

**Description of the growth hormone gene of the Australasian snapper, *Chrysophrys auratus*, and associated intra- and interspecific genetic variation**

Kate Irving<sup>1</sup>, Maren Wellenreuther<sup>2,3</sup> and Peter A. Ritchie<sup>1</sup>

<sup>1</sup> School of Biological Sciences, Victoria University of Wellington, Wellington, New Zealand

<sup>2</sup> The New Zealand Institute for Plant and Food Research Limited, Nelson, New Zealand

<sup>3</sup> School of Biological Sciences, University of Auckland, Auckland, New Zealand

\* Corresponding author: Peter Ritchie

E: [Peter.Ritchie@vuw.ac.nz](mailto:Peter.Ritchie@vuw.ac.nz)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: [10.1111/jfb.14810](https://doi.org/10.1111/jfb.14810)

## ABSTRACT

The Growth Hormone (GH) gene of the marine teleost, the Australasian snapper (*Chrysophrys auratus*), was identified and characterised from the reference genome showing it was approximately 5,577 bp in length and consisted of six exons and five introns. Large polymorphic repeat regions were found in the first and third introns, and putative transcription factor binding sites were identified. Phylogenetic analysis of the GH genes of perciform fish showed largely conserved coding regions and highly variable non-coding regions among species. Despite some exon sequence variation and an amino acid deletion identified between *C. auratus* and its sister species *Chrysophrys/Pagrus major*, the amino acid sequences and putative secondary structures were largely conserved across the Sparidae. A population-level assessment of 99 samples caught at five separate coastal locations in New Zealand revealed six variable alleles at the intron 1 site of the *C. auratus* GH gene. A population genetic analysis suggested that *C. auratus* from the five sample locations were largely panmictic, with no evidence for departure from Hardy Weinberg Equilibrium, and have a high level of heterozygosity. Overall these results suggest that the GH gene is largely conserved across the coding regions, but some variability could be detected.

## KEYWORDS

*Chrysophrys auratus*, growth hormone, polymorphism, genetic variation, growth.

## 1. Introduction

Growth hormone (GH) is a pluripotent hormone produced by the pituitary gland in teleosts and is, as in other vertebrates, one of the main regulators of postnatal somatic growth (Goodman, 1993; Reinecke *et al.*, 2005). GH works by binding to a single pass-transmembrane receptor, called the GH receptor (GHR), in target tissue. This ligand binding process induces receptor dimerization to produce an active trimeric complex (Pérez-Sánchez *et al.*, 2002). GH works by binding to GH receptors (GHR) in target tissues and can bind directly to the cells in tissues such as muscle (Vijayakumar *et al.*, 2012) and bone (Butler & Roith, 2001); however, its main target is the liver (Backeljauw & Hwa, 2016). Upon GHR binding, GH stimulates the synthesis and release of insulin-like growth factors (IGF), IGF1 or IGF2, from target tissues. IGFs released from the liver travel to target tissues such as bone, muscle, and heart (Moriyama *et al.*, 2000). IGFs are involved in the regulation of protein, lipid, carbohydrate, and mineral metabolism in cells, and the differentiation and proliferation of cells. These processes ultimately lead to growth (Moriyama *et al.*, 2000). The GH-IGF axis is complex and there is still a large amount that is unknown about its intricate control of growth.

Although the basic GH-IGF regulation of growth is conserved in vertebrates (Moriyama *et al.*, 2000), fish represent a unique model for studying the fundamental mechanisms regulating growth. In other vertebrates, such as mammals, somatic muscle growth is achieved through hypertrophy (increase in cell size) of muscle fibres formed prior to birth. This process is determinate, only occurring in early life stages (Fuentes *et al.*, 2013; Rowe & Goldspink, 1969). In fish however, muscle growth is indeterminate, occurring throughout the whole life cycle of a fish, and can be achieved through both hypertrophy and hyperplasia (cell proliferation) (Mommsen, 2001). In fish, GH participates in almost all major physiological processes in the body including the regulation of ionic and osmotic balance, lipid, protein, and carbohydrate metabolism, skeletal and soft tissue growth, reproduction and immune function.

Recent studies have indicated that GH affects several aspects of behaviour, including appetite, foraging behaviour, aggression, and predator avoidance, and thus it also has ecological consequences (Agellon *et al.*, 1988; Perez-Sanchez, 2000; Reinecke *et al.*, 2005). *GH* genes have been characterised in a large number of important aquaculture fish species (De-Santis & Jerry, 2007; Reinecke *et al.*, 2005) with the aim of developing markers for growth rate improvement. The cross-species DNA sequence information on the *GH* gene has enabled comparative analyses and these have shown that the gene is comprised of conserved coding regions but variable non-coding regions, such as the 5' flanking regions and intronic sequences (Almuly *et al.*, 2005; De-Santis & Jerry, 2007; Quéré *et al.*, 2010).

*GH* gene variation has been associated with growth traits in a range of terrestrial livestock (Knorr *et al.*, 1997; Kuhnlein *et al.*, 1997; Tambasco *et al.*, 2003) and aquaculture fish species (Almuly *et al.*, 2008; Jaser *et al.*, 2017; Sawayama & Takagi, 2015b). A number of variable repeat regions have been identified in the non-coding regions of *GH* genes in vertebrate species (Almuly *et al.*, 2000; Arango *et al.*, 2014; Sawayama & Takagi, 2015a). It has been shown that repeat regions within non-coding sections (i.e. introns and 5' regulatory regions) of genes can affect transcription, mRNA splicing, and gene silencing (Chorev & Carmel, 2012; Li *et al.*, 2002), and can, in turn, lead to variable phenotypic effects (Li *et al.*, 2004). Furthermore, transcription factor binding sites within intronic repeat regions have been shown to directly affect gene expression (Almuly *et al.*, 2008). For example, repeat elements in the first and third introns of *GH* have been developed as genetic markers for improved growth for the important aquaculture fish species gilt-head bream *Sparus aurata* (Linnaeus 1758) (Almuly *et al.*, 2008) and red seabream *Pagrus major* (Temminck & Schlegel 1843) (Sawayama & Takagi, 2015b) respectively.

The Australasian snapper *Chrysophrys auratus* (Forster 1801) is a long-lived, demersal species from the family Sparidae (Bernal-Ramírez *et al.*, 2003; Kailola, 1993) that is widely distributed

throughout the temperate to subtropical waters of New Zealand, Australia, Lord Howe Island, and the Norfolk Islands (Paulin, 1990). *C. auratus* supports one of New Zealand's largest commercial fisheries (Annala *et al.*, 2000; Colman, 2010; Parsons *et al.*, 2014; Paul, 1977) and is also highly sought after by recreational fishers (Parsons *et al.*, 2014). Understanding the genetic architecture and variation of the *C. auratus* *GH* gene (*caGH*) is of great importance for future breeding programmes in *C. auratus* and many of its closely related species that are currently used or developed for aquaculture, such as *P. major*, *S. aurata*, red porgy *Pagrus pagrus* (Linnaeus 1758), and common dentex *Dentex dentex* (Linnaeus 1758) (Basurco *et al.*, 2011). This study characterised the *GH* gene structure in *C. auratus* and identified several sites of variation. A phylogeny was reconstructed using the *GH* genes of 21 species from the Perciformes order and three outgroups, to investigate the evolutionary history of the *GH* gene in perciform fish and to further understand where the extensive variation within this gene arises. Finally, a limited population study was conducted to assess the genetic variation of a large repeat region in the first intron of the *caGH* gene.

## 2. Materials and Methods

### *Ethical Statement*

The research was conducted according to Plant and Food Research Ltd animal ethics procedures and under approval 21765 from the Victoria University of Wellington (VUW) Animal Ethics Committee, and complied with the VUW Code of Ethical Conduct and was undertaken within the provisions of the New Zealand Animal Welfare Act 1999.

## Characterisation of the *caGH* gene using genome-wide sequences

A comparative gene approach, using genetic information from the closest living relative of *C. auratus*, *P. major* as a sequence reference, was used to identify and annotate the *GH* gene in the *C. auratus* reference genome. These two species are considered sister taxa; however, the genus they belong to remains undetermined (Parsons *et al.*, 2014). The *P. major GH* gene was obtained from Sawayama and Takagi (2015a) (GenBank accession: AB904715). A BLAST search between the *P. major GH* gene (*pmaGH*) and the *C. auratus* reference genome (Wellenreuther *et al.*, unpublished) was carried out in Geneious v10 (Kearse *et al.*, 2012) under the default settings. Once located, the *C. auratus GH* gene (*caGH*) was annotated by eye using the *pmaGH* as a reference. This process revealed a DNA sequence data gap in the *C. auratus* reference genome at the beginning of the *caGH*. This gap spanned from before the start of the gene to the end of intron 1 (64,690 – 67,000 nucleotide positions, scaffold 819 of the *C. auratus* genome assembly version 1.0). The gap was filled using two different methods. Firstly, the 5' regulatory region and exon 1 were reconstructed using whole genome re-sequenced data from a sample of 12 wild-catch *C. auratus* individuals from three populations around New Zealand (two females and two males per population), namely Manukau Harbour (37.0000° S, 174.6667° E), Tasman Bay (40.9980° S, 173.1649° E) and Hauraki Gulf (36.5989° S, 175.1894° E). These re-sequenced genomes were unassembled and consisted of 100 bp sequence reads. A Geneious v10 BLAST search was carried out between the *pmaGH* sequence and each re-sequenced genome. BLAST hits for each re-sequenced genome were then mapped back to the *pmaGH* using the 'map to reference' function in Geneious v10. A consensus sequence of the *caGH* gene was obtained for each individual *C. auratus*. The 12 consensus sequences from these individuals were then aligned in Geneious v10 using a MUSCLE alignment and an overall consensus sequence of all 12 wild individuals was obtained. The DNA

sequence gap for the 5' regulatory region and the first exon was successfully filled using this method.

Second, intron 1 could not be interpreted using the re-sequenced data owing to the technical difficulties of determining the DNA sequence of a highly repetitive region, which caused the DNA sequence reads to be extremely variable, so direct Sanger sequencing was used to determine the DNA sequence. DNA was extracted from tissue samples of 20 wild juvenile *C. auratus* using a modified high salt extraction method (Aljanabi & Martinez, 1997). Intron 1 was amplified by polymerase chain reaction (PCR) using the forward primer SG1 (5'-AGAACCTGAACCAGACATGG-3') and reverse primer FIS (5'-AGGTGTTGAACTCTGCTGAC-3') (Almuly *et al.*, 2000). These primers bind at the 613-632 and 1148-1129 nucleotide intervals respectively (GenBank accession: MW972222). PCRs were carried out in 15- $\mu$ L volumes containing 1x PCR buffer (67 mM Tris-HCl (pH 8.8 at 25°C), 16 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>), 0.6  $\mu$ M Bovine Serum Albumin (BSA), 200  $\mu$ M of each dNTP, 0.4  $\mu$ M of each primer, 2 mM MgCl<sub>2</sub>, 1 unit of Biotaq DNA polymerase (Bioline Reagents Limited), and 1  $\mu$ L of template DNA (~50 ng). The PCR cycling conditions were as follows: an initial denaturation step at 95°C for 5 minutes followed by 35 cycles of 94°C for 30 seconds, 56°C for 30 seconds, 72°C for 60 seconds, and a final extension step at 72°C for 10 minutes. PCR products were electrophoresed in a 1.5% agarose gel at 90 V for 30 minutes. Gels were stained with ethidium bromide and viewed under UV light. Homozygote PCR products were then purified with ExoSAP-IT (Amersham Pharmacia Biotech) and their sequence determined using an ABI3730 DNA Analyzer (Massey Genome Services, Palmerston North). A complete intron 1 sequence was obtained and inserted into the correct position in the *caGH* gene in Geneious v10 using the *pmaGH* gene as a reference. This resulted in a full *GH* gene sequence for *C. auratus* that could subsequently be annotated and used for further analyses. It should be acknowledged that owing to the multiple methods required to fill the sequence gap in the *C.*

*auratus* reference genome, the *caGH* sequence presented in this study does not belong to a single individual and is instead a combination of sources and individuals brought together using the *GH* gene from *P. major*, the closest living relative of *C. auratus*, as a reference.

The *caGH* sequence was searched for polymorphic loci by eye and by using the Repeat Finder plugin in the reference *caGH* and the 12 re-sequenced genomes in Geneious v10 (Kearse *et al.*, 2012). Putative transcription factor binding sites and regulatory motifs were predicted using the EMBOSS plugin (Rice *et al.*, 2000); the settings allowed no mismatches and a minimum length of 4 bp for matches. The *caGH* amino acid sequence was determined and the secondary structure was predicted using Geneious v10 plugin EMBOSS Protein (Rice *et al.*, 2000).

#### *Phylogenetic analysis of the GH gene in perciform fish*

*GH* gene sequences from 20 fish species in the Perciformes order were obtained from GenBank and imported to MEGA v7 (Kumar *et al.*, 2016) along with the *caGH* sequence from this study (Table 1). Three outgroup species from the orders Pleuronectiformes, Siluriformes, and Salmoniformes, within the class Actinopterygii, were also imported to MEGA from GenBank (Table 1).

The *GH* sequences were aligned in MEGA using the MUSCLE alignment option. A MEGA model test was then carried out to determine the most appropriate model for subsequent phylogenetic analysis. The K2 + G + I (Kimura 2-parameter model (Kimura, 1980)) was indicated in the model test as the most appropriate model to use for this dataset. A maximum-likelihood tree was constructed in MEGA using the Kimura 2-parameter model with 1000 bootstrap replicates. Scaled diagrams of the basic *GH* gene structures (exon-intron sizes) were manually produced for each species. Additionally, *GH* sequences were translated into their amino acid sequences in Geneious v10 and aligned using a MUSCLE alignment.

Similarities between amino acid sequences, hydrophobicity, and polarity were assessed by

eye. Secondary structures were predicted using the EMBOSS Protein plugin (Rice *et al.*, 2000).

### *Intron 1 variation*

Intronic repeat variation was assessed in 99 juvenile *C. auratus* collected from five harbours in the north of New Zealand; Kaipara (36.3821° S, 174.2312° E; n = 20), Parengarenga (34.5456° S, 172.9338° E; n = 20), Rangaunu (34.9546° S, 173.9338° E; n = 19), Bay of Islands (35.2218° S, 174.1413° E; n = 20), Doubtless Bay (34.9277° S, 173.4387° E; n = 20). Fish were caught by NIWA researchers in 2013/14 under MPI Special Permit (#666) and we sampled the specimens that has been stored frozen for five years. Genomic DNA was extracted from samples using a modified high-salt extraction method (Aljanabi & Martinez, 1997). The repeat regions of intron 1 were amplified using PCR with the forward primer SG1 (5'-M13 - AGAACCTGAACCAGACATGG-3') and reverse primer FIS (5'-AGGTGTTGAACTCTGCTGAC-3') (Almuly *et al.*, 2000). Fluorescent labels (FAM) were incorporated into PCR products during the amplification step using the M13 labelling method (Schuelke, 2000). PCRs were carried out in 15- $\mu$ L volumes containing 1x PCR buffer (67 mM Tris-HCl (pH 8.8 at 25°C), 16 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>), 0.6  $\mu$ M BSA, 200  $\mu$ M of each dNTP, 0.1  $\mu$ M forward primer, 0.4  $\mu$ M reverse primer, 0.4  $\mu$ M FAM labelled M13 primer, 2 mM MgCl<sub>2</sub>, 1 unit of Biotaq DNA polymerase (Bioline Reagents Limited), and ~100 ng of DNA. Thermal cycling conditions were as follows: An initial step of 95°C for 5 minutes followed by 26 cycles of 94°C for 30 seconds, 63°C for 30 seconds, and 72°C for 60 seconds, followed by 8 cycles of 94°C for 30 seconds, 53°C for 45 seconds, and 72°C for 45 seconds, followed by a final extension step of 72°C for 10 minutes. PCR products were electrophoresed in a 1.5% agarose gel using 90 V for 30 minutes. Gels were stained with ethidium bromide and viewed under UV

light. Amplified fragments were sized on an ABI3730 Genetic Analyzer using the LIZ 1200 standard at the Massey Genome Services.

#### *Data analysis of the caGH intron 1 repeat region*

Allele peaks were viewed and analysed in Geneious v10 (Kearse *et al.*, 2012) and manually placed into size ‘bins’ for each allele size based on known core repeat sizes. The intron 1 repeat region is made up of varying combinations and repeats of the following core repeat sequences: 3’ - TGACCTGTCTGTCTCTCTCTCTCTCTGTC - 5’; 3’ - TGACCTGTCTGTCTCTC - 5’; 3’ - TGACGTGTCTGTCTCTCTCTCTTTCTCTCTGTC - 5’; 3’ - TGACCTGTCTGTCTGTCTGTCAGTGGTGCTGATGCTGTCGGTGCTGTCTC - 5’ (Figure 1). Alleles were named according to their respective peak sizes. Allele number ( $N_a$ ), observed heterozygosity ( $H_o$ ), expected heterozygosity ( $H_e$ ), and allele frequencies were calculated in Microsoft® Excel plugin GenAlEx 6.5 (Peakall & Smouse, 2006). Weir and Cockerham’s  $F$ -statistics  $\theta$  ( $F_{ST}$ ), and  $f$  ( $F_{IS}$ ) were calculated using FSTAT (Goudet, 2001; Weir & Cockerham, 1984) and deviations from the Hardy-Weinberg equilibrium (HWE) were assessed in GENEPOP (Raymond, 1995) using Fisher’s exact test.

### 3. Results

#### *Structure of the caGH gene*

The *caGH* (GenBank accession: MW972222) was identified on scaffold 819 of version 1.0 of the *C. auratus* genome. Nucleotide sequence analysis based on *pmaGH* annotation (Sawayama & Takagi, 2015a) (GenBank accession: AB904715) revealed that the *caGH* is approximately 5,577 bp in length. Its putative start site (start of exon I) is at nucleotide 66,741, and the end site is at nucleotide 72,318. The *caGH* is composed of six exons (I–VI) of length 71, 134, 114,

144, 147, 296 bp respectively, and five introns (1–5) of length 410, 713, 3384, 85, 79 bp respectively. A 567 bp long DNA sequence upstream of the first exon (66,174 – 66,741 bp) was also annotated based on the *pmaGH* sequence (Figure 1). The *GH* locus identified in this study was the only match in the BLAST search between the *C. auratus* genome and the *pmaGH*.

The 12 whole genome sequences of *C. auratus* showed that the first 150 bp of the 5' flanking region contains nine variable single-nucleotide sites. The first two repeat regions in the 5' flanking region were monomorphic in all 12 individuals, while the third repeat sequence in the 5' flanking region, (AC)<sub>15</sub>, ranged from 13 to 15 repeats in these individuals (Figure 1). Aside from these variable sites, the 5' flanking region, exon I, and the first 43 bp of intron 1 (before the repeat region) were completely conserved among all 12 individuals. Intraspecific variation at the intron 1 locus could also be assessed from the 10 intron 1 repeat sequences obtained through direct sequencing. The repeat region within this intron varied in length between individuals. Three individuals had a repeat region of 582 bp in length while the other seven individuals had repeat regions 393 bp long. Intron 2 was highly polymorphic which meant that finding a consensus sequence for each re-sequenced genome from the 12 wild individuals was unsuccessful in parts. The first repeat region in intron 2 with the motif 'CCCTGA' ranged from three to six repeats among individuals (Figure 1). The second repeat region in intron 2 with the motif 'GT' was monomorphic across all 12 individuals (Figure 1). As with intron 1, the highly repetitive nature of the intron 3 sequence meant that no consensus sequences could be constructed using the re-sequenced data. All remaining exons (II–VI) and introns (4–5) were entirely conserved across all re-sequenced and reference individuals.

The 5' regulatory region contains a typical TATA box located at nucleotides -106 to -112 upstream of the first codon of exon I. Two other TATA boxes were identified in the 5' flanking region at nucleotides -230 to -236 and nucleotides -323 to -329 upstream of the first codon.

There are two putative pituitary-specific transcription factor (Pit-1a) consensus sequences located upstream of the first codon at nucleotides -200 to -207 and nucleotides -124 to -131. Additionally, a cAMP response element (CRE) was identified at nucleotides -76 to -82 upstream of the first codon. These consensus sequences are also found throughout the full gene sequence. The repeat region in the 5' flanking region with the sequence motif (ACTGTTCTCT)<sub>10</sub> (Figure 1) contains a putative glucocorticoid response element (GRE) consensus sequence (TGTTCT). There are also several GRE sites in the 5' flanking region. No putative thyroid response element (TRE) or oestrogen response elements (ERE) were detected in the 5' flanking region. Alternatively, putative TREs (TGACCT) were located in intron 1 within the core repeat unit of this intron's repeat region. A putative GAGA factor consensus sequence (TCTCTC) was also found in this core repeat unit (Figure 1). Putative EREs were identified in intron 2, and exons III and IV. All other putative transcription factor binding sites and regulatory motifs identified using EMBOSS (Rice *et al.*, 2000), and the amino acid sequences and predicted secondary structures are included in the supplementary information (Supplementary information Table S1, Figure S1, and Figure S4 respectively). The *C. auratus* GH consists of 203 amino acids.

#### *Phylogenetic analysis of the GH gene in perciform fish*

The 21 perciform species were monophyletic with respect to the three outgroup taxa (Figure 2). All the species formed monophyletic groups within their respective families. *C. auratus* was grouped within the family Sparidae. Support for monophyly for all monophyletic groups was supported by high bootstrap values (94–100), while the relationships between families were not highly supported (19–100). All perciform fish species had a highly conserved *GH* gene structure consisting of six exons and five introns (Figure 3). Of the three outgroups, olive flounder *Paralichthys olivaceus* (Temminck & Schlegel 1846) and chum salmon

*Oncorhynchus keta* (Walbaum, 1792) conformed to this gene structure, while walking catfish *Clarias batrachus* (Linnaeus 1758) consisted of only five exons and four introns. *P. olivaceus* was more closely related to the Perciformes family than the other two outgroups, which was reflected in the structures of the exons. The lengths of exons II–V were the same across 19 of the 21 perciform species, and *P. olivaceus*. Sister species, *C. auratus* and *P. major* shared these conserved lengths for exons II, IV, and V; however, both species had an exon III that was 3 bp shorter than those of the other perciform species. While the exons were highly conserved, there was high variability across species in the length of introns. Introns 4 and 5 were similar in length across species, while introns 1–3 varied extensively (Figure 3). This variability in intron length is responsible for the large range of gene sizes within perciform fish.

Alignment of the amino acid sequences of each species showed a close similarity among perciform fish (Supplementary information Figure S1). All perciform fish had a sequence of 204 amino acids, excluding *P. major* and *C. auratus*, which had 203 owing to a 3 bp deletion in their third exons (one codon/amino acid). *P. major* and *C. auratus* are lacking the 67<sup>th</sup> amino acid in the sequence (leucine). Where this amino acid has been deleted, most perciform fish have the sequence isoleucine-phenylalanine-leucine while *P. major* and *C. auratus* have isoleucine-phenylalanine-[gap]. Following the leucine (hydrophobic), the majority of perciform fish have a glutamine (hydrophilic). However, following the amino acid deletion site, *P. major* and *C. auratus* have a proline (hydrophobic). This results in *P. major* and *C. auratus* and all other perciform fish having three consecutive hydrophobic amino acids (isoleucine and phenylalanine = hydrophobic and non-polar) at this site despite the missing leucine (Supplementary information Figure S2). The proline following the deleted leucine (non-polar) in *P. major* and *C. auratus* is also non-polar compared to the glutamine (polar) in all other perciform fish. This also leads to three consecutive non-polar amino acids at this site (Supplementary information Figure S3). Therefore, despite variability in amino acid sequence,

the overall secondary structure of perciform *GH* genes is relatively conserved across species (Supplementary information Figure S4).

#### *Intron 1 variation*

Population statistics including number of alleles, observed heterozygosity, and expected heterozygosity for the intron 1 locus are detailed in Table 2. The number of intron 1 alleles in each of the wild *C. auratus* populations ranged from three to six with three of the alleles occurring in all populations (alleles: 485, 550, 765). Observed heterozygosities ranged from 0.500 – 0.800 and expected heterozygosities ranged from 0.546 – 0.678. Weir and Cockerham's F-statistic  $\theta$  ( $F_{ST}$ ) across all five sampled populations was -0.011 and  $f$  ( $F_{IS}$ ) was -0.053, suggesting that *C. auratus* from the five sample sites are largely panmictic and have a high level of heterozygosity. Results of the Hardy-Weinberg exact test are also presented in Table 2. Neither the sample sites nor the global test across all individuals at the intron 1 locus exhibited significant  $p$ -values and therefore offer no evidence for departure from HWE.

#### 4. Discussion

This study identified and characterised the *GH* gene structure in *C. auratus* and identified several sites of variation. The *caGH* is approximately 5,577 bp in length and composed of six exons and five introns. This gene structure is conserved across a large number of teleost species (Almuly *et al.*, 2000; Chen *et al.*, 2004). The *caGH* gene identified in this study was the only BLAST match in the *C. auratus* reference genome for the *pmaGH* gene, suggesting that there is only one copy of the *GH* gene. Comparison of 13 *C. auratus* individuals revealed conserved exon sequences and highly variable non-coding regions. Repeat regions were identified only within the non-coding regions of *caGH* (5' flanking region, introns 1–3). This pattern appears to be consistent, not only within fishes, but also across other animal groups (Quére *et al.*, 2010).

In this study, these repeat sequences were responsible for a large majority of the intraspecific genetic variation within *C. auratus*. Similar high levels of intraspecific repeat variation within promoters and introns have been reported in a range of species (Almuly *et al.*, 2008; Arango *et al.*, 2014; Nie *et al.*, 2005; Quéré *et al.*, 2010; Zhang *et al.*, 2009).

Putative transcription factor binding sites and regulatory motifs were identified in *caGH*. It has been suggested that teleost *GH* genes are regulated by the synergistic effects of multiple regulatory factors in the promoter region, including Pit-1a, CRE, GRE, TRE, ERE (Moriyama *et al.*, 2006; Zhang *et al.*, 2009). The 5' flanking region of *caGH* contained putative binding sequences for Pit-1a, CRE, and GRE. Consensus sequences for ERE and TRE were not found in the 5' flanking region/promoter. Alternatively, consensus sequences for ERE were found in intron 2 and exons III and IV, and putative TRE binding sequences were identified within the core repeat unit in the intron 1 repeat region. The presence of the TRE regulatory motif in the intron 1 repeat sequence has also been described in two close relatives of *C. auratus*, *P. major* (Sawayama & Takagi, 2015a) and *S. aurata* (Almuly *et al.*, 2000). Furthermore, TRE binding within the first introns of other genes has been shown to directly influence the expression of those genes (Fu *et al.*, 2006; Plateroti *et al.*, 2006). The intron 1 core repeat sequence in *caGH* also contains a consensus sequence that binds a GAGA-factor that is known to regulate transcription at multiple levels (Lomaev *et al.*, 2017). The presence of these transcription factor binding sites in the first intron of *caGH* suggests the importance of this intron for gene expression. As these binding elements occur within a repeat sequence, the number of repeats in intron 1 (and therefore transcription factor binding sites) may alter the expression levels of this gene. An amino acid sequence analysis and predicted secondary structure analysis of *caGH* were conducted on the predicted polypeptide that was 203 amino acids long, which is similar to those in other teleost fish species.

Accepted Article

A phylogeny based on the *GH* genes of 21 species from the Perciformes order and three outgroups was reconstructed using a maximum likelihood method. The phylogeny was constructed to investigate the evolutionary history of the *GH* gene in perciform fish and to further understand where the extensive variation within this gene arises. It also offers insight into where *C. auratus* and its *GH* gene sequence lies in the evolution of this order. Perciformes is the most diverse of all fish orders and the largest of the vertebrate orders (Nelson, 2006). In this study, the 21 species classified under the order Perciformes are monophyletic with respect to the three outgroups. Families are grouped monophyletically, and general relationships are supported by other molecular phylogenies (Betancur-R *et al.*, 2017; Wei *et al.*, 2014). However, the exact relationships between some families are not well supported in this phylogeny and remain unclear in other studies.

While these higher-level classifications are still unclear, several overall trends can be observed in these species. Firstly, the gene structure of *C. auratus* comprising six exons and five introns is conserved across all other species in this study (20 perciform and two outgroups) with the exception of the third outgroup, *C. batrachus* (order: Siluriformes), which has five exons and four introns. This is consistent with the current idea that an intron was inserted into what was originally exon V, leading to the formation of a fifth intron and sixth exon (Moriyama *et al.*, 2006; Zhang *et al.*, 2009). It is suggested this insertion took place only within the ray-finned fish, after the evolutionary separation of Cypriniforms but before the divergence of the Salmoniformes, Perciformes, and Terradontiformes (Moriyama *et al.*, 2006; Zhang *et al.*, 2009) which is consistent with the phylogeny in this study. The lengths of the six exons are highly conserved among species in this study, in particular, exons II–V are equal in length across perciforms. An exception to this, the third exons of *P. major* and *C. auratus GH*, are 3 bp shorter than the third exons of all other perciforms in this study. Amino acid sequence and secondary structure analyses on the predicted polypeptide sequences of *caGH* and all other

species in this study showed that the shortened third exon in *P. major* and *C. auratus* is associated with one less amino acid (leucine) at that site.

Despite the loss of one amino acid, the predicted secondary structures of *P. major* and *C. auratus* remain fundamentally consistent with those of other perciform species in this study. The findings in the present study suggest that this is due to the alteration of amino acids surrounding the deletion through evolution, such that the deletion is tolerated with negligible disturbance to protein function. The areas of a protein that are prone to insertion or deletion events are not well understood; however, some studies have observed that deletions are often followed by compensatory substitutions elsewhere in the amino sequence (Choi *et al.*, 2012; Jones, 2005). The conservation of the amino acid sequence hydrophobicity, polarity, and secondary structure in GH throughout perciform evolution suggests not only common ancestry, but also selection for a particular architecture that ensures protein functionality (Pál *et al.*, 2006; Taylor *et al.*, 2004). While the lengths of the exons are highly conserved, there is extensive variation in the lengths of the introns across species. This variation in intron length is largely responsible for the variable GH gene lengths between perciform fish. This appears to be consistent across many vertebrate species groups (Almuly *et al.*, 2000).

#### *Genetic variation at intron 1 locus*

Analysis of the variation at the intron 1 repeat locus in wild *C. auratus* across five sample sites in the north of New Zealand showed that it is polymorphic. There was little differentiation between sites, a high level of heterozygosity, and conformation to HWE at all sites. Repeats in non-coding regions have been shown to affect transcription, mRNA splicing, and gene silencing (Chorev & Carmel, 2012; Li *et al.*, 2002). These functions affect gene expression and can, in turn, lead to variable phenotypic effects (Li *et al.*, 2004). For example, repeat sequences in the introns of the GH gene have been linked to changes in growth in multiple species

(Almuly *et al.*, 2008; Arango *et al.*, 2014; Nie *et al.*, 2005; Sawayama & Takagi, 2015b). It has been reported that this kind of gene regulation occurs frequently in 5' proximal introns in particular, and occurs across multiple genes and species (Chorev & Carmel, 2012). This is consistent with the findings of this study.

The exact mechanisms that allow intron repeat sequences to influence gene expression are still largely undefined. One hypothesis proposes that owing to the slow, energetically expensive nature of transcription, longer repeat regions are more costly to transcribe and thus lead to lower levels of gene expression (Castillo-Davis *et al.*, 2002). The reduced expression of gene variants with long introns has been observed in the first introns of the Friedreich ataxia (Ohshima *et al.*, 1998) and *GH* genes (Almuly *et al.*, 2008). It is also thought that repeat elements may affect DNA packaging and have an important role in marking sites of interaction of linked loci, leading to their 3D compartmentalization and coordinated regulation of gene expression (Kumar *et al.*, 2010). Therefore, it is conceivable that the number of repeats in an intronic repeat region can have a direct effect on the expression of that gene. The current study identified large repeat regions in introns 1 and 3 of *caGH* that exhibited repeat number variations. It should be noted that the variation at the intron 1 locus described in this study was based on a simple analysis scoring 'alleles' entirely by length. Due to the large diversity of repeat types in this intron and the small sample size used in this study, it is likely that there are alleles that were not discriminated. The methods in this study were intended to give an initial estimate of variation levels. Further analysis with a larger number of individuals and additional allele scoring information such as DNA sequence data will provide a more precise measure of intron 1 variation between individuals and across populations. A change in the scoring method may also alter the reported estimates of Hardy-Weinberg Equilibrium. Repeat regions in introns 1 and 3 of *GH* genes have been identified as significant markers for growth in several

fish species (Almuly *et al.*, 2008; Li *et al.*, 2017; Sawayama & Takagi, 2015b). These regions should be investigated in *C. auratus* as candidate markers for growth in the future.

### Conclusions

In this study, the *caGH* gene was characterised and repeat regions were identified within the 5' flanking region and introns 1–3, and introns 1 and 3 contained large repeat sequences. Our study demonstrates that the *GH* coding regions, amino acid sequences, and predicted secondary structures are highly conserved across perciform species. The extensive sequence variation present between species lies within the non-coding regions of this gene. Previous studies have highlighted the frequent occurrence of repeat regions in gene promoters and 5' proximal introns and their ability to influence gene expression. The findings of this study highlight important regions of *caGH* that may affect *GH* expression. These findings provide important information for future studies of the relationship between these polymorphic regions and growth traits in *C. auratus* and other related species. If these regions are shown to be related to growth performance, they may be useful genomic regions for developing markers that correspond to growth and can be utilised in marker-assisted selective breeding of *C. auratus* and other perciform species.

### 5. Acknowledgements

This project was supported with funding from the VUW University Research Fund and MBIE Endeavour Funds C11X1603 to Enhancing production of New Zealand's seafood sector using accelerated breeding techniques. This work was supported via strategic research funds from the School of Biological Sciences at Victoria University of Wellington. We are grateful for the samples and assistance that was provided by Mark Morrison (NIWA), and in

particular thanks to Parengarenga Incorporation for boat ramp access to Parengarenga Harbour, Ānahera Herbert-Graves (CEO) of Ngāti Kahu Rūnanga and Whiti Awarau of Ngai Takoto for Rangaunu Harbour, and the Integrated Kaipara Harbour Management Group (in particular, Willie Wright) for supporting the Kaipara Harbour sampling.

## 6. Supporting Information

TABLE S1. List of putative transcription factors and regulatory motifs in the *Chrysophrys auratus* growth hormone gene.

FIGURE S1. Similarity between perciform growth hormone amino acid sequences.

FIGURE S2. Hydrophobicity of perciform growth hormone amino acid sequences.

FIGURE S3. Polarity of perciform growth hormone amino acid sequences.

FIGURE S4. Predicted secondary structures of perciform growth hormone proteins.

## 7. Contributions

All authors were responsible for the idea and direction behind the paper. Data was generated by K. Irving and M. Wellenreuther, and analysed by K. Irving. All authors contributed to the preparation of the manuscript.

## 8. Significance Statement

Understanding the genetic mechanisms contributing to fish growth is advantageous for fisheries and aquacultures. Using the economically important teleost fish species *Chrysophrys auratus* from New Zealand, this work characterises a gene sequence for candidate growth gene, growth hormone, and describes the genetic variation in the study

species and closely related perciform fishes. The genetic variation described in this study is relevant to a wide range of fish species worldwide providing direction for future growth studies.

Accepted Article

## 9. References

- Agellon, L. B., Davies, S. L., Chen, T. T., & Powers, D. A. (1988). Structure of a fish (rainbow trout) growth hormone gene and its evolutionary implications. *Proceedings of the National Academy of Sciences*, 85(14), 5136-5140.
- Aljanabi, S. M., & Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, 25(22), 4692-4693.
- Almuly, R., Cavari, B., Ferstman, H., Kolodny, O., & Funkenstein, B. (2000). Genomic structure and sequence of the gilthead seabream (*Sparus aurata*) growth hormone-encoding gene: identification of minisatellite polymorphism in intron I. *Genome*, 43(5), 836-845.
- Almuly, R., Poleg-Danin, Y., Gorshkov, S., Gorshkova, G., Rapoport, B., Soller, M., et al. (2005). Characterization of the 5' flanking region of the growth hormone gene of the marine teleost, gilthead sea bream *Sparus aurata*: analysis of a polymorphic microsatellite in the proximal promoter. *Fisheries Science*, 71(3), 479-490.
- Almuly, R., Skopal, T., & Funkenstein, B. (2008). Regulatory regions in the promoter and first intron of *Sparus aurata* growth hormone gene: Repression of gene activity by a polymorphic minisatellite. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, 3(1), 43-50.
- Annala, J., Sullivan, K., O'Brien, C., & Iball, S. (2000). Report from the Fisheries Assessment Plenary, May 1998: stock assessment and yield estimates. Available from the NIWA Library, Wellington, New Zealand.

- Arango, J., Echeverri, J., & López, A. (2014). Association between a polymorphism in intron 3 of the bovine growth hormone gene and growth traits in Holstein heifers in Antioquia. *Genetics and Molecular Research, 13*, 6191-6199.
- Backeljauw, P., & Hwa, V. (2016). Growth hormone physiology. In L. E. Cohen (Ed.), *Growth hormone deficiency* (pp. 7-20): Springer.
- Basurco, B., Lovatelli, A., & García, B. (2011). Current status of Sparidae aquaculture. *Sparidae: biology and aquaculture of Gilthead Sea Bream and other species*, 1-50.
- Bernal-Ramírez, J., Adcock, G., Hauser, L., Carvalho, G., & Smith, P. (2003). Temporal stability of genetic population structure in the New Zealand snapper, *Pagrus auratus*, and relationship to coastal currents. *Marine Biology, 142*(3), 567-574.
- Betancur-R, R., Wiley, E. O., Arratia, G., Acero, A., Bailly, N., Miya, M., et al. (2017). Phylogenetic classification of bony fishes. *BMC evolutionary biology, 17*(1), 162.
- Butler, A. A., & Roith, D. L. (2001). Control of growth by the somatotropic axis: growth hormone and the insulin-like growth factors have related and independent roles. *Annual Review of Physiology, 63*(1), 141-164.
- Castillo-Davis, C. I., Mekhedov, S. L., Hartl, D. L., Koonin, E. V., & Kondrashov, F. A. (2002). Selection for short introns in highly expressed genes. *Nature Genetics, 31*(4), 415-418.
- Chen, Y., Wang, Y., He, S., & Zhu, Z. (2004). Cloning and sequencing of the growth hormone gene of large yellow croaker and its phylogenetic significance. *Biochemical Genetics, 42*(9-10), 365-375.
- Choi, Y., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and indels. *PLoS One, 7*(10), e46688.
- Chorev, M., & Carmel, L. (2012). The function of introns. *Frontiers in Genetics, 3*.

- Colman, J. A. (2010). Food of snapper, *Chrysophrys auratus* (forster), in the Hauraki Gulf, New Zealand. *New Zealand Journal of Marine and Freshwater Research*, 6(3), 221-239.
- De-Santis, C., & Jerry, D. R. (2007). Candidate growth genes in finfish—Where should we be looking? *Aquaculture*, 272(1), 22-38.
- Fu, L., Tomita, A., Wang, H., Buchholz, D. R., & Shi, Y.-B. (2006). Transcriptional regulation of the *Xenopus laevis* stromelysin-3 gene by thyroid hormone is mediated by a DNA element in the first intron. *Journal of Biological Chemistry*, 281(25), 16870-16878.
- Fuentes, E. N., Valdés, J. A., Molina, A., & Björnsson, B. T. (2013). Regulation of skeletal muscle growth in fish by the growth hormone–insulin-like growth factor system. *General and Comparative Endocrinology*, 192, 136-148.
- Goodman, H. M. (1993). Growth hormone and metabolism. *The Endocrinology of Growth, Development, and Metabolism in Vertebrates*, 93-115.
- Goudet, J. (2001). FSTAT; a program to estimate and test gene diversities and fixation indices version 2.9. 3. <http://www.unil.ch/izea/software/fstat.html>.
- Jaser, S. K. K., Dias, M. A. D., Lago, A. d. A., Reis Neto, R. V., & Hilsdorf, A. W. S. (2017). Single nucleotide polymorphisms in the growth hormone gene of *Oreochromis niloticus* and their association with growth performance. *Aquaculture Research*, 48(12), 5835-5845.
- Jones, D. D. (2005). Triplet nucleotide removal at random positions in a target gene: the tolerance of TEM-1  $\beta$ -lactamase to an amino acid deletion. *Nucleic Acids Research*, 33(9), e80-e80.
- Kailola, P. J. (1993). *Australian fisheries resources*: Bureau of Resource Sciences, Dept. of Primary Industries and Energy; Fisheries Research and Development Corp.

- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647-1649.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular evolution*, 16(2), 111-120.
- Knorr, C., Moser, G., Müller, E., & Geldermann, H. (1997). Associations of GH gene variants with performance traits in F2 generations of European wild boar, Pietrain and Meishan pigs. *Animal Genetics*, 28(2), 124-128.
- Kuhnlein, U., Ni, L., Zadworny, D., & Fairfull, W. (1997). DNA polymorphisms in the chicken growth hormone gene: response to selection for disease resistance and association with egg production. *Animal Genetics*, 28(2), 116-123.
- Kumar, R. P., Senthilkumar, R., Singh, V., & Mishra, R. K. (2010). Repeat performance: how do genome packaging and regulation depend on simple sequence repeats? *Bioessays*, 32(2), 165-174.
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, msw054.
- Li, M.-J., Liu, W.-S., Luo, W., Zhang, X.-Q., Zhu, W.-L., Wang, J., et al. (2017). Polymorphisms and their association with growth traits in the growth hormone gene of yellow catfish, *Pelteobagrus fulvidraco*. *Aquaculture*, 469, 117-123.
- Li, Y.-C., Korol, A. B., Fahima, T., & Nevo, E. (2004). Microsatellites within genes: structure, function, and evolution. *Molecular Biology and Evolution*, 21(6), 991-1007.
- Li, Y. C., Korol, A. B., Fahima, T., Beiles, A., & Nevo, E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology*, 11(12), 2453-2465.

- Lomaev, D., Mikhailova, A., Erokhin, M., Shaposhnikov, A. V., Moresco, J. J., Blokhina, T., et al. (2017). The GAGA factor regulatory network: Identification of GAGA factor associated proteins. *PLoS One*, *12*(3), e0173602.
- Mommsen, T. P. (2001). Paradigms of growth in fish. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *129*(2-3), 207-219.
- Moriyama, S., Ayson, F. G., & Kawauchi, H. (2000). Growth regulation by insulin-like growth factor-I in fish. *Bioscience, Biotechnology, and Biochemistry*, *64*(8), 1553-1562.
- Moriyama, S., Oda, M., Takahashi, A., Sower, S. A., & Kawauchi, H. (2006). Genomic structure of the sea lamprey growth hormone-encoding gene. *General and Comparative Endocrinology*, *148*(1), 33-40.
- Nelson, J. (2006). Fishes of the world 4th edition. *Hoboken: John Wiley & Sons, Inc.*
- Nie, Q., Sun, B., Zhang, D., Luo, C., Ishag, N., Lei, M., et al. (2005). High diversity of the chicken growth hormone gene and effects on growth and carcass traits. *Journal of Heredity*, *96*(6), 698-703.
- Ohshima, K., Montermini, L., Wells, R. D., & Pandolfo, M. (1998). Inhibitory Effects of Expanded GAA·TTC Triplet Repeats from Intron I of the Friedreich Ataxia Gene on Transcription and Replication in Vivo. *Journal of Biological Chemistry*, *273*(23), 14588-14595.
- Pál, C., Papp, B., & Lercher, M. J. (2006). An integrated view of protein evolution. *Nature Reviews Genetics*, *7*(5), 337-348.
- Parsons, D. M., Sim-Smith, C. J., Cryer, M., Francis, M. P., Hartill, B., Jones, E. G., et al. (2014). Snapper (*Chrysophrys auratus*): a review of life history and key vulnerabilities in New Zealand. *New Zealand Journal of Marine and Freshwater Research*, *48*(2), 256-283.

- Paul, L. (1977). *The Commercial Fishery for Snapper Chrysophrys Auratus (Forster), in the Auckland Region, New Zealand, from 1900 to 1971*: Fisheries Research Division, New Zealand Ministry of Agriculture and Fisheries.
- Paulin, C. D. (1990). *Pagrus auratus*, a new combination for the species known as “snapper” in Australasian waters (Pisces: Sparidae). *New Zealand Journal of Marine and Freshwater Research*, 24(2), 259-265.
- Peakall, R., & Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Resources*, 6(1), 288-295.
- Perez-Sanchez, J. (2000). The involvement of growth hormone in growth regulation, energy homeostasis and immune function in the gilthead sea bream (*Sparus aurata*): a short review. *Fish Physiology and Biochemistry*, 22(2), 135-144.
- Pérez-Sánchez, J., Calduch-Giner, J., Mingarro, M., de Celis, S. V.-R., Gómez-Requeni, P., Saera-Vila, A., et al. (2002). Overview of fish growth hormone family. New insights in genomic organization and heterogeneity of growth hormone receptors. *Fish Physiology and Biochemistry*, 27(3-4), 243-258.
- Plateroti, M., Kress, E., Mori, J. I., & Samarut, J. (2006). Thyroid hormone receptor  $\alpha 1$  directly controls transcription of the  $\beta$ -catenin gene in intestinal epithelial cells. *Molecular and Cellular Biology*, 26(8), 3204-3214.
- Quééré, N., Guinand, B., Kuhl, H., Reinhardt, R., Bonhomme, F., & Desmarais, E. (2010). Genomic sequences and genetic differentiation at associated tandem repeat markers in growth hormone, somatolactin and insulin-like growth factor-1 genes of the sea bass, *Dicentrarchus labrax*. *Aquatic Living Resources*, 23(3), 285-296.
- Raymond, M. (1995). GENEPOP: Population genetics software for exact tests and ecumenism. Vers. 1.2. *Journal of Heredity*, 86, 248-249.

- Reinecke, M., Björnsson, B. T., Dickhoff, W. W., McCormick, S. D., Navarro, I., Power, D. M., et al. (2005). Growth hormone and insulin-like growth factors in fish: where we are and where to go. *General and Comparative Endocrinology*, *142*(1), 20-24.
- Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite: Elsevier Current Trends.
- Rowe, R., & Goldspink, G. (1969). Muscle fibre growth in five different muscles in both sexes of mice. *Journal of Anatomy*, *104*(Pt 3), 519.
- Sawayama, E., & Takagi, M. (2015a). Isolation and characterization of tandem repeat sequences in the growth hormone gene of the red seabream, *Pagrus major* (Temminck & Schlegel, 1843). *Journal of Applied Ichthyology*, *31*(4), 762-766.
- Sawayama, E., & Takagi, M. (2015b). Parental contribution and growth hormone gene polymorphism associated with growth phenotypes of red sea bream *Pagrus major* in mass production: A case study. *Aquaculture Reports*, *2*, 144-151.
- Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, *18*(2), 233-234.
- Tambasco, D., Paz, C., Tambasco-Studart, M., Pereira, A., Alencar, M., Freitas, A., et al. (2003). Candidate genes for growth traits in beef cattle crosses *Bos taurus* × *Bos indicus*. *Journal of Animal Breeding and Genetics*, *120*(1), 51-56.
- Taylor, M. S., Ponting, C. P., & Copley, R. R. (2004). Occurrence and consequences of coding sequence insertions and deletions in mammalian genomes. *Genome Research*, *14*(4), 555-566.
- Vijayakumar, A., Wu, Y., Sun, H., Li, X., Jeddy, Z., Liu, C., et al. (2012). Targeted loss of GHR signaling in mouse skeletal muscle protects against high-fat diet-induced metabolic deterioration. *Diabetes*, *61*(1), 94-103.

Wei, T., Sun, Y., Zhang, B., Wang, R., & Xu, T. (2014). A mitogenomic perspective on the phylogenetic position of the *Haplogenys* genus (Acanthopterygii: Perciformes) and the evolutionary origin of perciformes. *PLoS One*, 9(7), e103011.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358-1370.

Zhang, D., Shao, Y., Jiang, S., Li, J., & Xu, X. (2009). *Nibeia coibor* growth hormone gene: Its phylogenetic significance, microsatellite variation and expression analysis. *General and Comparative Endocrinology*, 163(3), 233-241.

Species	Authority	Accession number	Family	Order
<i>Chrysophrys auratus</i>	(Forster 1801)	MW972222 (This study)	Sparidae	Perciformes
<i>Pagrus major</i>	(Temminck & Schlegel 1843)	AB904715	Sparidae	Perciformes
<i>Sparus aurata</i>	(Linnaeus 1758)	U48221	Sparidae	Perciformes
<i>Dicentrarchus labrax</i>	(Linnaeus 1758)	GQ918491	Moronidae	Perciformes
<i>Seriola quinqueradiata</i>	(Temminck & Schlegel 1845)	D50368	Carangidae	Perciformes
<i>Siniperca scherzeri</i>	(Steindachner 1892)	EF441623	Percichthyidae	Perciformes
<i>Siniperca chuatsi</i>	(Basilevsky 1855)	EF205280	Percichthyidae	Perciformes
<i>Siniperca kneri</i>	(Garman 1912)	EF205281	Percichthyidae	Perciformes
<i>Lates calcarifer</i>	(Bloch 1790)	LCU16816	Latidae	Perciformes
<i>Epinephelus coioides</i>	(Hamilton 1822)	KR269816	Serranidae	Perciformes
<i>Channa diplogramma</i>	(Day 1865)	GQ214246	Channidae	Perciformes
<i>Channa marulius</i>	(Hamilton 1822)	GQ214245	Channidae	Perciformes
<i>Channa gachua</i>	(Hamilton 1822)	GQ214244	Channidae	Perciformes
<i>Channa punctata</i>	(Bloch 1793)	GQ214243	Channidae	Perciformes
<i>Channa striata</i>	(Bloch 1793)	EF447030	Channidae	Perciformes
<i>Tilapia nilotica</i>	(Linnaeus 1758)	M84774	Cichlidae	Perciformes
<i>Coptodon zillii</i>	(Gervais 1848)	KT387600	Cichlidae	Perciformes
<i>Sarotherodon galilaeus</i>	(Linnaeus 1758)	KT387599	Cichlidae	Perciformes
<i>Tilapia mossambica</i>	(Peters 1852)	AF0338061	Cichlidae	Perciformes
<i>Pseudosciaena crocea</i>	(Richardson 1846)	AY090592	Sciaenidae	Perciformes
<i>Nibea coibor</i>	(Hamilton 1822)	FJ375311	Sciaenidae	Perciformes
<b>Outgroups</b>				
<i>Paralichthys olivaceus</i>	(Temminck & Schlegel 1846)	D29737	Paralichthyidae	Pleuronectiformes
<i>Clarias batrachus</i>	(Linnaeus 1758)	AF416486	Clariidae	Siluriformes
<i>Oncorhynchus keta</i>	(Walbaum, 1792)	L04688	Salmonidae	Salmoniformes

<b>Sample site</b>	<b>N</b>	<b>Na</b>	<b>Ho</b>	<b>He</b>	<b>P-value</b>
Bay of Islands	20	6	0.800	0.678	0.3995
Doubtless Bay	20	3	0.500	0.546	0.8230
Kaipara	20	4	0.650	0.581	0.4244
Parengarenga	20	6	0.700	0.643	0.9863
Rangaunu	19	4	0.632	0.594	1.0000

<b>Locus</b>	<b>P-value</b>
Intron 1	0.9489

## Table and Figure Captions:

TABLE 1. List of species used in the phylogenetic analysis of GH genes in this study. Includes the species name and authority, Genbank accession number, and the taxonomic family and order classifications for each species.

FIGURE 1. Structure of the caGH gene. Dotted line = 5'-flanking region. Black boxes = exons. Lines between boxes = introns. Brackets below figure = size in base pairs. Grey boxes = repeat regions. Brackets above figure = sequence motifs of each repeat region. The intron 1 repeat region is made up of varying combinations and repeats of the following core repeat sequences: 3' - TGACCTGTCTGTCTCTCTCTCTCTGTC - 5'; 3' - TGACCTGTCTGTCTCTC - 5'; 3' - TGACGTGTCTGTCTCTCTCTTTCTCTCTGTC - 5'; 3' - TGACCTGTCTGTCTGTGTCAGTGGTGCTGATGCTGTCGGTGCTGTCTC - 5'. The intron 3 repeat region is made up of varying combinations and repeats of the following core repeat sequences: 3' - CACACAGTACTACATGATACTG - 5'; 3' - CACACAGTACTACATGATACTA - 5'; 3' - CACACGCACTACATGATACTACACAGTACTCATGATACTA - 5'.

FIGURE 2. Phylogenetic tree of Perciform species GH genes inferred by the Maximum Likelihood method based on the Kimura 2-parameter model (Kimura, 1980). Bootstrap values are shown next to branches. Respective taxonomic families are shown to the right of the species names. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

FIGURE 3. GH gene structures for species used in this study for phylogenetic analysis. Species are listed in the same order as the phylogenetic tree in Figure 2. Black boxes = exons. Lines between boxes = introns. Roman numerals = exon number.

TABLE 2. GH intron 1 allelic diversity and heterozygosity of wild samples of *Chrysophrys auratus*. N, sample size; Na, number of alleles; Ho, observed heterozygosity; He, expected heterozygosity. P-value, significance value for exact tests assessing deviation from HWE by sample site and locus in the GH intron 1 locus of wild *Chrysophrys auratus* samples.

### Supporting Information legends:

TABLE S1. List of putative transcription factors and regulatory motifs detected in *caGH* gene using EMBOSS (Rice et al., 2000). Table includes the name of each factor and where it was detected within the gene sequence.

FIGURE S1. Similarity between Perciform GH amino acid sequences. Numbers 1-24 = Species name in same order as phylogenetic tree (Figure 2). Black = more similar and white = less similar. Thin horizontal lines between amino acids = amino acid gap.

FIGURE S2. Hydrophobicity of Perciform GH amino acid sequences. Numbers 1-24 = Species name in same order as phylogenetic tree (Figure 2). Blue = hydrophilic and red = hydrophobic. Thin horizontal lines between amino acids = amino acid gap.

FIGURE S3. Polarity of Perciform GH amino acid sequences. Numbers 1-24 = Species name in same order as phylogenetic tree (Figure 2). Blue = positively charged; red = negatively charged; green = polar; yellow = non-polar. Thin horizontal lines between amino acids = amino acid gap.

FIGURE S4. Predicted secondary structures of Perciform GH proteins. Numbers 1-24 = Species name in same order as phylogenetic tree (Figure 2). Pink cylinders = alpha helices; yellow arrows = beta strands; grey waves = coil; blue arrows = turns. Amino acid sequence similarity in black and white below each structure. Black = more similar and white = less similar. Thin horizontal lines between amino acids = amino acid gap.

(Intron 1 repeat region - See figure legend)

(ACTGTTCTCT)<sub>10</sub>

(TATT)<sub>4</sub>

(AC)<sub>15</sub>

(CCCTGA)<sub>6</sub>

(GT)<sub>10</sub>

(Intron 3 repeat region - See figure legend)

(ATTG)<sub>4n</sub> (ATTG)<sub>6</sub>

(567)

(410)

(713)

(3384)

(85)

(79)

(71)

(134)

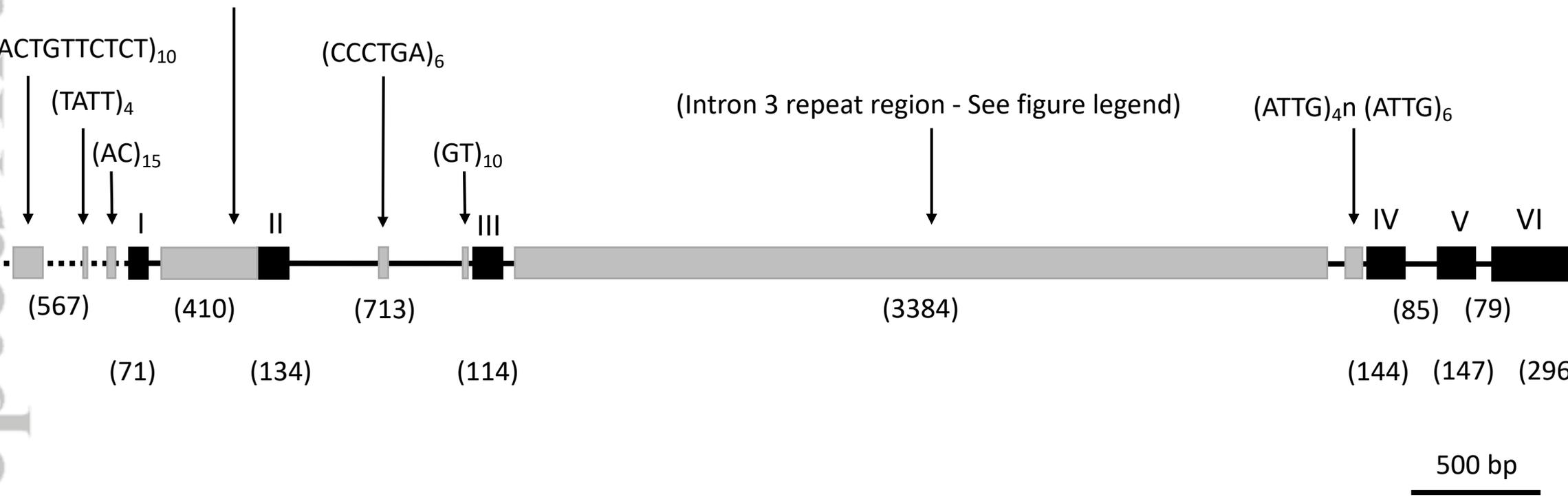
(114)

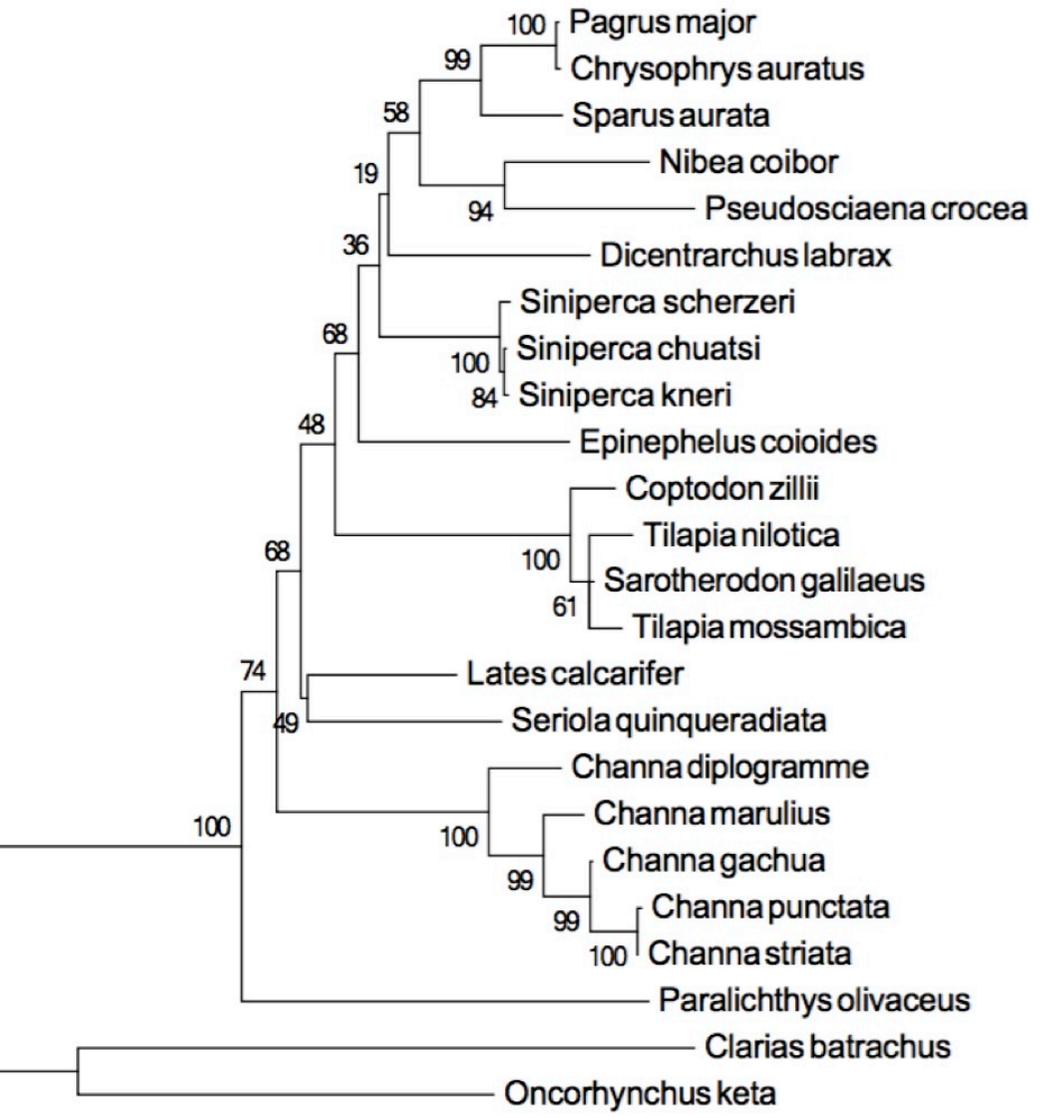
(144)

(147)

(296)

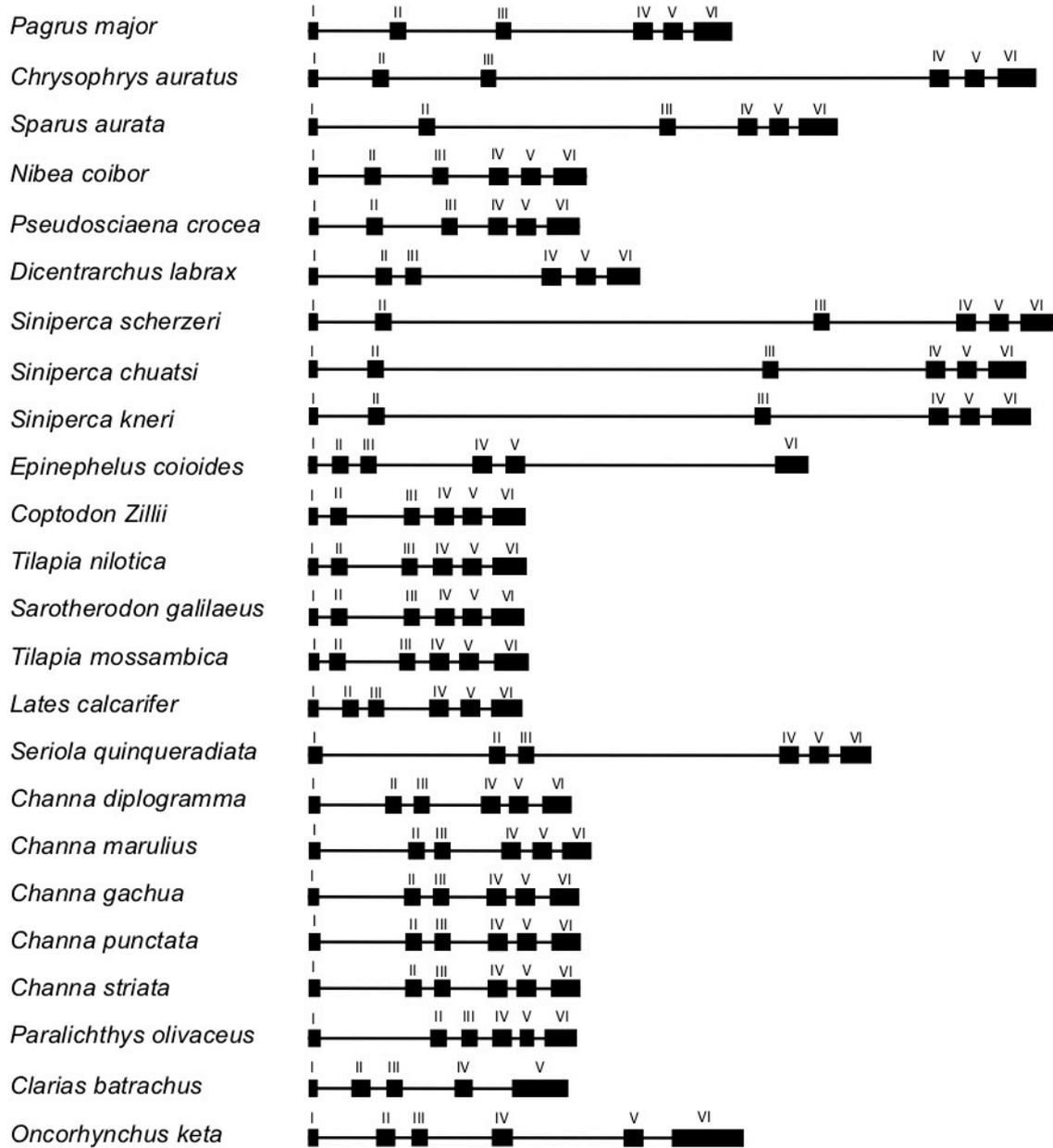
500 bp





0.10

- Sparidae
- Sciaenidae
- Moronidae
- Percichthyidae
- Serranidae
- Cichlidae
- Latidae
- Carangidae
- Channidae
- Paralichthyidae
- Siluriformes
- Salmoniformes



1 kb

**Significance Statement**

Understanding the genetic mechanisms contributing to fish growth is advantageous for fisheries and aquacultures. Using the economically important teleost fish species *Chrysophrys auratus* from New Zealand, this work characterises a gene sequence for candidate growth gene, growth hormone, and describes the genetic variation in the study species and closely related perciform fishes. The genetic variation described in this study is relevant to a wide range of fish species worldwide providing direction for future growth studies.