physiological processes involved in fish growth (De-Santis and Jerry 2007). Although key genes in this axis include the growth hormone (Irving et al. 2021) and insulin-like growth factors, several other factors, carriers, and receptors are implicated in this pathway. Several of our candidate SNPs were near the genes from the somatotropic axis, including the fibroblast growth factor 18 (FGF18) and growth factor 5 (GDF5). Both genes are essential regulators of cell growth, cell differentiation, morphogenesis, tissue growth, and tissue repair, especially during skeleton development (Jovelin et al. 2010; Nagayama et al. 2013), potentially affecting multiple traits. Concordantly, FGF18 has been associated with growth variation in large yellow croaker, *Larimichthys crocea* (Zhou et al. 2019). Two additional candidate genes detected, nucleophosmin 3 (NPM3) and transmembrane protein 132E (TMEM132E); both associated with the regulation of growth factors. NPM3 is predicted to be a histone chaperone protein modulating replication of DNA and gene expression during development (Wu et al. 2009). Because of proximity, it coactivates with the fibroblast growth factor 8 gene, and both regulate

development, cell growth, and proliferation in vertebrates (Kikuta et al. 2007). TMEM132E is a component of the cell membrane and is implicated in the regulation and transport of insulin-like growth factors. This gene is also involved in the posterior lateral line (Li et al. 2015) which, as a mechano-sensory organ, has physiologically demanding growth requirements (Sapède et al. 2002).

Biosynthesis and retention of proteins is fundamental for tissue development, and the efficiency of these processes determines growth rate (Fraser and Rogers 2007). Seven of our candidates are directly involved in metabolism of proteins (Supplementary Table 4), 6 of which are on chromosome 16. Of these, the prolyl 3-hydroxylase 1 (P3H1) and the serpin peptidase inhibitor clade H (SERPINH1) are notable in fishes for their implication in biosynthesis of collagen and assembly of collagen fibrils (Oecal et al. 2016; Tonelli et al. 2020). Collagen is essential for the structure and strength of bones, skin, muscle, and cartilage tissues, and therefore integrity during somatic growth (Fraser and Rogers 2007). Moreover, collagen metabolic processes have been associated with changes in growth rates and responses to temperature increments in the Australian snapper (Wellenreuther et al. 2019). Hence, variation in P3H1 and SERPINH1 could have a direct effect on the collagen metabolism efficiency and thus growth rate in this species. Other important processes for tissue growth include cell proliferation, differentiation, migration, and adhesion. From our candidates, 7 genes were associated with these cell processes (Supplementary Table 4), including laminin subunit alpha 3 (LAMA3), related extracellular matrix protein 2 (FREM2), and periostin (OSF2). These 3 genes have a regulatory function in cell migration and adhesion, and are involved in organ morphogenesis and tissue development of integuments, spinal cord, brain, eyes (Sztal et al. 2011), pharynx, fins (Carney et al. 2010), and the skeleton of fishes (Kessels et al. 2014). Finally, small difference in expression levels of specific genes have been shown to impact growth rate and metabolic efficiency (Fraser and Rogers 2007; Wellenreuther et al. 2019). Thus, variation in genes that act as transcript regularity factors can have important biological consequences. We detected several genes involved in gene regulation (see Supplementary Table 4). Faster growth can be achieved by reducing energetic cost of protein metabolism, increasing transcription regulation adeptness, and cell migration and adhesion efficiency, which together can make a larger proportion of energy and cell proliferation available for growth. We consider all the above-described genes as of potential importance in the growth patterns of not only Australian snapper but perhaps of other teleosts.

Applied relevance of genomic prediction of growth

Animal breeding has a long history and the underpinning science has been developed over decades, with significant work coming from the field of quantitative genetics (Falconer and Mackay 1996), as well as population genetics (Fisher 1918, 1930). The application of this to improve selective breeding in aquaculture has been revolutionized with the advent of high-throughput sequencing (Houston et al. 2020), and since then studies into the genomic basis of economically important traits have flourished (Valenza-Troubat et al. 2021a; Houston et al. 2020) and increasing efforts are now underway to disentangle how these genes are differentially expressed (Valenza-Troubat et al. 2021b). The most important contribution of our results related to the genomic basis of growth is the potential to inform individual breeding values in terms of growth efficiency. This information has direct applications to selective breeding programs of commercially important species, such as Australian snapper, but also for the breeding programs of the gilthead sea bream (*Sparus aurata*) in the Mediterranean Sea, and the red sea bream (*Pagrus major*) in Japan, which reside in the same family. The increment of growth rate and its energetic efficiency can directly reduce production time and cost, leading to higher economic returns (Ye et al. 2017; Valenza-Troubat et al. 2021a). However, the complexity and polygenetic nature of commonly involved traits has made it difficult to implement efficient genomic selective breeding programs (Goddard and Hayes 2009; Ashton et al. 2019b). Here, we partially dissected the genetic architecture of growth and, in doing so, we have provided a set of SNPs that can assist genome selection of elite broodstock lines for commercial breeding programs. Our results show that the use of reduced genome representation is sufficient to estimate breeding values (i.e. to predict phenotypes) with relative high accuracy even without pedigree information (Supplementary Table 3 and Supplementary Fig. 1). Owing to the inclusion of most of the causal mutations, and decreased limitation due to linkage between SNPs and causal mutations, the use of whole-genome resequencing (WGS) could increase predictive ability. However, empirical results comparing the use of WGS vs a set of SNPs show nil to marginal increase in genomic prediction accuracy (Lu et al. 2020; Yoshida and Yáñez 2021). This suggests WGS is unnecessary for this purpose and supports our suggestion that reduced genome representation provides accurate genomic prediction estimates to assist in the selection of elite broodstock lines. For systems where hundreds or thousands of samples may need to be evaluated over time, such as for Australian snapper aquaculture, this has significant cost implications. Recent advancements in the development of cost-effective genotyping technologies, such as SNP chips, are being developed for this, and related species (Montanari et al. 2022), and will prove crucial in enabling the cost-effective assessment of commercially relevant SNPs.

Our data have the potential to determine individual breeding values, but this can be extended to the population level, with broader evolutionary and ecological implications. Estimating a genotype value and its frequency in a population could make it possible to predict how the population will respond to a future environmental event in respect to the phenotypic trait in question (Hunter et al. 2021; McGaugh et al. 2021). This can be an important tool in the restocking of heavily affected populations, either by climatic change or by overfishing. There are several factors affecting growth in marine organisms, but one of the most important abiotic factors is temperature (Besson et al. 2016; Boltaña et al. 2017; Wellenreuther et al. 2019). In addition, many fish populations are under heavy exploitation, and there is strong evidence of selective harvesting gradually reducing the genetic potential for somatic growth in the population (Enberg et al. 2009; Denechaud et al. 2020), and thus magnifying climatic change impacts (Morrongiello et al. 2021; Wootton et al. 2021). Small changes in growth rates within a population can not only influence individual fitness but can also cause long-lasting shifts in population characteristics and demographic dynamics, including a reduction in fecundity, survival, and recruitment rates (Lorenzen 2016; Denechaud et al. 2020). For most marine fishes, mortality is considerably higher during early-life stages. At that time, individual phenotypes influence the probability of survival, and both field and laboratory research have shown this effect to be size dependent (Johnson et al. 2014). Moreover, reproductive success is often broken into 2 components: reproductive potential and offspring survival, and in many marine fish species both

components are strongly related to body size. Because female fish retain their oocytes internally during their development, maximum reproductive output will be subject to body size constraints (Lambert 2008; Ohlberger et al. 2020). Alternatively, reduced body size can result in small eggs that maximize the number produced, but with a significant reduction of offspring survival (Einum and Fleming 2000; Wootton et al. 2021). Since large organisms have relatively high survival probabilities and reproductive success, it is expected that by predicting the size composition of a population, we can determine its average survival and recruitment dynamics (Garrido et al. 2015). In other words, introducing size-enhanced genotypes to wild populations to modify natural size distributions could in turn increase the resilience to both climatic change and harvesting. Thus, size genomic prediction can be a valuable tool for fishery and conservation management, not only for Australian snapper, but also for other marine fishes.

Using breeding values to predict population response has potential complications, such as failure to incorporate the complexity of factors and uncertainty involved in trait measurement (Hadfield et al. 2010). A wide range of factors affects growth in marine fishes, such as temperature (Boltaña et al. 2017; Wellenreuther et al. 2019), nutritional state (Escalante-Rojas et al. 2020), and intra- and interspecies interactions (Mallard et al. 2020; Korman et al. 2021). Although these factors can induce phenotypic plastic changes to growth, there are also important genetic components in response to these factors (Wellenreuther et al. 2019; Escalante-Rojas et al. 2020). Adaptive variation in our candidate genes that regulate expression and synthesis of stress-related proteins, such as SERPINH1 and PRKAG2 (Wang et al. 2016; Causey et al. 2019), can potentially constrain these responses in Australian snapper and, therefore, the effects of external factors. In addition, Bayesian approaches, such as the 1 used here, can integrate the effects of unknown factors and uncertainty involved in genomic predictions in an efficient way (Ashraf et al. 2020). This suggests that our genomic predictions provide a reliable way to measure individual breeding values of Australian snapper from both wild and captive populations.

This study confirms that, despite the complex polygenetic architecture of growth in Australian snapper, reduced genome representation combined with a mix of bioinformatic approaches can detect candidate genes relevant to a quantitative trait. This opens the possibility of using Bayesian genomic prediction frameworks to measure individual breeding values for growth rates. This information is expected to assist the selective breeding programs for this and related species, and can be used to provide insights into growth changes experienced by wild populations following the exposure of anthrophonic stressors, such as climate change and fishing.

Data availability

As the genomic data of this species are from a taonga and thus culturally important species in Aotearoa New Zealand, the data have been deposited in a managed repository that controls access. Raw and analyzed data are available through the Genomics Aotearoa data repository at https://repo.data.nesi.org.nz/. This was done to recognize Māori as important partners in science and innovation and as intergenerational guardians of significant natural resources and indigenous knowledge.

Supplemental material is available at G3 online.

Acknowledgments

We would like to acknowledge the PFR staff who assisted with the breeding and husbandry operation of the snapper populations, in particular Warren Fantham, who oversees the rearing of finfish larvae, and Therese Wells, who manages the postjuvenile trevally husbandry.

Funding

This research was funded through the MBIE Endeavour Programme "Accelerated breeding for enhanced seafood production" (# C11X1603) to MW and by the Australian Research Council (LP180100756) to LB.

Conflicts of interest

None declared.

Literature cited

Akond Z, Ahsan MA, Alam M, Mollah MNH. Robustification of GWAS to explore effective SNPs addressing the challenges of hidden population stratification and polygenic effects. Sci Rep. 2021; 11(1):1–15.

Alexa A, Rahnenfuhrer J, Alexa MA, Suggests ALL. topGO: Enrichment Analysis for Gene Ontology. R package version 2.26.0; 2016. https://bioconductor.org/packages/release/bioc/html/topGO.html.

Andersen KH, Jacobsen NS, Farnsworth KD. The theoretical foundations for size spectrum models of fish communities. Can J Fish Aquat Sci. 2016;73(4):575–588.

Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Cambridge, (UK): Babraham Institute; 2011.

Ashraf B, Hunter DC, Bérénos C, Ellis PA, Johnston SE, Pilkington JG, Pemberton JM, Slate J. Genomic prediction in the wild: A case study in Soay sheep. *bioRxiv*; 2020. doi: 10.1111/mec.16262.

Ashton DT, Hilario E, Jaksons P, Ritchie PA, Wellenreuther M. Genetic diversity and heritability of economically important traits in captive Australasian snapper (*Chrysophrys auratus*). Aquaculture. 2019a;505:190–198.

Ashton DT, Ritchie PA, Wellenreuther M. 15 years of QTL studies in fish: challenges and future directions. Mol Ecol. 2017;26(6): 1465–1476.

Ashton DT, Ritchie PA, Wellenreuther M. High-density linkage map and QTLs for growth in snapper (*Chrysophrys auratus*). G3 (Bethesda). 2019b;9(4):1027–1035.

Aulchenko YS, Ripke S, Isaacs A, Van Duijn CM. GenABEL: an R library for genome-wide association analysis. Bioinformatics. 2007;23(10):1294–1296.

Baesjou J-P, Wellenreuther M.Genomic Signatures of Domestication Selection in the Australasian Snapper (Chrysophrys auratus). Genes. 2021;12(11):1737.

Barneche DR, Robertson DR, White CR, Marshall DJ. Genetic signatures of domestication selection in the Australasian snapper (*Chrysophrys auratus*). Science. 2018;360(6389):642–645.

Besson M, Vandeputte M, van Arendonk JAM, Aubin J, de Boer IJM, Quillet E, Komen H. Influence of water temperature on the economic value of growth rate in fish farming: the case of sea bass (*Dicentrarchus labrax*) cage farming in the Mediterranean. Aquaculture. 2016;462:47–55.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114–2120.

Boltaña S, Sanhueza N, Aguilar A, Gallardo-Escarate C, Arriagada G, Valdes JA, Soto D, Quiñones RA. Influences of thermal environment on fish growth. Ecol Evol. 2017;7(17):6814–6825.

Bowles E, Marin K, Mogensen S, MacLeod P, Fraser DJ. Size reductions and genomic changes within two generations in wild walleye populations: associated with harvest? Evol Appl. 2020;13(6): 1128–1144.

Bricknell I, Dalmo RA. The use of immunostimulants in fish larval aquaculture. Fish Shellfish Immunol. 2005;19(5):457–472.

Brzyski D, Peterson CB, Sobczyk P, Candès EJ, Bogdan M, Sabatti C. Controlling the rate of GWAS false discoveries. Genetics. 2017; 205(1):61–75.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10(1):421.

Carney TJ, Feitosa NM, Sonntag C, Slanchev K, Kluger J, Kiyozumi D, Gebauer JM, Coffin Talbot J, Kimmel CB, Sekiguchi K, *et al.* Genetic analysis of fin development in zebrafish identifies furin and hemicentin1 as potential novel fraser syndrome disease genes. PLoS Genet. 2010;6(4):e1000907.

Cartwright I, McPhail J, Knuckey I, Smith T, Rayns N, Steer M. National Snapper Workshop, Adelaide (12 to 14 November 2019); 2021.

Catanach A, Crowhurst R, Deng C, David C, Bernatchez L, Wellenreuther M. The genomic pool of standing structural variation outnumbers single nucleotide polymorphism by threefold in the marine teleost *Chrysophrys auratus*. Mol Ecol. 2019;28(6): 1210–1223.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. Mol Ecol. 2013; 22(11):3124–3140.

Causey DR, Kim J-H, Devlin RH, Martin SA, Macqueen DJ. The AMPK system of salmonid fishes was expanded through genome duplication and is regulated by growth and immune status in muscle. Sci Rep. 2019;9(1):1–11.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4(1):7. doi:10.1186/s13742-015-0047-8.

Cook D, Black S. Ngai Tahu Snapper Culture Trials: Growth Performance and Harvest Outcomes. A Plant & Food Research Report, Nelson, New Zealand; 2017.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, *et al.*; 1000 Genomes Project Analysis Group. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–2158.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, *et al.* Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2):giab008.

De-Santis C, Jerry DR. Candidate growth genes in finfish—where should we be looking? Aquaculture. 2007;272(1–4):22–38.

de Maturana LE, Ibáñez-Escriche N, González-Recio Ó, Marenne G, Mehrban H, Chanock SJ, Goddard ME, Malats N. Next generation modeling in GWAS: comparing different genetic architectures. Hum Genet. 2014;133(10):1235–1253.

Denechaud C, Smoliński S, Geffen AJ, Godiksen JA, Campana SE. A century of fish growth in relation to climate change, population dynamics and exploitation. Glob Change Biol. 2020;26(10): 5661–5678.

Dijoux S, Boukal DS. Community structure and collapses in multichannel food webs: role of consumer body sizes and mesohabitat productivities. Ecol Lett. 2021;24(8):1607–1618.

Dong L, Xiao S, Wang Q, Wang Z. Comparative analysis of the GBLUP, emBayesB, and GWAS algorithms to predict genetic values in large yellow croaker (*Larimichthys crocea*). BMC Genomics. 2016;17(1):1–11.

Einum S, Fleming IA. Highly fecund mothers sacrifice offspring survival to maximize fitness. Nature. 2000;405(6786):565–567.

Enberg K, Jørgensen C, Dunlop ES, Heino M, Dieckmann U. Implications of fisheries-induced evolution for stock rebuilding and recovery. Evol Appl. 2009;2(3):394–414.

Escalante-Rojas M, Martínez-Brown JM, Ibarra-Castro L, Llera-Herrera R, García-Gasca A. Effects of feed restriction on growth performance, lipid mobilization, and gene expression in rose spotted snapper (*Lutjanus guttatus*). J Comp Physiol B. 2020;190(3): 275–286.

Falconer DS, Mackay TFC. Introduction to Quantitative Genetics. Harlow, Essex (UK): Longmans Green; 1996.

Fisher R. The Genetical Theory of Natural Selection. Oxford (UK): Clarendon; 1930.

Fisher RA. On the correlation between relatives on the supposition of Mendelian inheritance. Trans R Soc Edinb. 1918;52(2):399–433.

Fraser KP, Rogers AD. Protein metabolism in marine animals: the underlying mechanism of growth. Adv Mar Biol. 2007;52:267–362.

Garrido S, Ben-Hamadou R, Santos AMP, Ferreira S, Teodósio MA, Cotano U, Irigoien X, Peck MA, Saiz E, Ré P, *et al.* Born small, die young: intrinsic, size-selective mortality in marine larval fish. Sci Rep. 2015;5(1):17065–17010.

Gjedrem T, Robinson N, Rye M. The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. Aquaculture. 2012;350–353:117–129.

Goddard ME, Hayes BJ. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet. 2009;10(6):381–391.

Gong J, Zhao J, Ke Q, Li B, Zhou Z, Wang J, Zhou T, Zheng W, Xu P. First genomic prediction and genome-wide association for complex growth-related traits in Rock Bream (*Oplegnathus fasciatus*). Evol Appl. 2021. https://doi.org/10.1111/eva.13218.

Hadfield JD, Wilson AJ, Garant D, Sheldon BC, Kruuk LE. The misuse of BLUP in ecology and evolution. Am Nat. 2010;175(1):116–125.

Hamazaki K, Iwata H. RAINBOW: haplotype-based genome-wide association study using a novel SNP-set method. PLoS Comput Biol. 2020;16(2):e1007663.

Hendry AP, Gotanda KM, Svensson EI. Human Influences on Evolution, and the Ecological and Societal Consequences. London (UK): The Royal Society; 2017.

Holt GJ. Larval Fish Nutrition. New York (NY): John Wiley & Sons; 2011.

Houston RD, Bean TP, Macqueen DJ, Gundappa MK, Jin YH, Jenkins TL, Selly SLC, Martin SAM, Stevens JR, Santos EM, *et al.* Harnessing genomics to fast-track genetic improvement in aquaculture. Nat Rev Genet. 2020;21(7):389–321.

Houston RD, Haley CS, Hamilton A, Guy DR, Tinch AE, Taggart JB, McAndrew BJ, Bishop SC. Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). Genetics. 2008;178(2):1109–1115.

Huang M, Ding L, Wang J, Ding C, Tao J. The impacts of climate change on fish growth: a summary of conducted studies and current knowledge. Ecol Indicators. 2021;121:106976.

Hunter DC, Ashraf B, Bérénos C, Johnston SE, Wilson AJ, Pilkington JG, Pemberton JM, Slate J. Using genomic prediction to detect microevolutionary change of a quantitative trait. *bioRxiv*; 2021. doi: https://doi.org/10.1101/2021.01.06.425564.

Irving K, Wellenreuther M, Ritchie PA. Description of the growth hormone gene of the Australasian snapper, *Chrysophrys auratus*, and

associated intra-and interspecific genetic variation. J Fish Biol. 2021;99(3):1060–1070.

Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, Sidiropoulos K, Cook J, Gillespie M, Haw R, *et al.* The reactome pathway knowledgebase. Nucleic Acids Res. 2020;48(D1): D498–D503.

Johnson DW, Grorud, K -Colvert S, Sponaugle BX. Semmens Phenotypic variation and selective mortality as major drivers of recruitment variability in fishes. Ecol Lett. 2014;17(6):743–755.

Johnston IA, Bower NI, Macqueen DJ. Growth and the regulation of myotomal muscle mass in teleost fish. J Exp Biol. 2011;214(Pt 10): 1617–1628.

Jovelin R, Yan Y-L, He X, Catchen J, Amores A, Canestro C, Yokoi H, Postlethwait JH. Evolution of developmental regulation in the vertebrate FgfD subfamily. J Exp Zool B Mol Dev Evol. 2010;314(1): 33–56.

Kessels MY, Huitema LFA, Boeren S, Kranenbarg S, Schulte-Merker S, van Leeuwen JL, de Vries SC. Proteomics analysis of the zebrafish skeletal extracellular matrix. PLoS One. 2014;9(3):e90568.

Kikuta H, Laplante M, Navratilova P, Komisarczuk AZ, Engström PG, Fredman D, Akalin A, Caccamo M, Sealy I, Howe K, *et al.* Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. Genome Res. 2007; 17(5):545–555.

Kingman GAR, Vyas DN, Jones FC, Brady SD, Chen HI, Reid K, Milhaven M, Bertino TS, Aguirre WE, Heins DC, *et al.* Predicting future from past: the genomic basis of recurrent and rapid stickleback evolution. Sci Adv. 2021;7(25):eabg5285.

Koot E, Wu C, Ruza I, Hilario E, Storey R, Wells R, Chagné D, Wellenreuther M. Genome-wide analysis reveals the genetic stock structure of hoki (*Macruronus novaezelandiae*). Evol Appl. 2021;14(12):2848–2863.

Korman J, Yard MD, Dzul MC, Yackulic CB, Dodrill MJ, Kennedy TA. Changes in prey, turbidity, and competition reduce somatic growth and cause the collapse of a fish population.Ecol Monogr. 2021;91(1):e01427.

Koshikawa S, Giorgianni MW, Vaccaro K, Kassner VA, Yoder JH, Werner T, Carroll SB. Gain of cis-regulatory activities underlies novel domains of wingless gene expression in *Drosophila*. Proc Natl Acad Sci U S A. 2015;112(24):7524–7529.

Lambert Y. Why should we closely monitor fecundity in marine fish populations? J Northw Atl Fish Sci. 2008;41:93–106.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357–359.

Laufkötter C, Zscheischler J, Frölicher TL. High-impact marine heatwaves attributable to human-induced global warming. Science. 2020;369(6511):1621–1625.

Le Rouzic A, Renneville C, Millot A, Agostini S, Carmignac D, Édeline É. Unidirectional response to bidirectional selection on body size II. Quantitative genetics. Ecol Evol. 2020;10(20):11453–11466.

Li H. Toward better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics. 2014;30(20):2843–2851.

Li J, Zhao X, Xin Q, Shan S, Jiang B, Jin Y, Yuan H, Dai P, Xiao R, Zhang Q, *et al.* Whole-exome sequencing identifies a variant in TMEM 132 E causing autosomal-recessive nonsyndromic hearing loss DFNB 99. Hum Mutat. 2015;36(1):98–105.

Li Y, Levran O, Kim JJoo, Zhang T, Chen X, Suo C. Extreme sampling design in genetic association mapping of quantitative trait loci using balanced and unbalanced case-control samples. Sci Rep. 2019;9(1):15504.

Lorenzen K. Toward a new paradigm for growth modeling in fisheries stock assessments: embracing plasticity and its consequences. Fisheries Research. 2016;180:4–22.

Lu S, Liu Y, Yu X, Li Y, Yang Y, Wei M, Zhou Q, Wang J, Zhang Y, Zheng W, *et al.* Prediction of genomic breeding values based on pre-selected SNPs using ssGBLUP, WssGBLUP and BayesB for Edwardsiellosis resistance in Japanese flounder. Genet Sel Evol. 2020;52(1):1–10.

Mallard F, Le Bourlot V, Le Coeur C, Avnaim M, Péronnet R, Claessen D, Tully T. From individuals to populations: how intraspecific competition shapes thermal reaction norms. Funct Ecol. 2020; 34(3):669–683.

Manunza A, Casellas J, Quintanilla R, González-Prendes R, Pena RN, Tibau J, Mercadé A, Castelló A, Aznárez N, Hernández-Sánchez J, *et al.* A genome-wide association analysis for porcine serum lipid traits reveals the existence of age-specific genetic determinants. BMC Genomics. 2014;15(1):758–712.

Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, Derks EM. A tutorial on conducting genome-wide association studies: quality control and statistical analysis. Int J Methods Psychiatr Res. 2018;27(2):e1608.

McGaugh SE, Lorenz AJ, Flagel LE. The utility of genomic prediction models in evolutionary genetics. Proc Biol Sci. 2021;288(1956): 20210693.

Monk CT, Bekkevold D, Klefoth T, Pagel T, Palmer M, Arlinghaus R. The battle between harvest and natural selection creates small and shy fish. Proc Natl Acad Sci USA. 2021;118(9):e2009451118.

Montanari S, Jibran R, Deng C, David C, Koot E, Kirk C, Bassil NV, Morrison-Whittle P, Worthington ML, Pradelles J, *et al.* A multispecies SNP chip enables diverse breeding and management application; 2022 (In preparation).

Morrongiello JR, Horn PL, Ó Maolagáin C, Sutton PJ. Synergistic effects of harvest and climate drive synchronous somatic growth within key New Zealand fisheries. Glob Chang Biol. 2021;27(7): 1470–1484.

Moser G, Lee SH, Hayes BJ, Goddard ME, Wray NR, Visscher PM. Simultaneous discovery, estimation and prediction analysis of complex traits using a Bayesian mixture model. PLOS Genet. 2015;11(4):e1004969.

Murua H, Rodriguez-Marin E, Neilson JD, Farley JH, Juan-Jordá MJ. Fast versus slow growing tuna species: age, growth, and implications for population dynamics and fisheries management. Rev Fish Biol Fisheries. 2017;27(4):733–773.

Nagayama T, Okuhara S, Ota MS, Tachikawa N, Kasugai S, Iseki S. FGF18 accelerates osteoblast differentiation by upregulating Bmp2 expression. Congenit Anom (Kyoto). 2013;53(2):83–88.

Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D. Population genomics of marine fishes: identifying adaptive variation in space and time. Mol Ecol. 2009;18(15):3128–3150.

Oecal S, Socher E, Uthoff M, Ernst C, Zaucke F, Sticht H, Baumann U, Gebauer JM. The pH-dependent client release from the collagen-specific chaperone HSP47 is triggered by a tandem histidine pair. J Biol Chem. 2016;291(24):12612–12626.

Ogier E, Jennings S, Fowler A, Frusher S, Gardner C, Hamer P, Hobday AJ, Linanne A, Mayfield S, Mundy C, *et al.* Responding to climate change: participatory evaluation of adaptation options for key marine fisheries in Australia's South East. Front Mar Sci. 2020;7: 97.

Ohlberger J, Schindler DE, Brown RJ, Harding JMS, Adkison MD, Munro AR, Horstmann L, Spaeder J. The reproductive value of large females: consequences of shifts in demographic structure for population reproductive potential in Chinook salmon. Can J Fish Aquat Sci. 2020;77(8):1292–1301.

Oke KB, Cunningham CJ, Westley PAH, Baskett ML, Carlson SM, Clark J, Hendry AP, Karatayev VA, Kendall NW, Kibele J, *et al.*

Recent declines in salmon body size impact ecosystems and fisheries. Nat Commun. 2020;11(1):1–13.

Papa Y, Oosting T, Valenza-Troubat N, Wellenreuther M, Ritchie PA. Genetic stock structure of New Zealand fish and the use of genomics in fisheries management: an overview and outlook. N Z J Zool. 2021;48(1):1–31.

Parsons DM, Sim-Smith CJ, Cryer M, Francis MP, Hartill B, Jones EG, Le Port A, Lowe M, McKenzie J, Morrison M, *et al.* Snapper (*Chrysophrys auratus*): a review of life history and key vulnerabilities in New Zealand. N Z J Mar Freshwater Res. 2014;48(2):256–283.

Parsons DM, Bian R, McKenzie JR, McMahon SJ, Pether S, Munday PL. An uncertain future: effects of ocean acidification and elevated temperature on a New Zealand snapper (*Chrysophrys auratus*) population. Mar Environ Res. 2020;161:105089.

Peters RH. The Ecological Implications of Body Size. Cambridge (UK): Cambridge University Press; 1986.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One. 2012;7(5):e37135.

Pinsky ML, Eikeset AM, Helmerson C, Bradbury IR, Bentzen P, Morris C, Gondek-Wyrozemska AT, Baalsrud HT, Brieuc MSO, Kjesbu OS, *et al.* Genomic stability through time despite decades of exploitation in cod on both sides of the Atlantic. Proc Natl Acad Sci USA. 2021;118(15):e2025453118.

Robledo D, Fernández C, Hermida M, Sciara A, Álvarez-Dios JA, Cabaleiro S, Caamaño R, Martínez P, Bouza C. Integrative transcriptome, genome and quantitative trait loci resources identify single nucleotide polymorphisms in candidate genes for growth traits in turbot. Int J Mol Sci. 2016;17(2):243.

Sandoval-Castillo J, Robinson, NA Hart, AM Strain, LW Beheregaray LB. Seascape genomics reveals adaptive divergence in a connected and commercially important mollusc, the greenlip abalone (*Haliotis laevigata*), along a longitudinal environmental gradient. Mol Ecol. 2018;27(7):1603–1620.

Sapède D, Gompel N, Dambly-Chaudière C, Ghysen A. Cell migration in the postembryonic development of the fish lateral line. Development. 2002;129(3):605–615.

Sharpe DM, Hendry AP. SYNTHESIS: life history change in commercially exploited fish stocks: an analysis of trends across studies. Evol Appl. 2009;2(3):260–275.

Silvy Y, Guilyardi E, Sallée J-B, Durack PJ. Human-induced changes to the global ocean water masses and their time of emergence. Nat Clim Change. 2020;10(11):1030–1036.

Sinclair-Waters M, Ødegård J, Korsvoll SA, Moen T, Lien S, Primmer CR, Barson NJ. Beyond large-effect loci: large-scale GWAS reveals a mixed large-effect and polygenic architecture for age at maturity of Atlantic salmon. Genet Sel Evol. 2020;52(1):9–11.

Sohail M, Maier RM, Ganna A, Bloemendal A, Martin AR, Turchin MC, Chiang CW, Hirschhorn J, Daly MJ, Patterson N, *et al.* Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. eLife. 2019;8:e39702.

Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang T, Bork P, *et al.* The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 2021;49(D1):D605–D612.

Sztal T, Berger S, Currie PD, Hall TE. Characterization of the laminin gene family and evolution in zebrafish. Dev Dyn. 2011;240(2):422–431.

Tonelli F, Cotti S, Leoni L, Besio R, Gioia R, Marchese L, Giorgetti S, Villani S, Gistelinck C, Wagener R, *et al.* Crtap and p3h1 knock out zebrafish support defective collagen chaperoning as the cause of their osteogenesis imperfecta phenotype. Matrix Biol. 2020;90:40–60.

Tung A, Levin M. Extra-genomic instructive influences in morphogenesis: a review of external signals that regulate growth and form. Dev Biol. 2020;461(1):1–12.

UniProt. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Research. 2021;49(D1):D480–D489.

Uusi-Heikkilä S, Whiteley AR, Kuparinen A, Matsumura S, Venturelli PA, Wolter C, Slate J, Primmer CR, Meinelt T, Killen SS, *et al.* The evolutionary legacy of size-selective harvesting extends from genes to populations. Evol Appl. 2015;8(6):597–620.

Vadstein O, Bergh Ø, Gatesoupe F-J, Galindo-Villegas J, Mulero V, Picchietti S, Scapigliati G, Makridis P, Olsen Y, Dierckens K, *et al.* Microbiology and immunology of fish larvae. Rev Aquacult. 2013;5:S1–S25.

Valenza-Troubat N, Montanari S, Morrison-Whittle P, Ritchie PA, Wellenreuther M. Evaluating new species for aquaculture: a genomic dissection of growth in the New Zealand silver trevally (*Pseudocaranx georgianus*). *Evol Appl*. 2021a (In press).

Valenza-Troubat N, Davy M, Storey R, Wylie MJ, Ritchie PA, Wellenreuther M. Differential expression analyses reveal extensive transcriptional plasticity induced by temperature in New Zealand silver trevally (*Pseudocaranx georgianus*). *Evol Appl*. 2021b (In press).

Valenza-Troubat N, Montanari S, Ritchie PA, Wellenreuther M. Unravelling the complex genetic basis of growth in trevally (*Pseudocaranx georgianus*). G3 (Bethesda). 2021c (Accepted for publication).

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, *et al.* From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43(1):11.10.1–11.10.33.

Wang C, Lo L, Feng F, Zhu Z, Yue G. Identification and verification of QTL associated with growth traits in two genetic backgrounds of Barramundi (*Lates calcarifer*). Anim Genet. 2008;39(1):34–39.

Wang L, Wan ZY, Bai B, Huang SQ, Chua E, Lee M, Pang HY, Wen YF, Liu P, Liu F, *et al.* Construction of a high-density linkage map and fine mapping of QTL for growth in Asian seabass. Sci Rep. 2015;5(1):16358–16310.

Wang Y, Liu Z, Li Z, Shi H, Kang Y, Wang J, Huang J, Jiang L. Effects of heat stress on respiratory burst, oxidative damage and SERPINH1 (HSP47) mRNA expression in rainbow trout *Oncorhynchus mykiss*. Fish Physiol Biochem. 2016;42(2):701–710.

Wellenreuther M, Hansson B. Detecting polygenic evolution: problems, pitfalls, and promises. Trends Genet. 2016;32(3):155–164.

Wellenreuther M, Luyer JL, Cook D, Ritchie PA, Bernatchez L. Domestication and temperature modulate gene expression signatures and growth in the Australasian snapper *Chrysophrys auratus*. G3 (Bethesda). 2019;9(1):105–116.

Wootton HF, Audzijonyte A, Morrongiello J. Multigenerational exposure to warming and fishing causes recruitment collapse, but size diversity and periodic cooling can aid recovery. Proc Natl Acad Sci USA. 2021;118(18):e2100300118.

Wright S. The genetical structure of populations. Ann Eugen. 1949;15(1):323–354.

Wu L, Yang Y, Li B, Huang W, Wang X, Liu X, Meng Z, Xia J. 2019 First genome-wide association analysis for growth traits in the largest coral reef-dwelling bony fishes, the Giant grouper (*Epinephelus lanceolatus*). Mar Biotechnol (NY). 2019;21 (5):707–717.

Wu N, Li C-J, Gui J-F. Molecular characterization and functional commonality of nucleophosmin/nucleoplasmin in two cyprinid fish. Biochem Genet. 2009;47(11–12):749–762.

Yang J, Jiang H, Yeh C-T, Yu J, Jeddeloh JA, Nettleton D, Schnable PS. Extreme-phenotype genome-wide association study (XP-GWAS): a method for identifying trait-associated variants by sequencing pools of individuals selected from a diversity panel. Plant J. 2015; 84(3):587–596.

Yang W, Wang Y, Jiang D, Tian C, Zhu C, Li G, Chen H. ddRADseq-assisted construction of a high-density SNP genetic map and QTL fine mapping for growth-related traits in the spotted scat (*Scatophagus argus*). BMC Genomics. 2020;21(1):1–18.

Ye B, Wan Z, Wang L, Pang H, Wen Y, Liu H, Liang B, Lim HS, Jiang J, Yue G, *et al.* Heritability of growth traits in the Asian seabass (*Lates calcarifer*). Aquac Fish. 2017;2(3):112–118.

Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, Frayling TM, Hirschhorn J, Yang J, Visscher PM, *et al.*; GIANT Consortium. Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. Hum Mol Genet. 2018;27(20):3641–3649.

Yoshida GM, Yáñez JM. Increased accuracy of genomic predictions for growth under chronic thermal stress in rainbow trout by prioritizing variants from GWAS using imputed sequence data. Evol Appl. 2021. https://doi.org/10.1111/eva.13240.

Zhou Z, Han K, Wu Y, Bai H, Ke Q, Pu F, Wang Y, Xu P. Genome-wide association study of growth and body-shape-related traits in large yellow croaker (*Larimichthys crocea*) using ddRAD sequencing. Mar Biotechnol (NY). 2019;21(5):655–670.

Communicating editor: D.-J. de Koning