An investigation into growth-related genes in the Australasian snapper, *Chrysophrys auratus*

Kate Rose Irving

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

Master of Science

in Ecology and Biodiversity

Victoria University of Wellington

2018

Abstract

Growth is a complex quantitative trait that is controlled by a variety of genetic and environmental factors. Due to its significance in animal breeding programmes, growth is a commonly studied trait in agriculture and aquaculture species. The Australasian snapper, *Chrysophrys auratus*, supports significant commercial and recreational fisheries in New Zealand and has the potential to be developed as a new aquaculture species. However, the relatively slow growth rate of *C. auratus* is a constraint and little is known of the specific regulation of growth in this species. The overall aim of this thesis research was to use genome sequence data and transcriptomics to investigate the loci that influence growth rate of *C. auratus*.

In Chapter Two, the *C. auratus* Growth Hormone (GH) gene was identified in the reference genome and the structure and polymorphisms were characterised using re-sequenced data. The *GH* gene was approximately 5,577 bp in length and was comprised 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 conserved non-coding regions and highly variable non-coding regions. The amino acid sequences and putative secondary structures were also largely conserved across this order. In Chapter Three, the genetic variation of two large intronic repeat regions were assessed in wild *C. auratus* populations and shown to be polymorphic. The intron 1 locus was then assessed in slow- and fast-growing *C. auratus* for associations with growth rate. No significant differences were detected in the variation between groups; however, trends seen in the results corroborated other studies of an association between shorter introns and increased gene expression. Further investigation with a larger sample size is needed. A high level of heterozygosity was detected in all populations used in this study and may be due to negative selection acting on one allele (485). In Chapter Four, gene expression data was compared between *C. auratus* at high (21 °C) and low (13 °C) temperatures to investigate how the gene regulation of growth is influenced by temperature. The high temperature treatment (HTT) was characterized by a large number of differentially expressed genes associated with biosynthesis, skeletal muscle components, and catalytic activity while the low temperature treatment (LTT) had an upregulation of genes associated with important degradation pathways. The results of this study also suggest the action of negative feedback on growth regulation in the HTT, which may be a result of chronic heat stress.

This thesis research represents one of the first studies to explore the genetic regulation of growth in *C. auratus* and makes a significant contribution to the field of research into growth, not only in *C. auratus*, but also other fish species. The findings presented in this thesis may be applied to a selective breeding programme of *C. auratus* that aimed to increase the growth rate, and consequently, improve its economic viability as a commercial aquaculture species in New Zealand.

Acknowledgements

I would like to start by thanking my supervisor, Dr Peter Ritchie, for your support throughout my project. Your door was always open if I needed it and your continued encouragement, especially throughout the more challenging times, was much appreciated.

I would also like to thank Maren Wellenreuther for your significant input to my project. I appreciate the time you invested in me and in my work. Thank you for supplying samples and other resources for my research. A big thank you also to Jérémy Le Luyer for supplying the data for Chapter Four of my thesis research and for answering all of my questions.

To Ange, Jana, and Balam, thank you for all of your help in the lab. The skills you taught me and the wisdom you passed on are much appreciated. A special thank you to Alex, who not only helped me a great deal in lab, but also endured endless questioning from me. Thank you for putting up with me and for reading over my work.

Thank you to Jean Goodman Dance studios for offering some stress relief and space from the stresses of university life. A big thank you to Nick, James, and Eleanor at Wellington Diving Club for your support throughout my project, for putting up with me when I was stressed, and most of all, for providing me with a great deal of entertainment.

To my amazing friends Imogen, Rachel, Susan, and Esther, I cannot express how much I appreciate all of the support, distraction, and encouragement you have given me. I could not have done this without you.

I am also incredibly grateful to have had the support of Roald, Jess, and Josh throughout the last two years. I don't know how I would have survived this without you all. Thank you for making everything bearable, for answering questions, and for reading my work. But most of all, thank you for making me laugh through good times and bad.

And to my family, immediate and extended, thank you for your enduring support. Thank you, Debbie, for reading my work and offering academic advice throughout my research. Thank you to Josh and Kaitlyn for keeping in touch and for being so encouraging. Thank you, Ben, Kathryn, and Angus, for all of your visits and for having me to stay. Thank you for all of

your support and for always being there for me when I needed it. It is fair to say that I would never have made it this far without my incredible parents, Tricia and Adrian. Thank you so much for supporting me in every way possible. I am so grateful for all that you have done for me.

Finally, I would like to thank Victoria University of Wellington for providing me with a Victoria Masters by Thesis Fees scholarship to complete my research.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

4EBP1	EIF4E (Eukaryotic Translation Initiation Factor 4E) Binding Protein
ACTB	Actin Beta
ACTB2	Actin Beta 2
ACTC	Actin, Alpha, Cardiac Muscle 1
ACTS	Actin, Alpha 1, Skeletal Muscle
AKT	Akt Serine/Threonine Kinase 1
ATF3	Activating Transcription Factor 3
ATF4	Activating Transcription Factor 4
ATF5	Activating Transcription Factor 5
ATF6	Activating Transcription Factor 6
BCL2	B-Cell Lymphoma 2
BSA	Bovine Serum Albumin
caGH	*Chrysophrys auratus* Growth Hormone
CDN1A	Cyclin Dependent Kinase Inhibitor 1A
CHAC1	ChaC Glutathione Specific Gamma
CPM	Counts Per Million
CRE	Cyclic-AMP Response Element
DEG	Differentially Expressed Gene
ER	Endoplasmic Reticulum
ERE	Estrogen Response Element
ERK	Extracellular Signal-Regulated Kinase
FBX32/MAFBX	F-Box Protein 32/Muscle Atrophy F-Box Protein
FDR	False Discovery Rate
FGFR3	Fibroblast Growth Factor Receptor 3
FOXO	Forkhead Box O
FOXO1	Forkhead Box O1
FOXO3	Forkhead Box O3
GEI	Genotype-Environment Interaction
GH	Growth Hormone
GHBP	Growth Hormone Binding Protein
GHR	Growth Hormone Receptor
GHRH	Growth Hormone Releasing Hormone
GLM	Generalised Linear Model
GO	Gene Ontology
GRE	Glucocorticoid Response Element
GWAS	Genome-Wide Association Study
HBB	Haemoglobin Subunit Beta
HBAA	Haemoglobin Subunit Aplha 1
HBAB	Haemoglobin Subunit Alpha 2
HSP	Heat Shock Protein
HSP70	Heat Shock Protein 70

HSP90	Heat Shock Protein 90
HSP90A	Heat Shock Protein 90A
HTT	High Temperature Treatment
HWE	Hardy-Weinberg Equilibrium
IGF	Insulin-like Growth Factor
IGF1	Insulin-like Growth Factor 1
IGF2	Insulin-like Growth Factor 2
IGFBP	Insulin-like Growth Factor Binding Protein
IGFR	Insulin-like Growth Factor Receptor
IP3	Inositol-Trisphosphate 3
IP3KA	Inositol-Trisphosphate 3-Kinase A
KLF15	Kruppel like factor 15
KS6B1	70-kDa Ribosomal Protein S6 Kinase
LTT	Low Temperature Treatment
M4K3	Mitogen-Activated Protein Kinase Kinase Kinase Kinase 3
MAPK	Mitogen-Activated Protein Kinase
MAPK3	Mitogen-Activated Protein Kinase 3
MK13	Mitogen-Activated Protein Kinase 13
MRF	Myogenic Regulatory Factor
mRNA	Messenger RNA (Ribonucleic Acid)
MSTN	Myostatin
MTOR	Mechanistic Target of Rapamycin
MURF1/TRI63	Muscle Specific Ring Finger Protein 1/Tripartite Motif Containing 63
MYF5	Myogenic Factor 5
MYL9	Myosin Light Chain 9
P3C2A	Phosphatidylinositol-4-Phosphate 3-Kinase Catalytic Subunit Type 2 Alpha
P3C2B	Phosphatidylinositol-4-Phosphate 3-Kinase Catalytic Subunit Type 2 Beta
PCNA	Proliferating Cell Nuclear Antigen
PCR	Polymerase Chain Reaction
PFR	Plant and Food Research
PHLA3	Plecktsrin Homology Like Domain Family A Member 3
PI3K	Phosphoinositide 3-Kinase
PI3R4	Phosphoinositide-3-Kinase Regulatory Subunit 4
Pit-1a	Putative Pituitary-Specific Transcription Factor
PLC	Phospholipase C
pmaGH	*Pagrus major* Growth Hormone
QTL	Quantitative Trait Locus
SDS	Sodium Dodecyl Sulphate
SESN2	Sestrin 2
SMAD	Smad Protein

SNP	Single Nucleotide Polymorphism
SOX4	SRY (Sex Determining Region)-Box 4
TPM	Transcripts Per kilobase Million
TRAF2	TNF (Tumour Necrosis Factor) Receptor Associated Factor 2
TRE	Thyroid Response Element
UPR	Unfolded Protein Response
VNTR	Variable Number Tandem Repeat
WGD	Whole Genome Duplication

Chapter One: General introduction

1.1. Background

The vast majority of phenotypic variation is thought to be a result of quantitative genetic variation where individual phenotypic traits are determined by the cumulative effects of multiple regions of a genome and environmental factors (Mackay, 2001). Growth is a commonly studied quantitative trait due to its significance in animal breeding programmes such as for agriculture and aquaculture. Consequently, the control of growth traits and related pathways are relatively well described in vertebrate species used as livestock. Vertebrate growth is a complex and highly coordinated process. Growth traits can be influenced by a multitude of environmental factors including temperature, photoperiod, food availability, and disease (Johnston, 2006; Reinecke, 2010) and also by underlying genetic variation (Mullis, 2005). The primary system regulating somatic growth in vertebrates is the growth hormone (GH) – insulin-like growth factor (IGF) axis and is highly conserved across vertebrate species (Moriyama, Ayson, & Kawauchi, 2000).

1.2. Vertebrate growth

A basic summary of the GH-IGF axis is presented in Figure 1.1. GH functions under the control of the hypothalamic hormones GH-releasing hormone (GHRH) and GH-inhibiting hormone, also known as somatostatin. When GHRH is released from the hypothalamus it travels to the pituitary and binds receptors promoting the production and secretion of GH. Conversely, the release of hypothalamic somatostatin inhibits the production and release of GH from the pituitary (Moriyama et al., 2000). Once released from the pituitary, GH travels through the bloodstream and binds GH receptors (GHR) on cells in target tissues. While in circulation, GH is typically bound by GH-binding proteins (GHBP) which are able to transport and regulate the bioavailability and half-life of GH (Edens & Talamantes, 1998; Moriyama et al., 2000). GH can bind directly to the cells in target 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 IGFs (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). Similarly to GH, IGFs are typically bound by binding proteins (IGFBP) that regulate their bioactivity and half-lives (Kajimura & Duan, 2007). As stated, GH can also bind directly to target tissues promoting the synthesis and release of IGFs.

Locally produced IGFs may act on target tissues in an autocrine or paracrine fashion. 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). IGF stimulates the action of these processes through binding to IGF receptors (IGFR). The GH-IGF axis also involves a negative feedback loop in which higher levels of circulating GH and IGFs lead to an increase in the release of somatostatin from the hypothalamus and consequently, reduced secretion of GH from the pituitary (Backeljauw & Hwa, 2016). Furthermore, it was recently discovered that GH, traditionally believed to be produced only in the pituitary, can be expressed in peripheral tissues which may suggest an additional pathway to growth control (Harvey, Martínez-Moreno, Luna, & Arámburo, 2015). The GH-IGF axis is complex and there is still a large amount that is unknown about its intricate control of growth.

Figure 1.1. Summary of basic GH-IGF axis. Transmembrane receptors mediate action of hormones, growth hormone (GH), and insulin-like growth factor (IGF). GH and IGF circulate in the blood bound to specific binding proteins. Diagram sourced from Moriyama et al. (2000).

1.3. Regulation of skeletal muscle mass in teleost fish

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, Valdés, Molina, & Björnsson, 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). Furthermore, the complexity of the GH-IGF axis in fish is increased due to a whole genome duplication (WGD) event that is thought to have occurred in the common ancestor of all teleosts (Glasauer & Neuhauss, 2014). This has resulted in both ortholog and paralog genes being retained in extant teleost species (Amores et al., 1998; Jaillon et al., 2004). The effects of the WGD on systems such as the GH-IGF axis are largely unknown; however, multiple copies of genes involved in this axis have been detected in some species such as salmon (Lappin, Shaw, & Macqueen, 2016) and tilapia (Ber & Daniel, 1993).

Myogenesis, or skeletal muscle formation, is under the specific control of myogenic regulatory factors (MRFs). These MRFs control specification, activation, and differentiation of myogenic cells (Fuentes et al., 2013). The maintenance of already formed muscle fibres is a comparatively complex and dynamic process regulated by a balance between protein synthesis and protein degradation. Muscle growth occurs when protein synthesis outweighs protein degradation while muscle atrophy occurs when protein degradation exceeds protein synthesis (Johnston, Bower, & Macqueen, 2011; Schiaffino, Dyar, Ciciliot, Blaauw, & Sandri, 2013). A basic summary of the known control of skeletal muscle growth by GH in fish is shown in Figure 1.2. There are two major signalling pathways regulating protein synthesis in the muscle tissue of teleosts. The IGF/Akt serine threonine kinase 1(AKT)/ mechanistic target of rapamycin (MTOR) pathway exerts positive regulation of protein synthesis while the myostatin (MSTN)/SMAD pathway negatively regulates protein synthesis and growth (Figure 1.2). AKT and MTOR are the major effectors of IGF signalling, and MSTN is a recognised negative regulator of growth that is produced in skeletal muscle (Lee & McPherron, 2001). Protein degradation and atrophy is also regulated through two major signalling pathways, the ubiquitin-proteasomal pathway, and the autophagic-lysosomal pathway (Schiaffino et al., 2013).

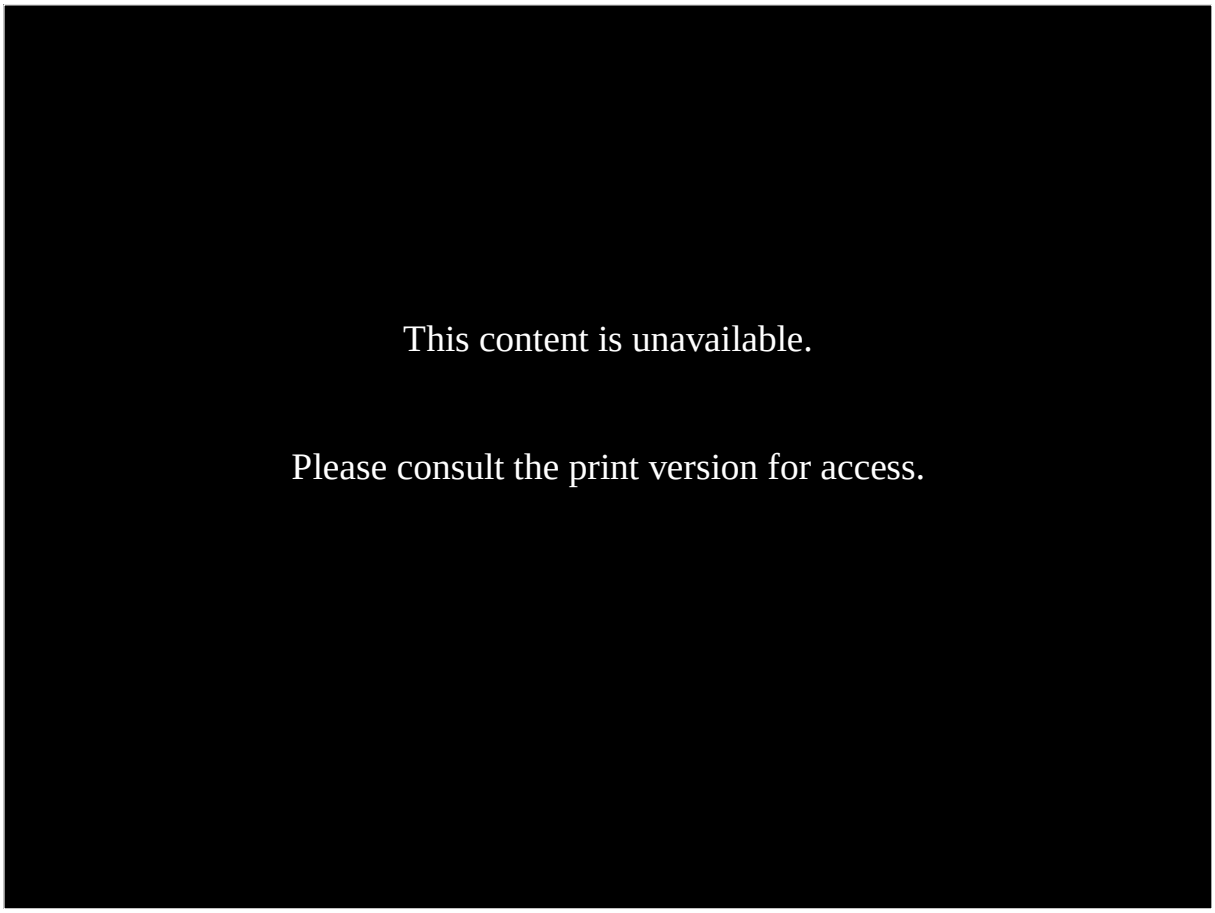

Figure 1.2. Regulation of skeletal muscle growth by the GH system in teleost fish. GH stimulates muscle growth in fish through proliferation, hypertrophy and/or hyperplasia. These processes are controlled by the regulation of several genes in the MSTN, atrophy, GH, IGF systems, and genes for MRFs. GH may control the expression of *IGF1* via the JAK2/STAT5 pathway. The effects of *GH* on *IGF1* expression in fish muscle appear to be related to the presence of higher contents of tGHR1 (truncated form GH receptor) than flGHR1 (full length GH receptor), suggesting impairment of GH signalling pathways (e.g. JAK2/STAT5). White symbols (i.e. molecules and boxes) and question marks (?) represent unstudied molecules or biological processes in fish skeletal muscle. Absent molecules are indicated in black- dashed symbols. Symbols (i.e. molecules and boxes) coloured in blue and green denote positive and negative modulators of fish muscle growth, respectively. Figure sourced from Fuentes et al. (2013).

These protein degradation pathways are under the control of specific transcription factors and lead to the transcription of atrophy genes, or atrogenes. The ubiquitin-proteasomal pathway involves the marking of proteins and other cell constituents with ubiquitin, consequently targeting them for degradation in the proteasome (Glickman & Ciechanover, 2002). The autophagic-lysosomal pathway is the system by which cytoplasmic materials are delivered to and degraded in the lysosome (Mizushima & Komatsu, 2011).

As described, GH typically affects growth in target tissues through the indirect action of liver-produced IGFs. IGFs have two major intracellular signalling pathways, the phosphoinositide 3-kinase (PI3K)/AKT pathway and the mitogen-activated protein kinase

(MAPK)/Extracellular signal-regulated kinase (ERK) pathway (Glass, 2003). Induction of the PI3K/AKT pathway leads to protein synthesis (Rommel et al., 2001), myoblast differentiation (Coolican, Samuel, Ewton, McWade, & Florini, 1997), and muscle hypertrophy (Bodine, Stitt, et al., 2001). This pathway regulates these processes by activation of MTOR which, when activated, regulates the 70-kDa ribosomal protein S6 kinase (KS6B1 or P70S6K), and the EIF4E binding protein, (4EBP1) (Glass, 2003). The second major IGF signalling pathway, MAPK/ERK, stimulates cell proliferation (Coolican et al., 1997) and terminal differentiation (Li & Johnson, 2006) in muscles. IGF signalling promotes growth while simultaneously inhibiting protein degradation. AKT phosphorylates forkhead box O (FOXO) proteins which are specific transcription factors that promote the transcription of atrogenes and lead to the induction of protein degradation pathways. Phosphorylation of FOXO prevents its entry to the nucleus, blocking the transcription of atrogenes and preventing protein degradation (Calnan & Brunet, 2008). These major IGF signalling pathways are summarised in Figure 1.3.

The pathways described here make up the major signalling pathways of the GH-IGF axis. However, as stated, the regulation of growth is exceedingly complex and may be influenced by many other pathways and stimuli. For example, it is known that growth and the GH-IGF axis are affected by environmental factors such as nutrition, temperature, photoperiod, and behaviour (Björnsson et al., 2002; Gabillard et al., 2005). Environmental factors are capable of enhancing or repressing the expression of genes underlying different growth traits (Danzmann, Kocmarek, Norman, Rexroad, & Palti, 2016). As poikilotherms, fish are particularly sensitive to temperature which can affect virtually all biological processes in fish including growth (Hochachka & Somero, 2002). Generally, warmer temperatures are associated with increased growth; though, responses to temperature differ between species. Increased temperatures are typically associated with an increase in plasma GH and IGFs in fish suggesting that the GH-IGF axis plays a role in mediating the effects of temperature on growth (Gabillard et al., 2005). Due to the complexity of this system and the myriad of environmental factors able to influence it, the specific effects of temperature on the growth rate of fish are still largely undefined.

Figure 1.3. Modulation of skeletal muscle growth by the IGF system in teleost fish. Muscle derived IGFBPs may control IGF action in fish muscle; however, it is unknown which IGFBPs are positive or negative modulators of IGFs actions. IGFs stimulate muscle hypertrophy; however, it is unknown whether muscle-derived (mdIGF1) or liver-derived IGF1 (ldIGF1) is the main promoter of muscle growth. Binding of IGFs to their receptors triggers the activation of MAPK/ERK and PI3K/AKT/TOR signalling pathways. These pathways stimulate the transcription of genes involved in the IGF system while, in parallel, IGF also inhibits muscle atrophy via the PI3K/AKT/TOR and PI3K/AKT/FOXO pathways. Amino acids can also stimulate the transcription of *IGF1*, likely through TOR. TOR can also stimulate the activation of P70S6K and 4EBP1 leading to hypertrophy. IGFs affect the gene expression of *MRFs* and components of the MSTN system; however, the signalling pathways involved in this are unknown. White-symbols (i.e. molecules and boxes) and question marks (?) represent unstudied molecules or biological processes in fish skeletal muscle. Symbols (i.e. molecules and boxes) coloured in blue and green denote positive and negative modulators of fish muscle growth, respectively. Figure sourced from Fuentes et al. (2013).

1.4. Growth and selective breeding

Understanding growth is of great significance for animal breeding programmes aiming to increase production efficiency. Improving growth rate reduces the time required for raising animals and allows the animal product to reach market earlier, hence lowering the cost per unit (Gjedrem, Robinson, & Rye, 2012). Moreover, improving the production efficiency of animal breeding programmes is becoming increasingly important as demand for protein for human consumption rises with increased global population, and as efforts are made to reduce the environmental impacts of animal breeding programmes (Bostock et al., 2010; Diouf, 2009). Selective breeding involves identifying variants associated with a desirable trait and breeding

them in such a way as to increase the frequency of variants with the desired trait in future generations. Though selective breeding for genetic improvement in plants and livestock production has a long history, the genetic improvement of aquaculture stocks through selective breeding lags behind that of terrestrial agriculture species (Bostock et al., 2010; Gjedrem & Robinson, 2014). While most terrestrial agriculture has been subject to genetic improvement, less than 10% of global aquaculture production in 2012 was based on genetically improved stocks (Gjedrem et al., 2012). A lack of genetic resources and poor hatchery control has led to the considerable reduction in the performance of many farmed species through inbreeding, genetic drift, and uncontrolled hybridization (Bostock et al., 2010). Because of this decreased performance and reduced viability, many stocks routinely require supplementation from wild populations (Bostock et al., 2010; McAndrew & Napier, 2011). Aquaculture is the fastest growing food production industry with fin-fish making up the largest proportion of aquaculture species and 73.8 million tonnes harvested in 2014 (FAO, 2016). In order to keep up with the demand of this growing industry, breeding programmes must start to employ genetic tools to control levels of inbreeding and the selection of economically relevant traits, no longer depending on restocking from wild populations. These objectives can be achieved through genetic improvement and the application of selective breeding techniques in aquaculture fish stocks (Gjedrem et al., 2012).

Fish have a high potential for genetic improvement. For example, approximately 17% of consumed energy is used for biomass production in fish, while homeotherms such as terrestrial farm animals use only 2% (Smith, 1992). Fish also exhibit growth rates four to five times higher than that of farm animals (Gjedrem & Baranski, 2010) and have a favourably high fecundity (Gjedrem & Robinson, 2014). On average, the genetic gain of traditional selective breeding on fish body weight is 12.7% per generation (Gjedrem & Robinson, 2014). It is possible, with this level of genetic gain, to double growth rate in just six generations, as shown in Nile tilapia (Bentsen, Eknath, Rye, Thodesen, & Gjerde, 2005) and Atlantic salmon (Thodesen, Grisdale-Helland, Helland, & Gjerde, 1999). Selection for growth rate has also been associated with correlated responses in feed conversion rate (Thodesen et al., 1999), and disease resistance (Overturf, LaPatra, Towner, Campbell, & Narum, 2010; Rye, LilleviK, & Gjerde, 1990) which are also commonly selected traits in selective breeding programmes. The collective selection of growth rate, feed conversion efficiency, and disease resistance (survival) has been shown to reduce space, water, and in particular, feed requirements (Gjedrem et al., 2012). Therefore, genetically improved stocks could expect more fish to reach market size in less time while also

experiencing reduced production costs. This increase in efficiency can therefore result in large economic returns, and as genetic improvements are cumulative over generations, these returns could be expected to increase over time (Gjedrem et al., 2012). These have often been the key objectives of selective breeding programmes, leading to an increasing number of studies that investigate how growth is regulated fish.

1.5. Genetic marker technologies and selective breeding for growth

Traditional selective breeding techniques focus on selecting males and females with a phenotype that, when mated, produce offspring with the desired phenotype. Limitations of traditional breeding techniques relate to the ability to routinely phenotype selection-candidates so that appropriate selection decisions can be made early in life (Dekkers, 2012). Moreover, some traits of interest are only able to be measured late in life, can only be recorded on one sex, require animals to be sacrificed, or expose animals to conditions that would hinder the ability to market their germplasm (Dekkers, 2012). Selection of some traits may also result in the undesired selection or loss of other correlated traits (Williams, 2005). These limitations can be reduced using genetic marker technologies which assist in the screening and selection of offspring at an early age that are predicted to have increased production performance later in life (Dekkers, 2012; Williams, 2005). These techniques typically involve the identification of genetic markers that appear to be associated with certain desired traits such as growth. At present, DNA markers applied to selective breeding programmes typically include variable number tandem repeats (VNTRs) or single nucleotide polymorphisms (SNPs) (Liu & Cordes, 2004; McAndrew & Napier, 2011). VNTRs consist of short sequences that are tandemly repeated on a DNA strand. At a VNTR locus, the number of tandem repeats can vary considerably between individuals and populations leading to high levels of polymorphism. In contrast, SNPs are changes in a single base in the DNA sequence. Molecular markers are classified as either type I, markers associated with genes of known function, or type II, markers associated with genomic segments of unknown function (Liu & Cordes, 2004; McAndrew & Napier, 2011).

A focus of research into growth has been centred around identifying type I markers associated with candidate genes for growth. A number of candidate genes for growth have been reported for fish with particular emphasis on *GH*, *IGF1*, *IGF2*, and *MSTN*. Both VNTRs and SNPs associated with growth traits have been identified in fish for *GH* (Almuly, Skopal, &

Funkenstein, 2008; Jaser, Dias, Lago, Reis Neto, & Hilsdorf, 2017) and *IGF2* (Juhua et al., 2010; Li et al., 2012), while only SNPs have been associated with growth traits in *IGF1* (Feng, Yu, & Tong, 2014) and *MSTN* (Wang et al., 2014). Genome-wide scans to look for statistical associations between a large number of genetic loci and desirable production traits can now be carried out using genome-wide association (GWAS), and quantitative trait loci (QTL) studies. Genome-wide studies such as these have been carried out in economically important aquaculture species such as Atlantic salmon (Gutierrez, Yáñez, Fukui, Swift, & Davidson, 2015), rainbow trout (Wringe et al., 2010), and big-head carp (Fu, Liu, Yu, & Tong, 2016). These genome-wide approaches are useful for the identification of loci associated with quantitative traits, such as growth, which are influenced by a large number of genes with small effect (Wellenreuther & Hansson, 2016).

The development of high-throughput DNA sequencing technologies has significantly improved the ability to obtain full sets of RNA sequences, which are more commonly referred to as transcriptomes (Wang, Gerstein, & Snyder, 2009). Transcriptome sequencing enables gene transcripts to be identified and quantified. This information can be used to investigate the differences in gene expression levels between individuals in a population. When combined with an experimental treatment, transcriptome sequencing can be used to detect differentially expressed genes (DEG) that are associated with a particular treatment. The identification of DEGs has been used to identify candidate genes for growth and improve the understanding of gene pathways that influence growth rate. Temperature, due to its strong effect on poikilotherms, is a commonly used experimental treatment in fish transcriptome studies (Liu et al., 2013; Scott & Johnston, 2012; Smith, Bernatchez, & Beheregaray, 2013; Tan et al., 2012; Tomalty et al., 2015).

In comparison to mammals, the application of these high throughput DNA sequencing techniques to fish has only occurred relatively recently and many commercially important species still lack basic genomic information. Comparative genomics is a powerful tool for transferring knowledge from one model species to a closely related non-model species that is lacking genomic information (Sarropoulou & Fernandes, 2011). The analyses that can be compared between species include polymorphism discovery, small RNA identification, and gene and genome annotation (Sarropoulou & Fernandes, 2011). Using a comparative approach can accelerate the rate at which genetic knowledge can be gained for emerging aquaculture species. As the GH-IGF axis is largely conserved from mammals to fish, aquaculture species

have largely benefited from the extensive knowledge of mammal growth systems by using comparative gene approaches (Moriyama et al., 2000). As the demand for improved aquaculture production increases, it is expected that the application of these genetic techniques to aquaculture species will also increase (Gjedrem et al., 2012). The knowledge gained from these studies can be used to assist in selective breeding programmes aimed at increasing the growth rate of a target species.

1.6. Study species: The Australasian snapper, *Chrysophrys auratus*

The Australasian snapper, *C. auratus*, is a long-lived, demersal species from the family Sparidae (Bernal-Ramírez, Adcock, Hauser, Carvalho, & Smith, 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, Sullivan, O'Brien, & Iball, 2000; Colman, 2010; Parsons et al., 2014; Paul, 1977) and is also highly sought after by recreational fishers (Parsons et al., 2014). Although there is currently no commercial aquaculture for *C. auratus*, techniques for rearing *C. auratus* effectively in captivity are being developed at Plant and Food Research (PFR) at their Seafood Research Facility in Nelson. To develop *C. auratus* as an economically viable aquaculture species, efforts must be made to improve desirable production traits such as growth rate. At present, little is known about the specific control of growth in *C. auratus*. It has been reported that growth rates of *C. auratus* are at their highest between spring and autumn, and at their lowest in winter (Francis, 1994). Body size has also been observed to vary over their geographical distribution (Paul, 1976; Paul & Tarring, 1980). It is often hypothesised that geographical variation of growth rate and body size in *C. auratus* is a result of differing sea temperatures throughout the seasons and geographic distributions (Francis, 1994; Murphy, Jenkins, Hamer, & Swearer, 2013). However, other factors such as food availability, photoperiod, behaviour, and genetic variation can also vary seasonally and with geographic location and may contribute to this variation in growth rate. There is currently a lack of published studies relating to the genetic control of growth in *C. auratus*. Further studies are required to understand the role of genetic and environmental factors on the growth rates of *C. auratus.*

1.7. Thesis aims

The overall aim of this thesis is to use genome sequence data and transcriptomics to investigate the loci that influence growth rate of *C. auratus.* The identification of the key genes associated with the control of growth rates could be used as part of a marker set in a genomics-enabled selective breeding programme.

Chapter Two

In this chapter, a comparative gene approach is used to identify and annotate the *GH* gene in *C. auratus*. The gene structure and polymorphisms are characterised. The evolutionary history of the *GH* gene in the Perciformes order is investigated and changes in the structure between *C. auratus* and other Perciforms is described.

Chapter Three

Chapter Three reports the genetic variation found in wild populations of *C. auratus* for two large repeat regions in the first and third introns of the *GH* gene identified in Chapter Two. Two groups that were identified as slow- and fast-growing *C. auratus* individuals were then sampled and their first intron repeat regions genotyped to test for an association with growth rate.

Chapter Four

In Chapter Four, a gene expression dataset that was obtained from two experimental populations (high- and low-temperature treatments) of *C. auratus* was used to test for differences in gene expression and to identify important loci for growth.

Chapter Five

Chapter Five presents a general discussion and synthesis of the previous chapters, relating the implications of these results to the development of *C. auratus* as an aquaculture species

Chapter Two: Characterisation of the growth hormone gene in *Chrysophrys auratus* and the evolutionary history of growth hormone in Perciform species

2.1. Introduction

Growth is an important phenotypic trait for aquaculture production of fish and has been the subject of an extensive number of genetic studies relating to selective breeding (Gjedrem et al., 2012; Hulata, 2001). Growth is a complex, quantitative trait controlled by a large number of genes (Mullis, 2005) as well as environmental factors (Mackay, 2001; Moriyama et al., 2000). A number of candidate genes that are important for growth have been identified in both livestock and aquaculture species including *GH, GHR, IGF,* and *MSTN* genes (De-Santis & Jerry, 2007). One of the most well-studied genes for growth is *GH*. GH is one of the main regulators of postnatal somatic growth and is known to stimulate anabolic processes such as protein synthesis, cell proliferation, skeletal muscle growth and many other physiological processes (Goodman, 1993; Reinecke et al., 2005). The action and regulation of GH is highly complex with it being involved in a number of hormone signalling pathways, cellular pathways, and feedback loops (Backeljauw & Hwa, 2016).

GH gene variants have long been associated with enhanced growth traits in livestock such as boar (Knorr, Moser, Müller, & Geldermann, 1997), chicken (Kuhnlein, Ni, Zadworny, & Fairfull, 1997), and beef cattle (Tambasco et al., 2003). More recently, polymorphisms within the *GH* gene have been associated with enhanced growth traits in aquaculture fish species (Almuly et al., 2008; Jaser et al., 2017; Sawayama & Takagi, 2015b). The importance of growth rate in selective breeding programmes has meant that *GH* genes have been characterised in a large number of important aquaculture fish species (De-Santis & Jerry, 2007; Reinecke et al., 2005). Characterisation of fish *GH* genes has revealed that a large number of teleost species share a conserved gene structure comprised of six exons and five introns. Among fish, exon sequences appear to be largely conserved, while the introns vary considerably (Almuly, Cavari, Ferstman, Kolodny, & Funkenstein, 2000; Chen, Wang, He, & Zhu, 2004). Specifically, repetitive DNA frequently occurs within the 5' flanking and intronic regions of the *GH* gene (Almuly et al., 2005; De-Santis & Jerry, 2007; Quéré et al., 2010). Variability of repeat numbers in non-coding regions have been shown to affect transcription, messenger RNA (mRNA) splicing, and gene silencing (Chorev & Carmel, 2012; Li, Korol, Fahima, Beiles, & Nevo, 2002). Such repeat sequences with associations with growth have been identified in the

GH genes of several fish species including *Sparus aurata* (Almuly et al., 2008), *Pagrus major* (Sawayama & Takagi, 2015b), and *Salmo salar* (Gross & Nilsson, 1999). Although genetic studies are becoming more common, they are still lacking for many economically important fish species. Comparative genetics is a powerful tool that enables the transfer of findings from studies of model species to closely related non-model species (Sarropoulou & Fernandes, 2011). Due to the conserved nature of the *GH* gene in fish, there is a large potential for comparative studies to accelerate the rate at which genetic advances can be made for emerging aquaculture species.

C. auratus supports one of New Zealand's largest commercial (Annala et al., 2000; Colman, 2010; Parsons et al., 2014; Paul, 1977) and recreational fisheries (Parsons et al., 2014). There is currently no commercial aquaculture for *C. auratus* in New Zealand; however, techniques for growing *C. auratus* effectively in captivity are currently being developed. At present, little is known about the specific control of growth in *C. auratus.* A reference genome has been constructed for *C. auratus* (unpublished); though, specific genetic studies on *C. auratus* growth are lacking in published literature.

The aims of the research presented in this chapter are:

1) To identify and characterise the *GH* gene structure and polymorphism in *C. auratus.*
2) To describe the evolutionary changes of *GH* in the Perciformes order.

2.2. Methods

2.2.1. Characterisation of the *GH* gene sequence in *C. auratus*

To identify the *GH* gene in *C. auratus*, a comparative gene approach was used. Genetic information from the closest living relative of *C. auratus*, the red sea bream (*P. major*), was used to identify and annotate the *GH* gene in *C. auratus.* These two species are considered sister taxa; however the genus they belong to remains undetermined (Parsons et al., 2014). Although both species have been classified under several names, they are most commonly referenced as *Chrysophrys auratus* and *Pagrus major*. 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 (unpublished) was carried out in Geneious v10 (Kearse et al., 2012) under the default settings (Program:

Megablast; results: hit table; retrieve: matching region with annotations; and a maximum of 100 hits). Once located, the *C. auratus GH* gene (*caGH*) was annotated manually using the *pmaGH* as a reference. This process revealed a DNA sequence 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 bp, scaffold 819 of the *C. auratus* genome version 1). The gap was filled using two different methods.

First, the 5' regulatory region and exon 1 were reconstructed using whole genome re-sequenced data from 12 wild *C. auratus* individuals from three populations around New Zealand (two females and two males per population), namely Manakau Harbour, Tasman Bay and Hauraki Gulf. These re-sequenced genomes were unassembled and consisted of 100 bp fragments. A Geneious BLAST search (default settings) was carried out between the *pmaGH* sequence and each re-sequenced genome. The results were binned into 'hits' and 'no hits'. BLAST hits for each re-sequenced genome were then mapped back to the *pmaGH* using the 'map to reference' function in Geneious. The *pmaGH* was set as the reference sequence and other settings were left as default (Mapper: Geneious; Sensitivity: medium sensitivity/fast; Fine tuning: Iterate up to five times; Trim before mapping: 'do not trim'). A consensus sequence of the *caGH* gene, automatically determined in Geneious, was obtained for each individual *C. auratus*. The 12 consensus sequences from these individuals were then aligned in Geneious 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 due to the technical difficulties of determining the DNA sequence of a highly repetitive region, so direct Sanger sequencing was used to determine the DNA sequence. DNA was extracted from tissue samples of 20 juvenile *C. auratus* using a modified high salt extraction method (Aljanabi & Martinez, 1997). Samples came from wild *C. auratus* collected from the north of New Zealand by NIWA in 2013/14 for a separate project and subsequently stored in 85% ethanol at -20 °C. Approximately 20 mg of tissue was added to 400 μL of extraction buffer (0.4 M NaCl, 10 mM Tris-HCl pH 8.0, and 2 mM EDTA pH 8.0), 40 μL of 20% sodium dodecyl sulphate (SDS), and 20 μL of 10 mg/mL proteinase K and incubated at 56 °C for 1-2 hours to digest tissue. After the tissue was fully digested, 300 μL of 6 M sodium chloride was added to each tube.

Tubes were mixed thoroughly and centrifuged for 15 minutes at 12000 rpm to collect all unwanted proteins and other matter in a pellet. The aqueous layer was retained and centrifuged again. The aqueous layer was kept and 900 mL of isopropanol was added before samples were stored at -20 °C for 1 hour. Tubes were then centrifuged at 4 °C for 20 minutes and the aqueous layer was removed. The pellets were washed with 70% ethanol and centrifuged at 4 °C for 10 minutes before the supernatant was discarded. Pellets were then dried in a vacuum concentrator and resuspended in Tris-EDTA buffer (10mM TrisHCl pH 8.0, 1mM EDTA). The DNA was quantified using a Nanodrop ND1000 spectrophotometer (Thermo Fisher Scientific).

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). PCR reactions were carried out in 15 μL volumes containing 1x PCR buffer (67 mM Tris-HCl (pH 8.8 at 25 °C), 16 mM $(NH_4)_2SO_4$), 0.6 μM Bovine Serum Albumin (BSA), 200 μM of each dNTP, 0.4 μM of each primer, 2 mM $MgCl_2$, 1 unit of Biotaq DNA polymerase (Bioline Reagents Limited), and 1 μL 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 run on 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 Parmacia Biotech) and sequenced on a ABI3730 DNA Analyzer (Massey Genome Services, Palmerston North). Once a complete intron 1 sequence was obtained it was inserted into the correct position in the *caGH* gene in Geneious using the *pmaGH* gene as a reference. This resulted in a full *GH* gene sequence for *C. auratus* that could subsequently be annotated and analysed.

The *caGH* sequence was searched for polymorphic loci manually 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, Longden, & Bleasby, 2000); the settings allowed no mismatches and a minimum length of 4bp for matches. The *caGH* amino acid sequence was determined and the secondary structure was predicted using Geneious plugin EMBOSS Protein (Rice et al., 2000).

2.2.2. Phylogenetic analysis of the *GH* gene in Perciform fish

GH gene sequences from 20 fish species in the order Perciformes were obtained from Genbank and imported to MEGA v7 (Kumar, Stecher, & Tamura, 2016) along with the *caGH* sequence from this study (Table 2.1). Three outgroups from the orders Pleuronectiformes, Siluriformes, and Salmoniformes, within the class Actinopterygii, were also imported to MEGA from Genbank (Table 2.1).

Table 2.1. List of species used in the phylogenetic analysis of *GH* genes in this study. Includes the Genbank accession number, and the taxonomic family and order classifications for each species.

Species	Accession number	Family	Order
Chrysophrys auratus	This study	Sparidae	Perciformes
Pagrus major	AB904715	Sparidae	Perciformes
Sparus aurata	U48221	Sparidae	Perciformes
Dicentrarchus labrax	GQ918491	Moronidae	Perciformes
Seriola quinqueradiata	D50368	Carangidae	Perciformes
Siniperca scherzeri	EF441623	Percichthyidae	Perciformes
Siniperca chuatsi	EF205280	Percichthyidae	Perciformes
Siniperca kneri	EF205281	Percichthyidae	Perciformes
Lates calcarifer	LCU16816	Latidae	Perciformes
Epinephelus coioides	KR269816	Serranidae	Perciformes
Channa diplogramma	GQ214246	Channidae	Perciformes
Channa marulius	GQ214245	Channidae	Perciformes
Channa gachua	GQ214244	Channidae	Perciformes
Channa punctata	GQ214243	Channidae	Perciformes
Channa striata	EF447030	Channidae	Perciformes
Tilapia nilotica	M84774	Cichlidae	Perciformes
Coptodon zillii	KT387600	Cichlidae	Perciformes
Sarotherodon galilaeus	KT387599	Cichlidae	Perciformes
Tilapia mossambica	AF0338061	Cichlidae	Perciformes
Pseudosciaena crocea	AY090592	Sciaenidae	Perciformes
Nibea coibor	FJ375311	Sciaenidae	Perciformes
Outgroups			
Paralichthys olivaceus	D29737	Paralichthyidae	Pleuronectiformes
Clarias batrachus	AF416486	Clariidae	Siluriformes
Oncorhynchus keta	L04688	Salmonidae	Salmoniformes

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. A maximum likelihood tree was constructed in MEGA using the Kimura 2-parameter model (Kimura, 1980) with 1000 bootstrap replications. Scaled diagrams of the basic *GH* gene structures (exon-intron sizes) were manually produced for each species to assist

with the phylogenetic analysis. Additionally, *GH* sequences were translated into their amino acid sequences in Geneious v10 and aligned using a MUSCLE alignment. Similarities between amino acid sequence, hydrophobicity, and polarity were assessed by eye. Secondary structures were predicted using the EMBOSS Protein plugin (Rice et al., 2000).

2.3. Results

2.3.1. Structure of the *caGH* gene

The *caGH* was identified on scaffold 819 of version 1 of the *C. auratus* genome. Nucleotide sequence analysis based on *pmaGH* annotation (Sawayama & Takagi, 2015a) (Accession AB904715) revealed that the *caGH* is approximately 5,577 bp in length. Its putative start site (start of exon I) was at 66,741 bp, and the end site was at 72,318 bp. The *caGH* is comprised 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 2.1).

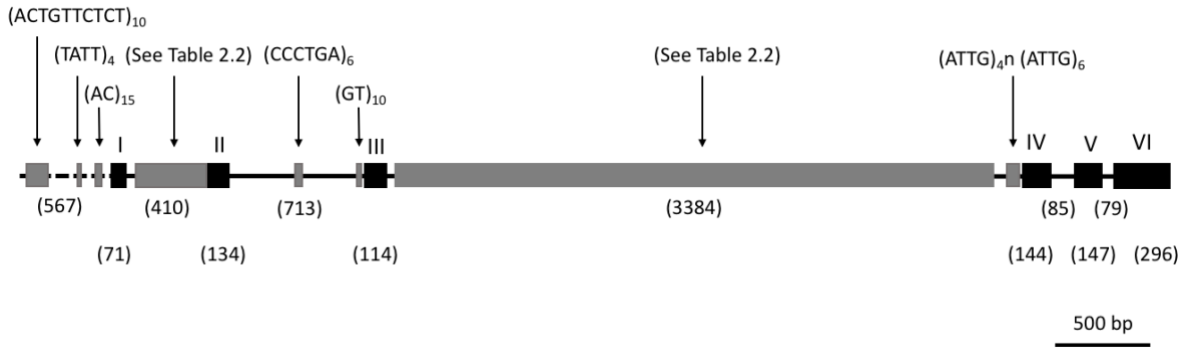

Figure 2.1. Structure of the *caGH* gene. Dotted line = 5'-flanking region. Black boxes = exons. Lines between boxes = introns. Brackets below figure = size in base pairs. Grey boxes = repeat regions. Brackets above figure = sequence motifs of each repeat region.

The *GH* locus identified in this study was the only match in the BLAST search between the *C. auratus* genome and the *pmaGH*. Using the Geneious plugin Repeat Finder, and *pmaGH* annotations as a reference, several repeat regions were identified in *caGH* (Figure 2.1). Introns 1 and 3 contained large repeat regions, the sequence structures of which are detailed in Table 2.2.

Table 2.2. Explanation of complex repeat regions in intron 1 and 3 of the *caGH* gene. Depicts individual core repeat sequences and the overall structure of the repeat region. Grey highlighted regions represent conserved sequences between each core repeat sequence within each repeat region.

Structure of complex repeat regions		
Intron 1 core repeats		**Length**
1	TGACCTGTCTGTCTCTCTCTCTCTCTGTC	29 bp
2	TGACCTGTCTGTCTCTC	17 bp
3	TGACGTGTCTGTCTCTCTCTCTTTCTCTCTGTC	33 bp
4	TGACCTGTCTGTCTGTCAGTGGTGCTGATGCTGTCGGTGCTGTCTC	46 bp
Overall repeat region structure: 1, 2, 3, $(2, 4)_5$		
Intron 3 core repeats		
$\mathbf{1_A}$	CACACAGTACTACATGATACTG	22 bp
$\mathbf{1_B}$	CACACAGTACTACATGATACTA	22 bp
2	CACACGCACTACATGATACTACACAGTACTCATGATACTA	41 bp
Overall repeat region structure: $(1_A, 1_A, 1_B, 2)_{28}$		

The process of reconstructing the DNA sequence gap using 12 re-sequenced *C. auratus* genomes and direct sequencing of 10 intron 1 repeat sequences allowed for several initial observations to be made about the intra-specific variation in the *caGH* gene sequence. The 12 re-sequenced genome sequences showed that the first 150 bp of the 5' flanking region contains nine variable single-nucleotide sites, five are parsimony-informative, and four are singleton sites. The first two repeat regions in the 5' flanking region (Figure 2.1) are monomorphic in all 12 individuals, while the third repeat sequence in the 5' flanking region, $(AC)_{15}$, ranged from 13-15 repeats in these individuals. Aside from these variable sites, the 5' flanking region was conserved among all 12 individuals. Exon I and the first 43 bp of intron 1 (before the repeat region) were completely conserved between these 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 repeat number between individuals. Three individuals had a repeat region of 582 bp in length while the other seven individuals had repeat regions 393 bp long. The three 582 bp sequences contained a cytosine as the 68^{th} nucleotide of the repeat region while the other seven sequences had a thymine at this site. 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 some areas. The first repeat region in intron 2 with the motif 'CCCTGA' ranged from three to six repeats among individuals (Figure 2.1). The second repeat region in intron 2 with the motif 'GT' was monomorphic across all 12 individuals (Figure 2.1). 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.

Putative transcription factor binding sites and regulatory motifs are listed in Table 2.3. The 5' regulatory region contains a typical TATA box located at -106 to -112 bp upstream of the first codon of exon I. Two other TATA boxes were identified in the 5' flanking region at -230 to -236 bp and -323 to -329 bp upstream of the first codon. There are two putative pituitary-specific transcription factor (Pit-1a) consensus sequences located upstream of the first codon at -200 to -207 bp and -124 to -131 bp. Additionally, a cAMP response element (CRE) was identified at -76 to -82 bp upstream of the first codon. These consensus sequences are also found throughout the full gene sequence (Table 2.3). The repeat region in the 5' flanking region with the sequence motif $(ACTGTTCTCT)_{10}$ (Figure 2.1) contains a putative glucocorticoid response element (GRE) consensus sequence. There are also numerous other putative GRE sites in the 5' flanking region. No putative thyroid response element (TRE) or estrogen 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) is also found in this core repeat unit (Table 2.2). 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) are listed in Table 2.3. The amino acid sequence and predicted secondary structure are illustrated in Figure 2.2. The *C. auratus* GH consists of 203 amino acids. Aside from the first codon, the translated *caGH* has three additional start codons (M = Methionine) at amino acids 7, 127, and 181 (Figure 2.2).

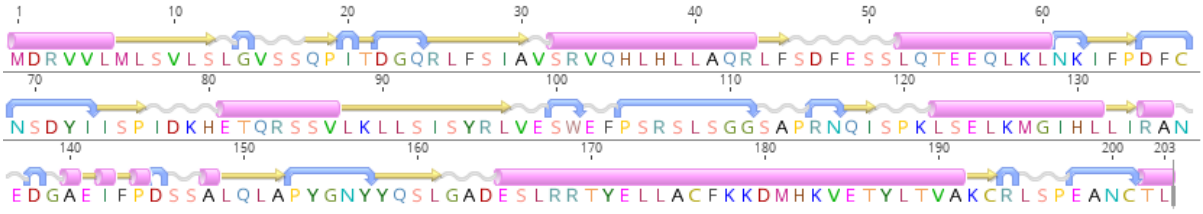

Figure 2.2. Amino acid sequence and predicted secondary structure of the *C. auratus* GH protein. Letters represent amino acids. Pink cylinders = alpha helices; yellow arrows = beta strands; grey waves = coil; blue arrows = turns.

Table 2.3. 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.

Putative transcription factor or regulatory motif	**Full name/Description**	**Occurrence in *caGH* (5' regulatory region, exons I-VI, introns 1-5)**
GR	Glucocorticoid receptor	5', I, II, 2, III, 3, IV, V, VI
GATA-1	Erythroid transcription factor 1	5',1, 2, III, 3, IV, 5
c-Ets-2	ETS – proto-oncogene 2	5', I, 2, III, IV, 4, V, VI
NF-E	Nuclear factor ets-like	5', 1, II, 2, IV, 4, V
Sp1	Specificity protein 1	5', I, II, 2, III, 3, IV, 4, V, VI
NF-1	Nuclear factor 1	5', I, II, 2, IV, V, VI
TFIID	Transcription factor II D	5', V
NF-E2	Nuclear factor, erythroid 2	5'
NFIII-e	NFIII-e	5'
ETF	EGFR-specific transcription factor	5'
Oct-1	Octamer binding transcription 1	5', 2, 4, VI
TBP	TATA-box binding protein	5', V
Pit-1a	Pituitary specific positive transcription factor 1	5', 2, 4
HiNF-A	Histone nuclear factor A	5', 2
PEA3	Polyoma enhancer activator 3	5', 4, V
ATF	Activating transcription factor	5', I, 1, II, 2, III, IV, V, VI
c-Myb	MYB proto-oncogene, transcription factor	1, 2, III, 3, VI
TRE	Thyroid response element	1
GAGA	GAGA transcription factor	1, II, 2, VI
CRE	cAMP response element	1, VI
Oct-1C	Octamer binding transcription 1C	1, II, 2, 3, 4, VI
PR	Progesterone receptor	II
CBF-B	Core binding factor beta subunit	II, III, 3, IV, VI
CAC-binding protein	CAC-binding protein	II, IV, VI
H4TF-1	Histone H4 gene-specific transcription factor 1	2, III, IV
ER	Estrogen receptor	2, III, IV
CACCC binding factor	CACCC binding factor	2, VI
Gamma CAC2	Gamma CAC2	2, IV, VI
NF-Y	Nuclear transcription factor Y	2, 3, IV
AP-2	Transcription factor activating protein 2	2
CTF	CCAAT box-binding transcription factor	2, 3, IV, V
H4TF-2	Histone H4 gene-specific transcription factor 2	2, 4, V
CP1	CTF transcription factor/Nuclear factor 1	2, 4, V
PU.1	PUrine-box binding factor 1	3
HNF-1	Hepatocyte nuclear factor 1	3
VHNF1	Hepatocyte nuclear factor 1 beta	3
GR-beta	Glucocorticoid receptor beta	3, 4
c-Myc	c-Myc proto-oncogene, BHLH transcription factor	3, 4, V
CCAAT binding factor	CCAAT binding factor	3, IV
PTF1-beta	Pancreas specific transcription factor 1b	IV
α-IRP	α-IRP	4, V
CDP2	Cut-like homeobox 2	4, V
SRF	Serum response factor	4, V
TGGCA – binding protein	TGGCA – binding protein	4, V
H1TF2	Histone 1 transcription factor 2	4, V
CP2	Transcription factor CP2	4, V
α-CBF	Alpha – core binding factor	4, V
Tal-1	Erythroid differentiation factor	V
HEB	Transcription factor 12	V
GATA-2	GATA binding protein 2	5
IRF-2	Interferon regulatory factor 2	VI

2.3.2. Phylogenetic analysis of the *GH* gene in Perciform fish

A phylogeny of 21 species from the order Perciformes was reconstructed using a maximum likelihood approach to assess the evolutionary relationships of the *GH* gene sequence in this order and how *caGH* compares to the *GH* genes of other closely related fish species (Figure 2.3). The 21 Perciform species were monophyletic with respect to the three outgroup taxa. All of the species formed monophyletic groups in their respective families. *C. auratus* was grouped within the family Sparidae. This monophyly was supported by high bootstrap values, while the relationships between families were not highly supported. All Perciform fish species had a highly conserved *GH* gene structure consisting of six exons and five introns (Figure 2.4). Of the three outgroups, *Paralichthys olivaceus* and *Oncorhynchus keta* conformed to this gene structure while *Clarias batrachus* 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* (*see* Appendix Table A.1). 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 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 relatively similar in length across species, while introns 1-3 varied extensively (Figure 2.4). This variability in intron length is responsible for the large range of gene sizes within Perciform fish.

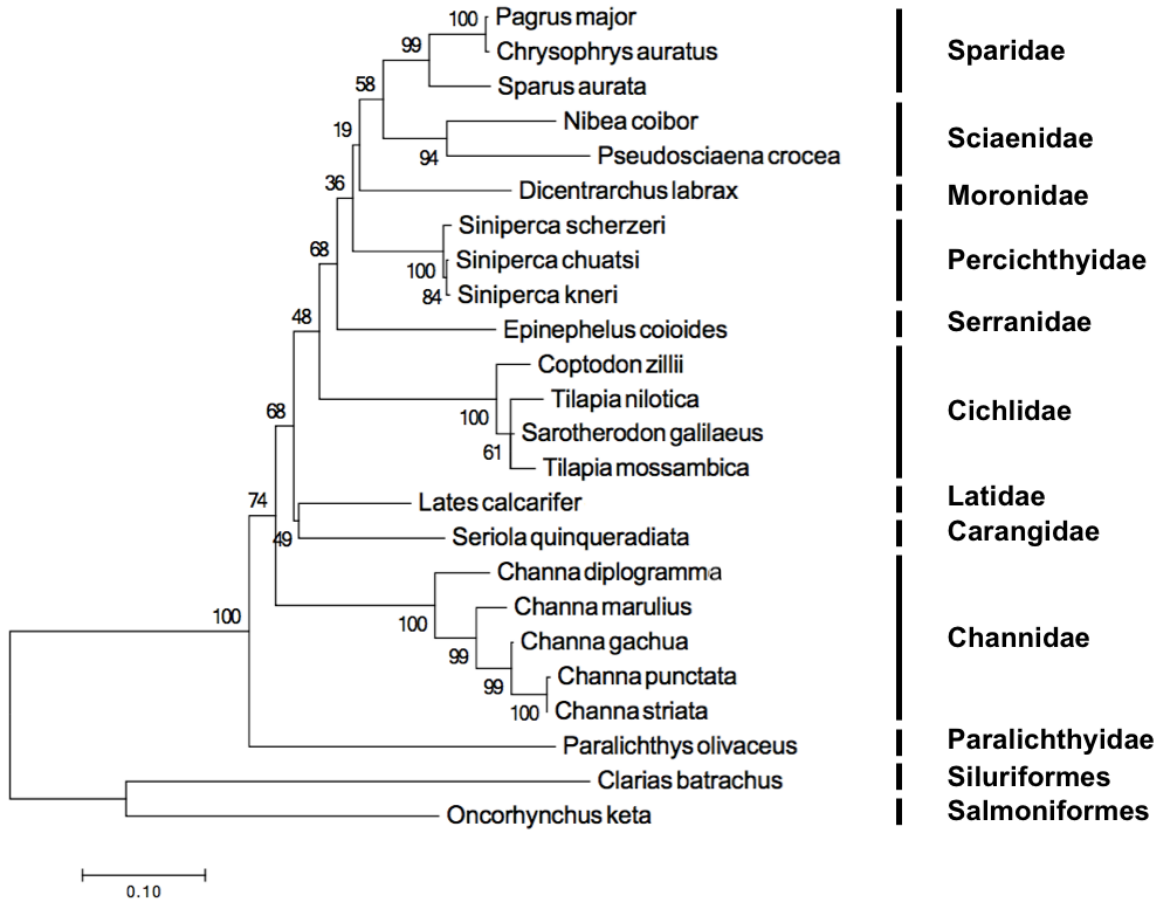

Figure 2.3. Phylogenetic tree of Perciform species *GH* genes inferred by the Maximum Likelihood method based on the Kimura 2-paramenter model (Kimura, 1980). Bootstrap values are shown next to branches. Respective taxonomic families are shown to the right of the species names. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

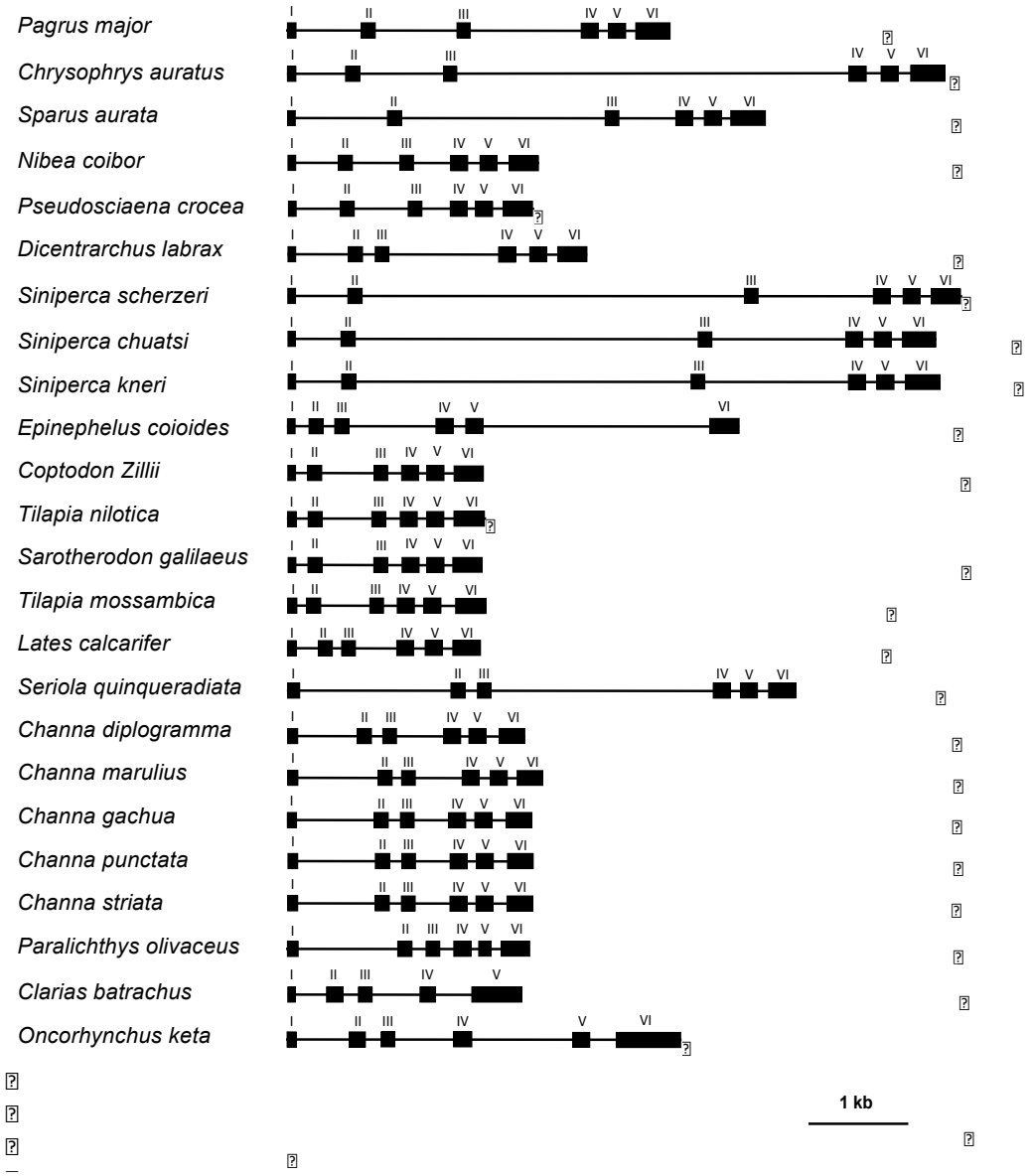

Figure 2.4. *GH* gene structures for species used in this study for phylogenetic analysis. Species are listed in the same order as the phylogenetic tree in Figure 2.3. Black boxes = exons. Lines between boxes = introns. Roman numerals = exon number. For information on sizes of exons and introns, see Appendix Table A.1.

To understand the evolutionary differences of GH as a functional protein, the DNA sequences were converted into amino acid sequences and their secondary protein structures were predicted. Alignment of these amino acid sequences shows an overall similarity among Perciform fish (Figure 2.5). All Perciform fish have a sequence of 204 amino acids excluding *P. major* and *C. auratus* which have 203 due to a 3 bp deletion in their third exons (= one codon/amino acid). There is a low level of variation in amino acid sequence identity between species (Figure 2.5). *P. major* and *C. auratus* are lacking the 67th amino acid in the sequence (leucine). Despite differences in amino acid sequence, the overall hydrophobicity and polarity of the amino acid sequences are highly conserved across all Perciform fish and the three

outgroups (Figures 2.6 & 2.7). 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 (See second gap in *P. major* and *C. auratus* amino acid sequence in Figure 2.6). The proline following the deleted leucine (non-polar) in *P. major* and *C. auratus* is also more non-polar than the glutamine (polar) in all other Perciform fish (see second gap in *P. major* and *C. auratus* amino acid sequence in Figure 2.7). This also leads to three consecutive non-polar amino acids at this site. With these three consecutive hydrophobic, non-polar amino acids, the predicted secondary structure at this site is conserved across all Perciform fish (Figure 2.8). Despite variability in amino acid sequence, the overall predicted secondary structure of perciform *GH* genes is relatively conserved across species.

Figure 2.5. Similarity between Perciform GH amino acid sequences. Numbers 1-24 = Species name in same order as phylogenetic tree (Figure 2.3). Black = more similar and white = less similar. Thin horizontal lines between amino acids = amino acid gap.

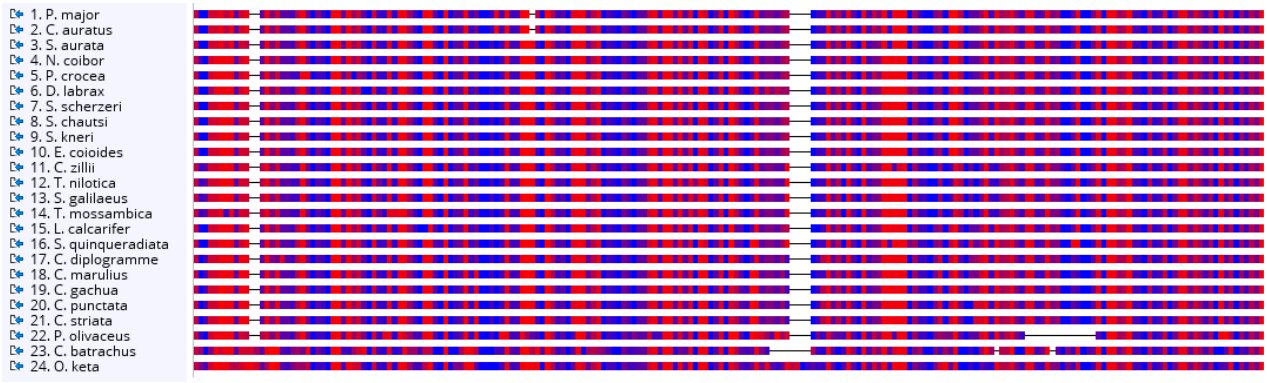

Figure 2.6. Hydrophobicity of Perciform GH amino acid sequences. Numbers 1-24 = Species name in same order as phylogenetic tree (Figure 2.3). Blue = hydrophilic and red = hydrophobic. Thin horizontal lines between amino acids = amino acid gap.

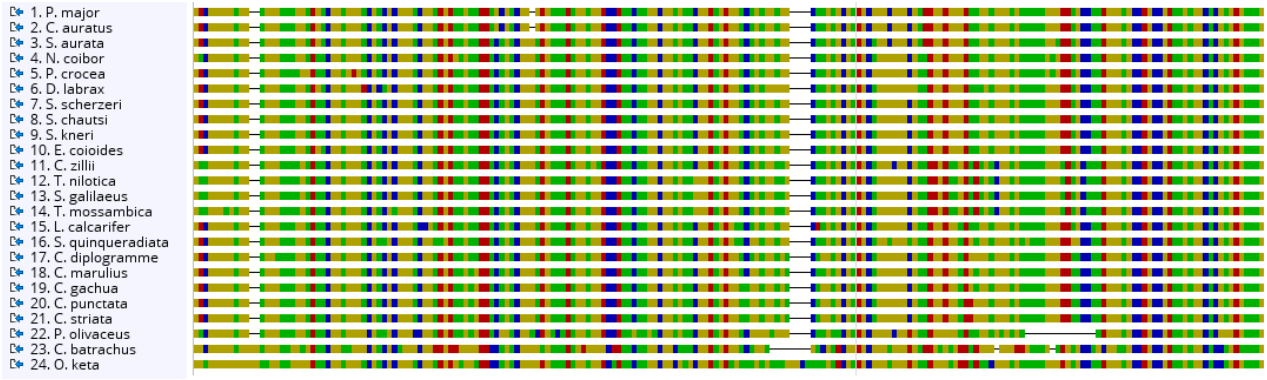

Figure 2.7. Polarity of Perciform GH amino acid sequences. Numbers 1-24 = Species name in same order as phylogenetic tree (Figure 2.3). Blue = positively charged; red = negatively charged; green = polar; yellow = non-polar. Thin horizontal lines between amino acids = amino acid gap.

Figure 2.8. Predicted secondary structures of Perciform GH proteins. Numbers 1-24 = Species name in same order as phylogenetic tree (Figure 2.3). 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.

2.4. Discussion

2.4.1. Structure of the *caGH* gene

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 is 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* for the *pmaGH* gene suggesting the presence of only one copy of the *GH* gene in this species. Duplicate copies of this gene have been identified in other teleost species such as salmon (McKay, Trautner, Smith, Koop, & Devlin, 2004) and tilapia (Ber & Daniel, 1993), likely due to a whole genome duplication event early in teleost evolution (Glasauer & Neuhauss, 2014). However, duplicate *GH* genes appear to be absent in some teleost species such as close relative to *C. auratus*, *S. aurata* (Almuly et al., 2000). 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 (Quéré 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 promotors and introns has been reported in a range of species (Almuly et al., 2008; Arango, Echeverri, & López, 2014; Nie et al., 2005; Quéré et al., 2010; Zhang, Shao, Jiang, Li, & Xu, 2009).

Strong evidence suggests that the locations of repeat sequences have some functional significance (Li et al., 2002). 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, Korol, Fahima, & Nevo, 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 is often cited in the literature that this kind of gene regulation occurs frequently in 5' proximal introns in particular and occurs across multiple genes and species (Chorev & Carmel, 2012). The exact mechanisms that allow intron repeat sequences to influence gene expression are still largely undefined. One hypothesis predicts that due 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, Mekhedov, Hartl, Koonin, & Kondrashov, 2002). The reduced expression of gene variants with long introns has been observed in the first introns of the Friedreich ataxia (Ohshima, Montermini, Wells, & Pandolfo, 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, Senthilkumar, Singh, & Mishra, 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. 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 further investigated in *C. auratus* as candidate markers for growth.

To further understand the regulation of *GH* expression in *C. auratus*, 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 promotor region, including Pit-1a, CRE, GRE, TRE, ERE (Moriyama, Oda, Takahashi, Sower, & Kawauchi, 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/promotor. 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, Tomita, Wang, Buchholz, & Shi, 2006; Plateroti, Kress, Mori, & Samarut, 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 other

teleost fish species (this study). The *C. auratus* amino acid sequence contains four methionines which translate to start codons. The presence of multiple translation start sites within a single mRNA transcript suggests the possibility of multiple GH protein isoforms; however this phenomenon remains poorly investigated (Bazykin & Kochetov, 2010).

2.4.2. Phylogenetic analysis of the *GH* gene in Perciform fish

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). Due to the sheer size of this order, the phylogenetic relationships are still debated today. 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, Sun, Zhang, Wang, & Xu, 2014). However, the exact relationships between some families are not well supported in this phylogeny and remain unclear in other studies. For example, the results of a recent large-scale study (Betancur-R et al., 2017) have split these 21 species, that are traditionally categorised under the Perciformes order, into several different orders. These proposed orders have been phylogenetically clustered into three distinct groups (Figure 2.9). In their study, the families Latidae (proposed order: incertae sedis - undefined), Carangidae (proposed order: Carangiformes), and Channidae (proposed order: Anabantifomes) are grouped with Pleuronectiformes (containing *P. olivaceus*, an outgroup in the current study) as more closely related to each other than to the other families in this study (group 1). This group shares a common ancestor with the Cichlidae family (proposed order: Cichliformes; group 2). These two groups, in turn, share a common ancestor with the remaining families of this study (group 3); Sparidae (proposed order: Spariformes), Sciaenidae (proposed order: incertae sedis), Moronidae (proposed order: incertae sedis), Percichthyidae (proposed order: Centrarchiformes), and Serranidae (proposed order: Perciformes).

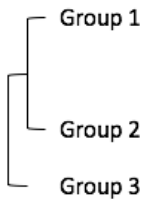

Figure 2.9. Visual representation of relationships of families in the current study according to Betancur-R et al. (2017). Group 1 = Latidae, Carangidae, and Channidae. Group 2 = Cichlidae. Group 3 = Sparidae, Sciaenidae, Moronidae, Percichthidae, and Serranidae.

As these higher-level classifications are still unclear, it is difficult to determine the evolutionary history of the *GH* gene and how its sequence has changed over time between these families. However, when taken together as one group, several overall trends can be observed in these species. (Note: the species in this study will continue to be classed as Perciform fish as per current accepted classifications). 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 2 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 theory that an intron was inserted into what was originally exon V leading to the creation 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 Salmoniformes, Perciforms, and Terradontiforms (Moriyama et al., 2006; Zhang et al., 2009) which is consistent with the findings of the current 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. 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). Based on the findings of this study and the literature on the importance of introns for gene expression, it is conceivable that this intronic variation may contribute to the diversity of growth phenotypes in Perciform species.

To further understand the evolution of GH as a functional protein, amino acid sequences and predicted secondary structures of these species were compared. Despite a low level of variation in amino acid sequence among Perciforms, the overall hydrophobicity and polarity of the

sequences are highly conserved across all Perciform fish and the three outgroups. All Perciform species had a sequence length of 204 amino acids with the exception of *P. major* and *C. auratus* which had 203 amino acids because of a 3 bp deletion (one codon/amino acid) in their third exons. Despite the loss of one amino acid (leucine), the predicted secondary structures of *P. major* and *C. auratus* remain fundamentally consistent with other Perciform species in this study. The findings in the present study suggests 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 sensitive to insertions or deletions are not well understood; however, some studies have observed that deletions are often followed by compensatory substitutions elsewhere in the amino sequence (Choi, Sims, Murphy, Miller, & Chan, 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 this architecture to ensure protein functionality (Pál, Papp, & Lercher, 2006; Taylor, Ponting, & Copley, 2004).

2.4.3. Conclusions

From the results of this study it can be concluded that across fish species, the *GH* coding regions, amino acid sequences, and predicted secondary structures are highly conserved. The extensive sequence variation present between species lies within the non-coding regions of this gene. In this study, the *caGH* gene was characterised and assessed for polymorphism. Repeat regions were identified within the 5' flanking region and introns 1-3. In particular, introns 1 and 3 contained large, complex repeats. 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 the foundation for further study into the relationship between these polymorphic regions and growth traits in *C. auratus*. If these regions prove 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 selection of *C. auratus*.

Chapter Three: Assessment of the *Chrysophrys auratus* growth hormone intron 1 and 3 repeat region variation and their association with growth performance

3.1. Introduction

One of the most extensively studied candidate genes for growth is the *GH* gene. GH is one of the main regulators of postnatal somatic growth and is known to stimulate anabolic processes such as protein synthesis, cell proliferation, skeletal growth, and many other physiological processes (Goodman, 1993; Reinecke et al., 2005). *GH* variation has been associated with improved growth traits in a range of 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). In recent years, *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 growth markers that can be used to guide selection for growth rate improvements. The cross-species DNA sequence information about the *GH* gene has enabled comparative analyses to be conducted. 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).

In particular, a number of variable repeat regions have been identified in the non-coding regions of *GH* genes in many vertebrate species (Almuly et al., 2000; Arango et al., 2014; Sawayama & Takagi, 2015a). It has been shown that repeat regions within non-coding sections 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). However, the molecular mechanisms to explain how repeat elements within these non-coding regions influence gene expression are largely unknown. One hypothesis predicts that due 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). It has also been suggested that repeat elements in non-coding regions 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). Furthermore, transcription factor binding sites within intronic repeat regions have been shown to directly affect gene expression (Almuly et al., 2008). Repeat elements in the first and third

introns of *GH* have been developed as genetic markers for improved growth for fish species *S. aurata* (Almuly et al., 2008) and *P. major* (Sawayama & Takagi, 2015b) respectively. In Chapter Two of this thesis, the *caGH* gene was characterised revealing large polymorphic intronic repeat regions in introns 1 and 3. These repeat regions were the same as those associated with growth in close relatives *S. aurata* and *P. major*.

The aims of the research presented in this chapter are:

1) To determine the level of variation at two large *GH* intronic repeat loci in *C. auratu*s. Specifically, wild population samples will be used to optimise the PCR amplification and genotyping methods and to assess the type and level of variable in the two intronic repeat regions.
2) Groups of slow- and fast-growing *C. auratus* will then be sampled and their *GH* intronic repeats genotyped to test for a correlation with growth traits.

3.2. Methods

3.2.1. Sample collection, DNA extraction and genotyping of the *caGH* intron 1 and 3 repeat regions

Intronic repeat variation was assessed in 99 juvenile *C. auratus* collected from five harbours in the north of New Zealand; Kaipara (n = 20), Parengarenga (n = 20), Rangaunu (n = 19), Bay of Islands (n = 20), Doubtless Bay (n = 20). These samples were collected by NIWA in 2013/14 for a separate project and were subsequently stored in 85% ethanol at -20 °C. Genomic DNA was extracted from samples using a modified high-salt extraction method (Aljanabi & Martinez, 1997). Approximately 20 mg of tissue was added to 400 μL of extraction buffer (0.4 M NaCl, 10 mM Tris-HCl pH 8.0, and 2 mM EDTA pH 8.0), which contained 40 μL of 20% SDS, and 20 μL of 10 mg/mL proteinase K. The solution was incubated at 56 °C and gently mixed for 1-2 hours. After the tissue was fully digested, 300 μL of 6 M sodium chloride was added to each tube to precipitate the proteins. Tubes were mixed thoroughly and centrifuged for 15 minutes at 12000 rpm to pellet the precipitated proteins and other unwanted matter. The clarified aqueous layer was transferred to another tube and centrifuged again. The aqueous layer was transferred to another tube and 900 mL of isopropanol was added and the samples were kept at -20 °C for 1 hour. Tubes were then centrifuged at 4 °C for 20 minutes and the aqueous layer was removed. The DNA pellet was washed with 70% ethanol and centrifuged at

4 °C for 10 minutes before the supernatant was discarded. Pellets were then dried in a vacuum concentrator and resuspended in Tris-EDTA buffer. The DNA was quantified using a Nanodrop ND1000 spectrophotometer (Thermo Fisher Scientific).

The repeat regions of intron 1 and 3 were amplified using the PCR with the primers listed in Table 3.1. 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 μL volumes containing 1x PCR buffer (67 mM Tris-HCl (pH 8.8 at 25 °C), 16 mM $(NH_4)_2SO_4$), 0.6 μM BSA, 200 μM of each dNTP, 0.1 μM forward primer, 0.4 μM reverse primer, 0.4 μM FAM labelled M13 primer, 2 mM $MgCl_2$, 1 unit of Biotaq DNA polymerase (Bioline Reagents Limited), and ~100ng of DNA. Reactions required a minimum of 100 ng of DNA due to the moderate level of degradation in these aged tissue samples. Thermal cycling conditions are outlined in Table 3.1.

Table 3.1. PCR primers and cycling conditions. M13 primer = 5' TGTAAAACGACGGCCAGT 3'. *GH* Intron 1 primers were taken from Almuly et al. (2000) and intron 3 primers from Sawayama and Takagi (2015a).

Genomic locus	**Primer (5'-3')**	**Cycling conditions**
GH intron 1	Forward: M13-AGAACCTGAACCAGACATGG Reverse: AGGTGTTGAACTCTGCTGAC	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.
GH intron 3	Forward: M13-GATTACATCATCAGCCCCATC Reverse: GTCTGAAATGTCTGTCCTGAAGC	An initial step of 95 °C for 5 minutes followed by 30 cycles of 94 °C for 30 seconds, 66 °C for 30 seconds, and 72 °C for 120 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. Due to the excessive polymorphism and long length of the intron 3 repeat region, the

amplification of this fragment was unreliable and the genotyping impractical with this method. Consequently, investigation into this locus was discontinued.

3.2.2. Data analysis of the *caGH* intron 1 repeat region

Allele peaks were viewed and analyzed in Geneious v 10 (Kearse et al., 2012) and manually placed into size 'bins' for each allele size based on known core repeat sizes. 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 θ (F_{ST}), and f (F_{IS}) were calculated using FSTAT (Goudet, 2001; Weir & Cockerham, 1984) and deviations from Hardy-Weinberg equilibrium (HWE) were assessed in GENEPOP (Raymond, 1995) using a Fisher's exact test.

3.2.3. Testing for an association between *GH* intron 1 repeat region variation and growth rate in *C. auratus*

C. auratus larvae were reared from a broodstock at the PFR fish culture facility. Fish were grown in 5000 L tanks with the same ambient temperature and food ration. Overstocking was prevented by transferring fish into additional 5000 L tanks as required. After 10 months, 21,891 juvenile *C. auratus* at a mean size of 37 g were transferred to the Ngai Tahu Seafood sea cage in Beatrix Bay, Marlborough to investigate the suitability of *C. auratus* for sea-cage aquaculture. Commercial grow-out of these individuals was then carried out for 17.5 months before the trial was concluded. Fish were then classed as slow- or fast-growing using the 5% extremes of the ends of the size distribution and tissue samples from groups of slow (n = 31) and fast (n = 37) growing *C. auratus* were collected. The aforementioned experimental setup was designed for a separate project; however, the slow- and fast-growing samples above were made available for the purposes of this thesis research.

DNA was then extracted from tissue samples using the same modified salt extraction method as previously outlined (see 3.2.1). The intron 1 locus was amplified by PCR using the same master mix, primers and conditions (Table 3.1) as above; however, PCR reactions for these samples required only ~50 ng of template DNA. PCR products were electrophoresed on 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 using the LIZ 1200 standard at Massey Genome Services on an ABI3730 Genetic Analyzer.

3.2.4. Data analysis testing for associations between *caGH* repeat region variation and growth rate

Allele peaks were viewed and analyzed in Geneious v 10 (Kearse et al., 2012) and manually placed into size 'bins' for each allele size based on known core repeat unit sizes (see Table 2.2). Alleles were named based on their respective peak sizes. GenAlEx 6.5 (Peakall & Smouse, 2006) was used to calculate allele number (N_a), observed heterozygosity (H_o), expected heterozygosity (H_e), and allele frequencies. Weir and Cockerham's F statistics θ (F_{ST}), and f (F_{IS}) were calculated using FSTAT (Goudet, 2001; Weir & Cockerham, 1984). Deviations from HWE were assessed by population and by locus in GENEPOP (Raymond, 1995) using the Fisher's exact test and each population was also tested for heterozygote deficiency and excess. Significant differences in genotype count and allele frequencies between slow- and fast-growing groups were assessed using Fisher's exact tests. Allele frequencies of the three alleles common to all populations (alleles: 485, 550, 765) were then compared between all populations from this study, wild and cultured, and chi-squared tests were carried out to assess for significant differences between populations.

3.3 Results

3.3.1. Assessment of *caGH* intron 1 and 3 repeat region variation

Population statistics including allele number, observed heterozygosity, and expected heterozygosity for the intron 1 locus are detailed in Table 3.2. The number of intron 1 alleles in 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.

Table 3.2. *GH* intron 1 allelic diversity and heterozygosity of wild samples of *C. auratus*. N, sample size; Na, number of alleles; Ho, observed heterozygosity; He, expected heterozygosity.

Sample site	N	Na	Ho	He
Bay of Islands	20	6	0.800	0.678
Doubtless Bay	20	3	0.500	0.546
Kaipara	20	4	0.650	0.581
Parengarenga	20	6	0.700	0.643
Rangaunu	19	4	0.632	0.594

Weir and Cockerham's F-statistic θ (F_{ST}) 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 presented in Table 3.3. Neither

the sample sites nor intron 1 locus exhibited significant p-values and therefore offer no evidence for departure from HWE.

Table 3.3. Significance values for exact tests assessing deviation from HWE by sample site and locus in the *GH* intron 1 locus of wild *C. auratus* samples.

Sample Site	**P-value**
Bay of Islands	0.3995
Doubtless Bay	0.8230
Kaipara	0.4244
Parengarenga	0.9863
Rangaunu	1.0000
Locus	**P-value**
Intron 1	0.9489

The intron 3 locus was difficult to amplify and too long to be accurately genotyped. However, visual inspection of PCR products after electrophoresis in agarose gels suggested that the intron 3 repeat region is polymorphic and has at least three alleles (approximately 900, 1600, and 2500 bp in length).

3.3.2. Testing for an association between the *GH* intron 1 repeat region variation and growth rate in *C. auratus*

All population statistics including allele number, observed heterozygosity, and expected heterozygosity are found in Table 3.4. The slow- and fast-growing snapper groups had three and four alleles respectively with three of these alleles occurring in both groups (485, 550, 765).

Table 3.4. *GH* intron 1 allelic diversity and heterozygosity of slow- and fast-growing *C. auratus* samples. N, sample size; Na, number of alleles; Ho, observed heterozygosity; He, expected heterozygosity.

Population	**N**	**Na**	**Ho**	**He**
Slow-growing	31	4.000	0.710	0.535
Fast-growing	37	3.000	0.514	0.450

Weir and Cockerham's F-statistic θ (F_{ST}) was 0.001 and f (F_{IS}) was -0.219 suggesting that there is little differentiation between the slow- and fast-growing groups and that there is an excess of heterozygotes. A Hardy-Weinberg exact test for heterozygote excess revealed that this heterozygote excess was significant in the slow-growing group ($p = 0.0066$) and the intron 1 locus overall ($p = 0.0112$). Results of the general Hardy-Weinberg exact test are presented in

Table 3.5. The slow-growing group shows a significant deviation from HWE ($p = 0.0207$) while the fast-growing group and intron 1 locus conform to the equilibrium ($p > 0.05$). Intron 1, while not statistically significant, is trending towards a significant deviation from HWE ($p = 0.0693$). The Fisher's exact test reported no significant difference in genotype count between the slow- and fast-growing groups ($p = 0.112$). Comparison of the three alleles common to both slow- and fast-growing groups showed no significant differences in allele frequency between the two groups (485, $p = 0.70$; 550, $p = 0.86$; 765, $p = 0.34$).

Table 3.5. Significance values for exact tests assessing deviation from HWE by sample site and locus in the *GH* intron 1 locus in slow- and fast –growing *C. auratus*. * = significant p-value ($p < 0.05$).

Population	**P-value**
Slow-growing	0.0207*
Fast-growing	0.6257
Locus	**P-value**
Intron 1	0.0693

These three alleles occurred in all populations, wild and cultured, and differed in their frequencies (Table 3.6). The 485 allele is found in its highest frequency in the Bay of Islands and its lowest frequency in the fast-growing group. The 550 allele is found in its highest and lowest frequencies in the fast-growing group and the Bay of Islands respectively. Finally, the 765 allele is at its highest frequency in Doubtless Bay and its lowest in the fast-growing group. Therefore, the fast-growing group has the lowest frequencies of alleles 485 and 765, and the highest frequency of allele 550. Importantly, the 485 allele retained a low frequency in all populations and was never detected as a homozygote. Results of chi-squared tests assessing differences in allele frequencies of all populations, wild and cultured, also found no significant differences between populations (485, $p = 0.67$; 550, $p = 0.96$; 765, $p = 0.37$).

Table 3.6. *caGH* intron 1 allele frequencies of common alleles in all populations and groups (BAY, Bay of islands; DOU, Doubtless Bay; KAI, Kaipara; PNG; Parengarenga; RAN, Rangaunu; Slow, slow-growing; Fast, fast-growing). Bold values indicate highest frequency for each allele and italicised values indicate lowest frequencies.

Allele	**BAY**	**DOU**	**KAI**	**PNG**	**RAN**	**Slow**	**Fast**
485	**0.125**	0.050	0.100	0.075	0.105	0.065	*0.041*
550	*0.425*	0.475	0.550	0.475	0.526	0.597	**0.689**
765	0.350	**0.475**	0.325	0.350	0.342	0.323	*0.270*

3.4. Discussion

3.4.1. Assessment of *caGH* intron 1 and 3 repeat region variation

In Chapter Two, large repeat regions were identified in the first and third introns of the *caGH* gene. This investigation aimed to assess the genetic variability at these loci and test for associations between these markers and growth rate in *C. auratus*. 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. The amplification and genotyping of the intron 1 locus in this study enabled the methodologies to be optimised and provide new information on the level of variation at the *GH* locus in wild populations of *C. auratus* populations.

Investigation of the intron 3 repeat region was discontinued due to unresolvable issues with amplification and genotyping at this locus. Nevertheless, some of the preliminary assays were able to give an estimate of the minimum level of variation at the intron 3 locus. This locus has previously been identified as a marker that is correlated with growth in studies of the closest relative species, *P. major* (Sawayama & Takagi, 2015b). Therefore, it merits further investigation; however, the current technical challenges will need to be resolved. Further investigation would likely require direct DNA sequencing of the amplified products. This would be difficult as common methods for long DNA fragments such as primer walking and shotgun sequencing would be impractical given the repetitive nature of the sequence. An alternative method could be used that has been specifically designed for sequencing long stretches of repetitive DNA, and involves the combined use of a linear vector to stabilize the cloning process, and the use of exonuclease III for obtaining progressive deletions of simple sequence repeat (SSR) rich fragments (De Bustos, Cuadrado, & Jouve, 2016).

3.4.2. Testing for an association between *GH* intron 1 repeat region variation and growth rate in *C. auratus*

Genetic variation in the *GH* intron 1 repeat region was assessed in slow- and fast-growing *C. auratus* in order to determine if the locus is related to the growth phenotypes in each group. The slow- and fast-growing snapper groups had three and four alleles respectively with three of those alleles occurring in both groups. Weir and Cockerham's F-statistic θ (F_{ST}) revealed little differentiation between slow- and fast-growing groups and the f (F_{IS}) and an exact test for

heterozygote excess revealed there to be an overall excess of heterozygotes in these groups. This high level of heterozygosity was identified in both the wild populations in part one and the slow- and fast-growing groups in part two and will be further discussed below in a comparison of the results from the wild and cultured *C. auratus*. Hardy-Weinberg exact tests revealed that the slow-growing group was significantly deviated from HWE but that the fast-growing group was not. This may be due to the higher level of heterozygosity observed in the slow-growing group compared to the fast-growing group. Overall, the intron 1 locus did not significantly deviate from HWE ($p = 0.069$). Fisher's exact tests to determine differences in genotype counts and allele frequencies between the two groups revealed no significant difference at the 5% significance level. However, a 0.05 cut off is an arbitrary threshold (Greenland et al., 2016). The p values from tests for overall HWE deviations and differences in genotype distribution in this study (0.069 and 0.112 respectively) were relatively close to the 5% significance level.

These results suggest that this locus needs further assessment as a potential marker for growth. The sample sizes for the slow- and fast-growing groups were small because of limitation on the size of the experimental system that was designed for a separate study. With larger sample sizes, it is likely that statistical tests will reveal more definitive results. Growth rates are also known to vary throughout development in *C. auratus* (Francis, 1994). The age of the individuals may have affected the results depending on whether the groups were in a 'growth spurt' phase when sampled. *GH* is known to be expressed at different rates throughout development (Fuentes et al., 2013). This may have led to different groups of individuals being classed as slow- or fast-growing if the sampling of these groups had occurred at a later life stage. This effect was taken into account in *GH* marker investigations in a study of *P. major* by sampling individuals 50 and 200 days post-hatch (Sawayama & Takagi, 2015b). Future investigations into *GH* markers in *C. auratus* could apply a similar approach.

3.4.3. Comparison between the genetic variation of the *GH* intron 1 locus in wild-caught and cultured *C. auratus* populations

Across all populations in this study, there were three alleles (485, 550, and 765) found in all sample sites and growth groups. Of highest frequency were the alleles 550 and 765. These varied in frequency across all groups but the differences between groups were not statistically significant. Interestingly, the fast-growing group exhibits the highest frequency of the shorter allele (550) and the lowest frequency of the longer allele (765) when compared to all other

populations including the slow-growing group. Therefore, the slow-growing group has a lower frequency of the short allele (550) and a higher frequency of the long allele (765) when compared to the fast-growing group. This phenomenon has also been observed in the first intron of *GH* in a close relative of *C. auratus*, *S. aurata.* Almuly et al. (2008) showed that longer repeat regions in the first intron were associated with repressed gene activity of *GH*. The reduced expression of gene variants with long introns has also been observed in the first intron of the other genes such as the Friedreich ataxia gene (Ohshima et al., 1998). Repeat regions in non-coding regions such as these 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 is evident that these non-coding regions are important for gene expression; however, the mechanism to explain how the length of an intron may affect gene expression is still largely unknown. One hypothesis predicts that due 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). Additionally, the core repeat sequence of the *GH* intron 1 repeat in *C. auratus* contains consensus sequences for important transcription factors TRE and a GAGA factor. TRE binding within the first introns of other genes has shown to directly influence the expression of those genes (Fu et al., 2006; Plateroti et al., 2006). GAGA-factors are known to regulate transcription at multiple levels (Lomaev et al., 2017) and are also likely to affect the expression of the *GH* gene. 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 and help to explain how the length of this intron may affect growth. In this study, the frequency of the short allele (550) was higher in the fast-growing group while the frequency of the long allele (765) was higher in the slow-growing group. Although there was no significant difference in allele frequencies between these groups, this observation is consistent with the intronic length mechanism. Based on these results, and those of other studies, it is possible to suggest that the long allele (765) may be associated with the reduced growth rate in the slow-growing *C. auratus* in this study. With a larger sample size, this hypothesis could be further investigated.

Another finding that deserves further consideration is the high level of heterozygosity in all *C. auratus* populations in this study. There are several potential explanations for heterozygote excess including non-specific DNA amplification, negative assortative mating, outbreeding, sampling error, sample size, sex linkage, and selection (Stevens, Salomon, & Sun, 2007; Waples, 2014). One or more of these may occur in the study populations. It is possible that the high level of heterozygosity is the result of non-specific DNA fragment amplification. However, if genomic loci other than the targeted *GH* gene were being amplified, it would be expected that these PCR products would appear in all individuals, which was not observed in this study. Furthermore, some alleles that were found as homozygotes (550 and 765) were sequenced in Chapter Two and confirmed as *GH* intron 1 sequences. The 485 allele was unable to be sequenced by the same method as it was never observed as a homozygote. Sequencing of the 485 allele in the future would involve manually cutting the PCR product out of the agarose gel for sequencing. Due to a whole genome duplication event early in teleost evolution (Glasauer & Neuhauss, 2014), it is possible that multiple copies of the *GH* gene exist in *C. auratus* and may also be undergoing PCR amplification. Multiple copies of *GH* genes have been identified in other species such as salmon (McKay et al., 2004) and tilapia (Ber & Daniel, 1993). However, when a BLAST search was carried out between the *C. auratus* genome and the *pmaGH* gene, only a single locus in the *C. auratus* genome was identified as a match (Chapter Two). Moreover, no evidence of a second *GH* gene was found in the closely related species, *S. aurata* (Almuly et al., 2000). Therefore, it is unlikely that a second copy of the *GH* gene exists in *C. auratus* and thus the excess heterozygosity is unlikely the result of the amplification of a *GH* pseudogene.

Negative assortative mating (individuals tend to mate more frequently with partners that are phenotypically different) and outbreeding may have occurred when the original aquaculture brood stock was assembled, if their source populations were very different. This is unlikely as all individuals contributing to this brood stock were sourced from the Tasman Sea. Moreover, this does not explain why the wild *C. auratus* populations also exhibit such a high level of heterozygosity. Sampling error may be important because of the small random sample of individuals for the experimental group or the fact that individuals with certain genotypes may be more likely to be sampled. The fact that the high level of heterozygosity was found in all groups despite different sampling methods for wild and cultured fish makes non-random sampling an unlikely explanation. Small sample size has also been shown to lead to heterozygote excess when allele frequencies differ between sexes (Robertson, 1965). Sex-

linked loci, where the expression of an allele is directly associated with the sex of an individual (and therefore sex chromosomes), can also cause heterozygote excess (Waples, 2014). In species with two sex chromosomes, individuals of the heterogametic sex will have one copy of a sex-linked gene and individuals of the homogametic sex will have two copies. When there is random mating and there are no other disturbing forces acting on a population, the genotype frequencies of homogametic individuals will be in HWE. When this is not the case, the homogametic sex will often show a heterozygote excess (Waples, 2014). There is some evidence to suggest *GH* may be sex-linked in some organisms (Forbes, Knudsen, North, & Allendorf, 1994); however, there is currently no evidence to suggest that *C. auratus* has sex chromosomes. Furthermore, investigation of other fish from the Sparidae family did not detect any evidence of sex chromosomes (Devlin & Nagahama, 2002).

In this study, the 485 allele was never detected as a homozygote. This could be caused by not sampling enough individuals to identify a 485/485 homozygous individual. However, with an overall sample size of 167 individuals, it would be expected that at least one homozygote would be sampled. Another possibility is overdominance selection, which would maintain a high number of advantageous heterozygotes in the population (Parsons & Bodmer, 1961). The number of heterozygotes containing the 485 allele is relatively low in all populations compared to other heterozygotes (e.g. 550/765) which does not provide strong support for selection as the main force acting on the allele. Negative selection could be acting on the 485 homozygote. The complete absence of homozygotes for this allele suggests that these individuals may not survive. Lethal alleles may also produce non-lethal, but low fitness phenotypes as heterozygotes. For example, Achondroplasia, the most common form of short limb dwarfism in humans, is caused by a mutation in the fibroblast growth factor receptor 3 (*FGFR3*) gene. In heterozygous form, the mutation causes the detrimental growth phenotypes of the disease, however when in homozygous form, the mutation causes death very early in development (Horton, Hall, & Hecht, 2007). If the 485 allele from this study has detrimental effects in heterozygotes the allele may be in the process of being purged from these populations which could explain the low frequencies. Though, it would be expected that if this were the case, it would also be reflected in the results of the Hardy-Weinberg exact test. A larger sample size may be more likely to show deviations from HWE at this locus. Furthermore, the overall low frequency of the 485 allele in *C. auratus* may have affected the probability of detecting homozygote individuals in this study and therefore further analysis with larger sample sizes is recommended. One other observation that might support the lethal allele hypothesis is that the

485 allele frequency is higher (although not significantly) in the slow-growing group than the fast-growing group (0.065 and 0.041 respectively). The slow-growing group also had the highest level of heterozygote excess. This appears to be consistent with the hypothesis that the allele has negative effects (slow growth) when in heterozygous form. The 485 allele is much shorter than the other detected alleles in this study and may have arisen through a deletion-type mutation. Large deletions in the first introns of other genes have been shown to cause detrimental consequences on gene function and health (Jansen, Cook, Song, & Park, 2000; Rahkonen et al., 2004). As previously discussed, the first introns of *GH* genes are likely to affect its expression and consequently this is a plausible reason for the possible detrimental effects of the 485 allele. *GH* has been linked to developmental processes in several species (Cao, Zhou, Wei, Li, & Gui, 2011; Chen et al., 2015; Jing et al., 2015) and if these processes are hindered by altered *GH* expression by this allele, it may explain how individuals could die before, or soon after hatching. Future studies might benefit from genotyping of the locus during embryonic development or immediately after hatching. Determining the underlying process that is influencing the absence of 485 homozygotes in this study could be useful when developing an aquaculture breeding programme.

3.4.3. Conclusions

This study assessed the genetic variation at a repeat region in the first intron of the *caGH* gene and searched for associations between intron 1 sequence variation and growth rate in this species. The intronic repeat region was found to be polymorphic and to conform to HWE in wild populations. The level of genetic variation at this locus was compared between slow- and fast-growing *C. auratus* groups to search for associations between variation and growth traits. The slow-growing group significantly deviated from HWE; however, overall differences in genotype and allele frequencies were largely insignificant between groups. Observations of allele frequencies suggest the long allele (765) may be associated with the reduced growth rate in the slow-growing *C. auratus*. Further analysis with a larger sample size is required. There is also the possibility that the 485 allele is experiencing negative selection. A range of promising results were reported for the development of an assay method for the *GH* gene in the *C. auratus*. If further investigations based on larger sample sizes reveal a significant association of genetic variation and growth traits, this marker would prove a useful tool for identifying individuals with high growth potential for selective breeding programmes.

Chapter Four: Growth and transcriptomic responses to temperature in *Chrysophrys auratus*

4.1. Introduction

Growth is a complex and dynamic process that is regulated by a number of genetic and environmental factors. Environmental factors are capable of enhancing or repressing the expression of genes underlying different growth traits (Danzmann et al., 2016). Temperature plays a fundamental role in poikilothermic animals such as fish as it can affect virtually all biological processes (Hochachka & Somero, 2002). Due to the effects of temperature on poikilotherms, an increasing number of studies are emerging that investigate how temperature affects important traits such as growth. These studies are also particularly useful for developing optimal conditions for rearing fish in captivity and predicting responses of fish species to climate change.

The development of high throughput, next-generation sequencing technologies has provided a novel method of transcriptome analysis called RNA sequencing (Wang et al., 2009). Recently, RNA sequencing has been used to investigate the molecular mechanisms involved in temperature adaptation and acclimation of teleost fish such as rainbowfish (Smith et al., 2013), catfish (Liu et al., 2013), chinhook salmon (Tomalty et al., 2015), rainbow trout (Tan et al., 2012) and zebrafish (Scott & Johnston, 2012). Although these studies shed light on the relationship between temperature and growth regulation, this remains a young field of research and the complex network of genes and processes controlling a fish's response to temperature are still largely undefined. To add to the complexity, fish muscle has the ability to achieve post-natal somatic growth through hyperplasia (increase in cell number) as well as hypertrophy (increase in cell size). Furthermore, fish growth can be indeterminate, occurring throughout the whole life cycle of a fish (Mommsen, 2001). These characteristics set fish apart from other vertebrates where post-somatic growth is determinate and occurs only through hypertrophy of muscle fibres formed prior to birth (Fuentes et al., 2013; Rowe & Goldspink, 1969). Therefore, fish represent a unique model for studying the fundamental mechanisms regulating growth.

The Australasian snapper, *C. auratus*, supports an important commercial fishery in New Zealand (Annala et al., 2000; Colman, 2010; Parsons et al., 2014; Paul, 1977) and methods for rearing *C. auratus* in captivity are being developed. Despite the economic importance of this

species, published genomic studies relating to growth in *C. auratus* are lacking. Researchers from the PFR Seafood Research Facility in Nelson, New Zealand have recently carried out the first broad scale transcriptomic study in *C. auratus* to investigate growth responses to high and low temperatures (M. Wellenreuther, unpublished). They have provided a list of DEGs (Differentially Expressed Genes) and their gene ontology (GO) annotations for the investigative purposes of this study.

The aims of the research presented in this chapter are:

1) To use a transcriptome dataset that was obtained from two experimental populations (high- and low-temperature treatments) of *C. auratus* to test for differences in gene expression.
2) To identify important genomic loci for growth in *C. auratus*.

4.2. Methods

The experimental setup, sampling, RNA sequencing, and production of the overall dataset were carried out by researchers at PFR for a separate project (M. Wellenreuther, unpublished). A list of DEGs and their GO annotations were provided for the investigative purposes of this study. The following methodology is a brief summary of the steps taken to produce the original dataset that was ultimately used in this thesis research.

4.2.1. Experimental setup

Experiments were carried out on two-year-old wild-caught and hatchery reared domesticated snapper (both sourced from populations in the Tasman Sea) held at the PFR Seafood Research Facility in Nelson, New Zealand. At the commencement of the experiment, 20 fish from each wild-caught and domesticated strains were weighed, measured, and moved into four 800 L tanks (2 x 10 wild-caught and 2 x 10 domesticated) and preconditioned at 17 °C for five days. One wild-caught and one domesticated treatment were exposed to a decrease of 1 °C per day while the other wild-caught and domesticated treatments were exposed to an increase of 1 °C per day until the desired temperature differential of either 13.0 °C or 21.0 °C was established (low temperature treatment, LTT, and high temperature treatment, HTT, respectively). The temperatures chosen, represent the thermal minima and maxima that *C. auratus* experience in the Tasman Sea. Once the desired temperatures were reached, the experiment was allowed to

run for 18 days. During the experiment, each treatment received the same food rations and dissolved oxygen levels were maintained at > 90%.

4.2.2. Sampling and RNA sequencing

At the end of the experiment, eight fish from each treatment were weighed and measured for identification and phenotyping, and tissue samples were taken from epaxial white muscle. Total RNA was extracted from six fish from each treatment using the Trizol LS Reagent (Life Technologies) according to manufacturer's instructions. RNA samples were individually prepared for sequencing using the Illumina Tru-Seq kit on an Illumina HiSeq 2000 sequencer (paired-end 100bp sequencing, 160 bp insert length) at the Beijing Genomics Institute Shenzhen, China.

4.2.3. Sequence data processing and de novo transcriptome assembly

Raw sequences were quality trimmed and filtered, and paired-end reads were assembled into transcripts (minimum length 200 bp) following the Trinity V2.2.0 *de novo* assembly pipeline (Haas et al., 2013) with a default k-mer size of 25-bp. The raw transcripts were filtered by open reading frames (length ≥ 300 nt), longest isoform matches and mapping rate (≥ 1 transcripts per kilobase million (TPM)). The remaining transcript sequences were searched against the Uniprot-Swissprot database. The de novo transcriptome was quality checked and each of the transcriptome filtering steps were validated statistically. For downstream expression analysis, low expression genes (log counts per million (CPM) < 1 in at least three individuals) were filtered out.

4.2.4. Differential gene expression and GO analysis

The genotype x environment interaction (GEI) was investigated using a generalised linear model (GLM) approach (temperature x genotype) with the '*glmFit*' function and a likelihood-ratio test implemented in the R package edger (Robinson, McCarthy, & Smyth, 2010). To quantify the additive effect of temperature and genotype, a GLM approach (Temperature + genotype) was used on the dataset with prior removal of genes in interaction. Genes were considered significantly expressed when false discovery rate (FDR) < 0.01 and log fold change $|\log FC| \geq 2$. GO annotations for associated DEGs were determined from the GO-basic database.

4.2.5. Methods carried out within this thesis research: Interpretation of dataset

CateGOrizer (Hu, Bao, & Reecy, 2008) was used to group DEGs and their associated GO annotations into broader hierarchical GO categories to determine the percentage of DEGs associated with each category. These were separated into the three main GO categories: biological process, molecular function, and cellular component. The literature was then surveyed to determine the identity and potential functions of each individual DEG with a BLAST match. Literature searches into genes associated with 'growth', 'temperature', and 'stress' were conducted to identify DEGs in the current study that may be involved in the growth responses to temperature in *C. auratus*.

4.3. Results

The experimental treatments showed an increase in growth rate of fish that were held at 21 °C compared to those kept at 13 °C (*see* Appendix Figure A.1). RNA-sequence data and differential expression analysis of these fish identified a total of 1396 DEGs, with 691 genes upregulated in the HTT and 705 upregulated in the LTT. A BLAST search of the 1396 DEGs was able to assign putative identities to 1251 transcripts while 145 transcripts were unable to be identified with a similar sequence in the database. The percentage of DEGs associated with certain gene ontologies (GO) in each temperature treatment are presented in Figures 4.1-4.3. These represent the three major categories of GO annotations: biological process (Figure 4.1), molecular function (Figure 4.2), and cellular component (Figure 4.3). Metabolism was the biological process with the largest number of DEGs and had a higher percentage of DEGs in the HTT than in the LTT. Other biological processes upregulated in the HTT compared to the LTT included biosynthesis, development, cell differentiation, protein metabolism, morphogenesis, cell proliferation, and protein biosynthesis (Figure 4.1). Additionally, approximately 20% of the upregulated DEGs in the HTT were associated with biosynthetic processes compared to approximately 5% for catabolism (Figure 4.1). The HTT also showed an increase in DEGs associated with carbohydrate metabolism and responses to stress compared to the LTT. Molecular functions that were increased in the HTT included binding, catalytic activity, hydrolase activity, nucleotide binding, enzyme regulatory activity, and peptidase activity (Figure 4.2). Genes associated with cytoskeleton, extracellular region, and endoplasmic reticulum (ER) cellular components were also upregulated in the HTT (Figure 4.3).

In the LTT, increased biological functions compared to HTT were those such as transport, cell communication, cell-cell signalling and response to stimuli (Figure 4.1). In the LTT the percentage of DEGs associated with biosynthesis were approximately double the number associated with catabolism (Figure 4.1). Molecular functions associated with the LTT included transporter activity, receptor binding, ion channel signal transducer activity, and kinase activity (Figure 4.2). The LTT also showed increased DNA binding, nucleic acid binding, and transcription factor binding when compared to the HTT. A number of genes associated with the plasma membrane cellular component were also upregulated in the LTT (Figure 4.3).

BLAST results from the RNA sequence data show that genes from the GH-IGF axis are being expressed in both temperature treatments. These include *GH*, *GHR*, *IGF1*, *IGF2*, and *IGFR*. These genes, however, are not differentially expressed between the two temperature treatments. Literature on individual DEGs identified in this study was surveyed and revealed a number of important genes identified to be involved with growth and specific responses to temperature in fish. These genes are outlined in Table 4.1.

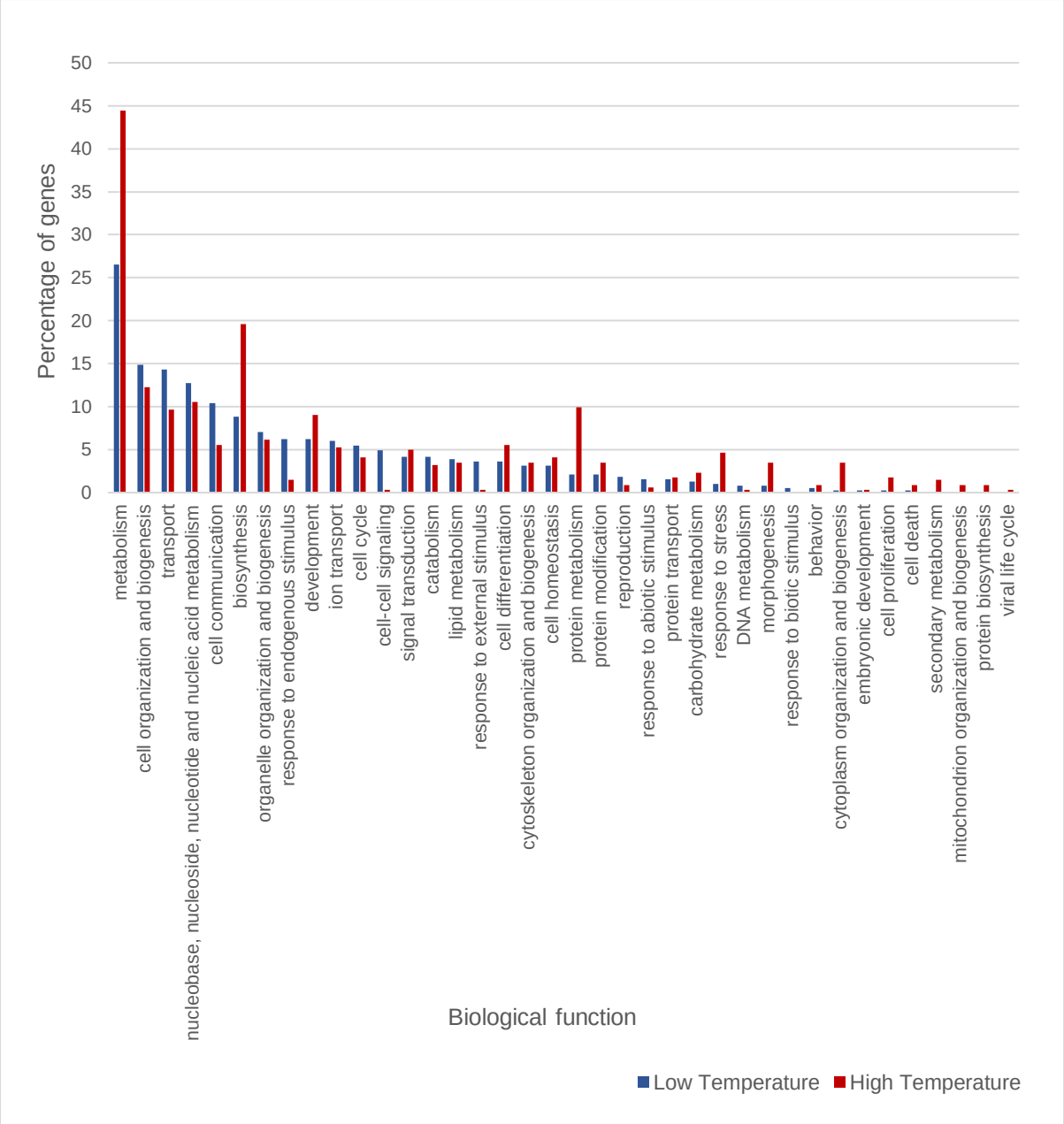

Figure 4.1. GO annotations of biological processes associated with upregulated DEGs in low (blue) and high (red) temperature groups of *C. auratus*.

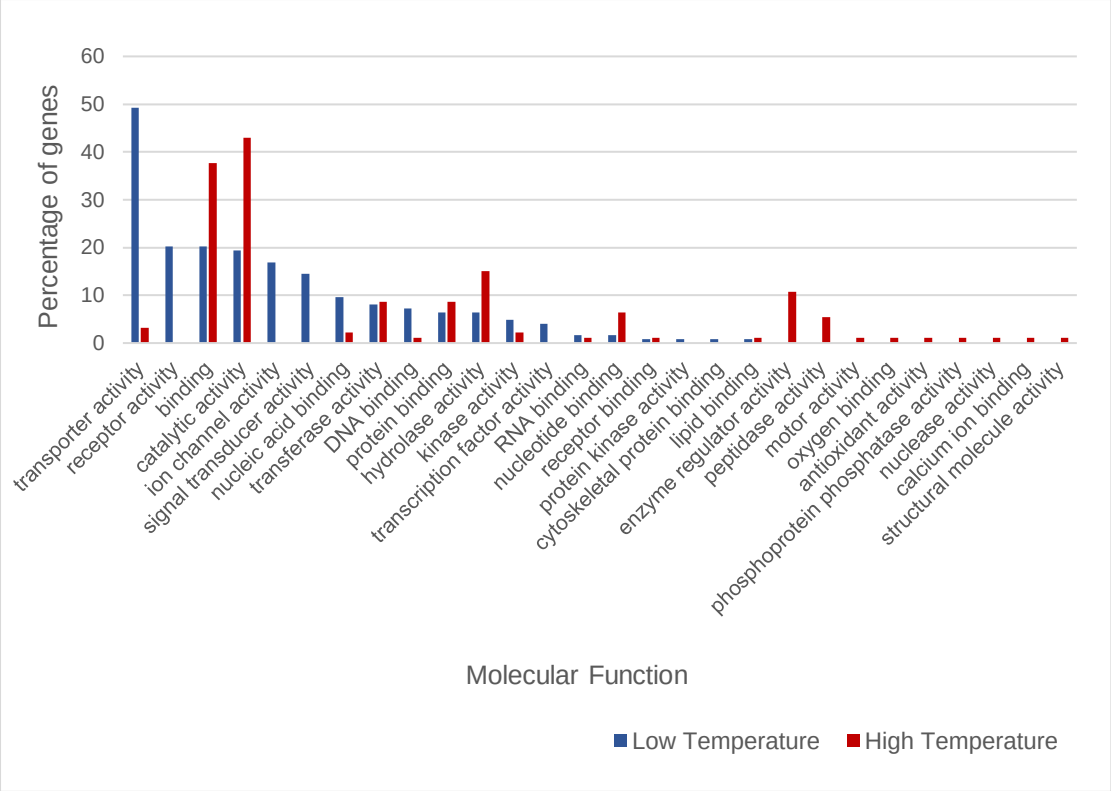

Figure 4.2. GO annotations of molecular functions associated with upregulated DEGs in low (blue) and high (red) temperature groups of *C. auratus*.

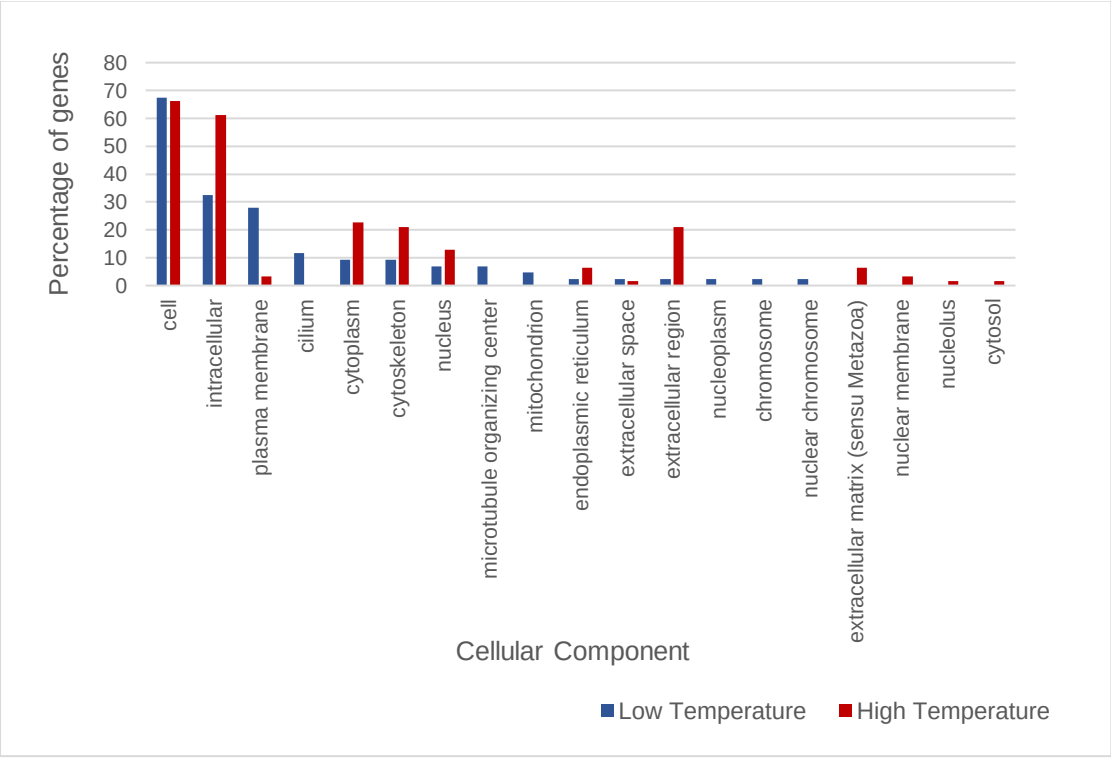

Figure 4.3. GO annotations of cellular components associated with upregulated DEGs in low (blue) and high (red) temperature groups of *C. auratus*.

Table 4.1. Differentially expressed genes in *C. auratus* identified as important for growth in response to temperature.

Gene abbreviation/group	Full name of associated protein/s	Temperature treatment (upregulated)
MYF5	Myogenic Factor 5	High
HSP90A	Heat Shock Protein 90a	High
MYL9	Myosin Light Chain 9	High
Actins (*ACTB, ACTB2, ACTC, ACTS)*	Actins	High
Collagens	Collagens	High
PCNA	Proliferating Cell Nuclear Antigen	High
FOXO1, 3	Forkhead box 01, 03	Low
FBX32/MAFBX/Atrogin-1	F-box protein 32	Low
KLF15	Kruppel Like Factor 15	Low
IGFBP1, 7	Insulin-like Growth Factor Binding Protein 1, 7	High
PHLA3	Plecktsrin Homology Like domain family A member 3	High
MURF1/TRI63	Muscle-specific Ring Finger protein 1	High
AKT	AKT Serine/Threonine Kinase 1	Low
P3C2A	Phosphatidylinositol-4 -Phosphate 3-Kinase Catalytic Subunit Type 2 Alpha	Low
P3C2B	Phosphatidylinositol-4 -Phosphate 3-Kinase Catalytic Subunit Type 2 Beta	Low
PI3R4	Phosphoinositide-3- Kinase Regulatory Subunit 4	Low
KS6B1	Ribosomal Protein S6 Kinase B1	Low
4EBP1	Eukaryotic Translation Initiation Factor 4E Binding Protein 1	Low
MAPK3	Mitogen-Activated Protein Kinase 3	Low
MK13	Mitogen-Activated Protein Kinase 13	Low
M4K3	Mitogen-Activated Protein Kinase Kinase Kinase Kinase 3	Low
IP3KA	Inositol-Trisphosphate 3- Kinase A	Low
HSP70	Heat Shock Protein 70	High
HSP90	Heat Shock Protein 90	High
Chaperones	Chaperone proteins	High and Low
ATF 3, 4, 5, 6	Activating Transcription Factors 3, 4, 5, 6	High
CHAC1	ChaC Glutathione Specific Gamma Glutamylcyclotransferase 1	High
TRAF2	TNF Receptor Associated Factor 2	
BCL2	B-cell lymphoma 2	
SOX4	SRY-box 4	High
CDN1A	Cyclin Dependent Kinase Inhibitor 1A	High
SESN2	Sestrin 2	High
HB Proteins (*HBB,HBAA, HBAB)*	Haemoglobin proteins	High

4.4. Discussion

The results from the experimental setup at PFR show that *C. auratus* exhibit a higher growth rate at 21 °C compared to 13 °C (*see* Appendix Figure A.1). The aim of this study was to determine whether there were differences in the transcriptomes of fish held at high and low temperatures and whether any differences were related to genes associated with growth. A dataset of 1396 DEGs and their GO annotations from the two temperature treatments was used to assess which genes associated with biological processes, molecular functions, and cellular components were upregulated in each temperature treatment. The most highly represented GO term was metabolism with the HTT exhibiting a higher percentage of DEGs associated with this biological process. It is well established that metabolic activity increases with temperature due to the increased rate of biochemical reactions (Brown, Gillooly, Allen, Savage, & West, 2004). Metabolic processes are commonly found to be differentially expressed in studies where fish were held at different temperatures (Kassahn, Crozier, Ward, Stone, & Caley, 2007; Smith et al., 2013; Tomalty et al., 2015). Other biological processes upregulated in the HTT compared to the LTT include biosynthesis, development, cell differentiation, protein metabolism, morphogenesis, cell proliferation, and protein biosynthesis. Growth, or hypertrophy, is recognised when anabolic processes outweigh catabolic ones (Johnston et al., 2011; Schiaffino et al., 2013). The percentage of DEGs in the HTT associated with biosynthetic processes (anabolic) is approximately four times the number associated with catabolic processes. Therefore, it is likely that the upregulation of these largely anabolic biological processes in the HTT fish is associated with growth rate. The HTT fish also exhibited upregulation of genes related to carbohydrate metabolism in comparison to the LTT. A shift toward a more carbohydrate-based metabolism in high temperatures is common in fish and represents a reallocation of energy resources to cope with higher temperatures (Kassahn et al., 2007; Windisch et al., 2014). Enhanced growth in *GH*-transgenic salmon has also been partially attributed to an increase in carbohydrate metabolism (Leggatt et al., 2009).

Fish in the HTT also had an increase in DEGs associated with several molecular functions when compared to fish at the LTT. These included binding, catalytic activity, hydrolase activity, nucleotide binding, enzyme regulator activity, and peptidase activity. These molecular functions are involved in an overall increase in catalytic activity in fish at higher temperatures which has been reported in similar temperature studies of fish (Smith et al., 2013; Tomalty et al., 2015). This result is consistent with the observed increase in DEGs associated with

metabolism in the HTT. Metabolism is a complex network of biochemical reactions, catalysed by enzymes, that leads to the regulation of substrate and product concentrations, and the rate of reactions (Brown et al., 2004). Therefore, it is logical that an increase in metabolic processes is accompanied by an increase in catalytic activity. Furthermore, metabolism is known to increase with temperature and body size (Brown et al., 2004) which is consistent with this study.

The response of fish to cold temperatures in this study was characterised by an increase in biological processes such as transport, cell communication, and cell-cell signalling. Similar patterns have been observed in other species of fish when they were exposed to cold temperatures (Liang, Chang, He, & Tang, 2015; Long et al., 2013); however the mechanisms behind their upregulation are largely unknown. Additionally, molecular functions that were upregulated in LTT compared to HTT include transporter activity, receptor binding, ion channel activity, signal transducer activity, and kinase activity. These molecular functions are consistent with the increased GO biological processes in cold temperatures in the current study. Although the response to low temperature is not well understood in fish, some evidence from other organisms suggests that these processes are induced in response to cold stress (Reed et al., 2017; Şahin-Çevik, 2013). The LTT fish also exhibited an upregulation of genes involved in the response to stimuli (endogenous, external, biotic, and abiotic stimuli) and genes associated with the cellular component, plasma membrane, compared to the HTT fish. This gives further support to the hypothesis that fish exposed to cold temperatures may upregulate pathways that recognise and initiate cellular responses to stimuli associated with cold temperatures. Furthermore, LTT fish show increased DNA binding, nucleic acid binding, and transcription factor binding. Several studies on fish responses to cold stress reported a large transcriptional response to cold which is consistent with the results of this study (Gracey et al., 2004; Windisch et al., 2014).

Cold temperatures are commonly associated with an increase in catabolic processes (Gracey et al., 2004; Liang et al., 2015; Long et al., 2013). In this study, the LTT had a higher percentage of DEGs associated with catabolism compared to the HTT. As previously stated, growth occurs when anabolic processes outweigh catabolic processes. Conversely, atrophy occurs when catabolism outweighs anabolism (Johnston et al., 2011; Schiaffino et al., 2013). In the LTT the percentage of DEGs associated with biosynthesis is approximately twice the number associated with catabolism. When compared to the HTT, where the number of DEGs associated with

biosynthesis were approximately four times that of catabolism, it is clear that this balance between growth and atrophy has shifted between the temperature treatments. In warm temperatures, biosynthetic processes are highly favoured over catabolic processes, while in low temperatures the balance between biosynthesis and catabolism is somewhat more equally balanced. This is consistent with the differences in growth rate observed between the two temperature treatments.

A closer inspection of the identities of the transcripts enabled a more in-depth assessment of the cellular processes leading to the different growth rates between high and low temperatures. In this study, RNA sequence data was retrieved from muscle tissue of *C. auratus*. Some key regulators of vertebrate growth, *GH*, *GHRs*, *IGFs*, and *IGFRs* were expressed in fish from both temperature treatments in this study, however, these genes were not differentially expressed between the HTT and LTT. Although differential expression of these genes has been observed in muscle tissue of fish suggesting direct action of GH on target tissues, the relative importance of direct and indirect GH pathways in fish is unresolved (Björnsson et al., 2002). In the indirect route, GH is secreted from the anterior pituitary and travels to the liver where it promotes the production of IGFs through GHR binding and subsequent signalling. From the liver, IGFs travel to their target tissues and initiate anabolic and growth promoting signalling pathways through binding of IGFRs (Moriyama et al., 2000). Therefore, the lack of differential expression of some of these genes in the muscle tissue found in this study is not indicative of a lack of differential function of the GH-IGF axis. Some evidence suggests that temperature promotes growth through IGF1 secretion from the liver in response to GH binding and that the autocrine/paracrine expression of IGFs in muscle tissue is not a key regulator of growth in response to temperature (Gabillard et al., 2005; Gabillard et al., 2003a). Further transcriptomic studies in multiple tissue types including the pituitary and liver would provide a deeper understanding of the complex and multifaceted regulation of growth in *C. auratus* and other teleost fish.

A literature survey of the identities and functions of the DEGs in *C. auratus* led to the identification of a number of genes with potential roles in growth rate. In particular, one of the major transcription factors for vertebrate myogenesis, myogenic factor 5 (MYF5), was upregulated in the HTT compared to the LTT. In myogenesis, *MYF5* is expressed in activated cells marking the beginning of biogenesis (Watabe, 1999). It is well established that expression of *MYF5* is an important myogenic factor and is commonly expressed in vertebrates with

improved growth (Fuentes et al., 2013). The expression of *MYF5* has been shown to increase in response to GH and IGF (Garikipati & Rodgers, 2012; Jiménez-Amilburu et al., 2013), further suggesting the possibility of indirect GH action on muscle tissue in this study. MYF5 is also responsible for the expression of another DEG that is upregulated in the HTT, heat shock protein 90A (*HSP90A*), which is required for correct sarcomere assembly in muscles (Du, Li, Bian, & Zhong, 2008; Hawkins et al., 2008). The HTT was also characterised by an upregulation of genes for structural components of muscle. Sarcomeric assembly during the later stages of differentiation is characterised by expression of myosin light and heavy chains, and actins (Fuentes et al., 2013; Johnston, 2006). In the HTT, myosin light chain 9 (*MYL9*) and several actin genes (*ACTB, ACTB2, ACTC, ACTS*) were upregulated in comparison to the LTT. The expression of a large number of collagen chains was also upregulated in the HTT. Upregulation of collagen chain proteins has been reported in fast-growing rainbow trout (Danzmann et al., 2016) and has shown increased expression induced by GH action in humans (Sjogren et al., 2007). This may explain the increased number of DEGs associated with cytoskeleton and extracellular region GO cellular components in the HTT. Therefore, the increased growth rate of *C. auratus* in HTT appears, logically, to be supported by an increase of structural components of muscle tissue. Fish muscle has the unique ability to grow through hyperplasia (cell proliferation) as well as by hypertrophy (cell enlargement) (Mommsen, 2001). The upregulation of the gene for proliferating cell nuclear antigen (*PCNA*) a marker for proliferating cells (Johnston, 2006), suggests the potential role of hyperplasia in the growth of *C. auratus* under high temperatures. IGF has also been implicated in the increased expression of *PCNA* (Bower & Johnston, 2010).

In the LTT, several well-known genes involved in atrophy were upregulated relative to the HTT. Of particular interest are the FOXO transcription factors. Members of the FOXO transcription factor superfamily are known for their regulation of cellular differentiation, growth, survival, cell cycle, metabolism, stress and tumour suppression pathways (Zhang, Tang, Hadden, & Rishi, 2011). In this study, two FOXO genes, *FOXO1* and *FOXO3*, were upregulated in the LTT. The overexpression of these genes is associated with atrophy (Kamei et al., 2004; Sandri et al., 2004). In particular, *FOXO3* expression is known to lead to atrophy by inducing autophagic/lysosomal and proteosomal degradation pathways (Zhao et al., 2007). One way that FOXOs induce these degradation processes is to activate the expression of atrophy genes, or atrogenes (Fuentes et al., 2013; Schiaffino et al., 2013). One of the best studied atrogenes is F-Box Protein 32/Muscle Atrophy F-Box Protein (*FBX32/MAFBX*). This

gene is an E3 ubiquitin ligase implicated in proteosomal degradation (Bodine, Latres, et al., 2001) and is also upregulated in the LTT in this study. Furthermore, kruppel like factor 15 (*KLF15*) another important gene involved in atrophy, is upregulated in the LTT. In coordination with FOXO1, KLF15 upregulates the expression of atrogenes such as *FBX32/MAFBX* (Pinheiro & Guimarães-Ferreira, 2015; Shimizu et al., 2011). Upregulation of *FOXO1, FOXO3, FBX32/MAFBX* and *KLF15* suggests the induction of degradation pathways is at least partially involved in the reduced growth rate of *C. auratus* in the LTT.

The upregulation of genes involved with biosynthetic processes in the HTT and the upregulation of genes involved in degradation pathways in the LTT offer an explanation of the differential growth rates seen in the fish in this study. The DEGs and associated GO terms that were identified suggest that the differential growth rates of *C. auratus* under different temperature treatments are, in part, due to shifts in the anabolism-catabolism balance. The increased growth rate of *C. auratus* in high temperatures is associated with a strong shift towards biosynthetic processes and growth. In comparison to the fast-growing *C. auratus*, fish in the LTT have a more balanced relationship between biosynthesis and anabolism with increased degradation processes. The LTT fish in this study experienced a very low growth rate, with some fish suffering a loss of condition (*see* Appendix Figure A.1). This is consistent with a more balanced relationship between growth and atrophy.

The literature survey of DEG identities also yielded some unexpected results. In the HTT, two IGF binding proteins were upregulated, *IGFBP1* and *IGFBP7*. IGFBPs are implicated in the positive and negative regulation of IGF function and are important in the regulation of GH signalling (Fuentes et al., 2013). The *IGFBP* genes upregulated in the HTT in this study have been identified as negative regulators of IGFs. *IGFBP1* and *IGFBP7* have been shown to inhibit growth signalling by competing with IGFs for binding at IGF receptors (Evdokimova et al., 2012; Jansson, Andersson, Uhlén, Nilsson, & Kördel, 1998; Kajimura & Duan, 2007). Furthermore, the gene for a known repressor of AKT activity, Plecktsrin Homology Like Domain Family A Member 3 (*PHLA3*) (Kawase et al., 2009), is also significantly upregulated in HTT fish. AKT is one of the main players in IGF signalling, and has an important role in promoting growth (Fuentes et al., 2013; Schiaffino et al., 2013). Thus, the upregulation of *PHLA3* in the HTT suggests further negative regulation of growth. In addition to this apparent negative regulation of growth in the HTT, there was also an upregulation of muscle specific ring finger protein 1/tripartite motif containing 63 (*MURF1/TRI63*), a well-known ubiquitin

ligase associated with atrophy (Bodine, Latres, et al., 2001). This gene is one of the important atrogenes that is usually expressed in coordination with *FBX32/MAFBX* after activation by FOXOs and KLF15 (Pinheiro & Guimarães-Ferreira, 2015; Shimizu et al., 2011). Therefore, reports of upregulation of this gene in fast-growing animals are rare in the literature. One study offers a possible explanation for the upregulation of *MURF1/TRI63* in the HTT of the current study. Consitt, Saneda, Saxena, List, and Kopchick (2017) demonstrated an increase in *MURF1/TRI63* in mice that were chronically overexpressing *GH*. As *GH* expression has been shown to increase with temperature in some fish species (Gabillard et al., 2003b; McCormick, Moriyama, & Björnsson, 2000; Ricordel, Smal, & Le Bail, 1995), it is plausible that *C. auratus* in the HTT are overexpressing pituitary *GH* during their chronic exposure to high temperatures leading to an increase in *MURF1/TRI63* and subsequent degradation pathways in muscle tissue. The GH-IGF axis is known to be subjected to multiple negative feedback loops (Backeljauw & Hwa, 2016). If chronic exposure to heat is in fact leading to the long-term overexpression of pituitary *GH* in *C. auratus*, it is possible that these negative feedback loops have also been upregulated. This could explain the upregulation of genes such as *IGFBPs* and *PHLA3* in the HTT.

There was also a relative downregulation of several important IGF signalling proteins in the HTT fish compared to the LTT. IGF has two major intracellular signalling pathways, the PI3K/AKT pathway and the MAPK/ERK pathway (Glass, 2003). Induction of the PI3K/AKT pathway leads to protein synthesis (Rommel et al., 2001), myoblast differentiation (Coolican et al., 1997), and muscle hypertrophy (Bodine, Stitt, et al., 2001). This pathway regulates these processes by activation of MTOR, which when activated, regulates the KS6B1, and 4EBP1 (Glass, 2003). *KS6B1* and *4EBP1* genes were both downregulated in HTT compared to the LTT. In the HTT, fish experienced a downregulation of *AKT*, and the following *PI3K* genes; phosphatidylinositol 4 phosphate 3 kinase catalytic subunit type 2 alpha and beta (*P3C2A*, *P3C2B*) and phosphoinositide 3 kinase regulatory subunit 4 (*PI3R4*). Furthermore, *KS6B1* and *4EBP1* were also downregulated in the HTT relative to the LTT. The second major IGF signalling pathway, MAPK/ERK, stimulates cell proliferation (Coolican et al., 1997) and terminal differentiation (Li & Johnson, 2006) in mammal muscles. In the HTT, mitogen-activated kinase 3 (*MAPK3*), mitogen-activated kinase 13 (*MK13*), and mitogen-activated kinase kinase kinase kinase 3 (*M4K3*) were downregulated. IGF also stimulates several other minor signalling pathways such as the phospholipase C (PLC)/Inositol-Trisphosphate 3 (IP3) pathway (Valdés et al., 2013). Inositol-Trisphosphate 3 Kinase A (*IP3KA*), an IP3 kinase was

also downregulated in the HTT relative to the LTT. The downregulation of such a large number of genes involved in IGF signalling provides further support for the presence of negative feedback on GH-IGF signalling in the HTT fish. Further investigations of transcriptomes in other tissues such as the pituitary and the liver are required to determine if this apparent negative feedback on growth is a result of chronic overexpression of *GH* under high temperatures or is caused by other factors.

Growth rates can also be negatively affected by stress as it tends to cause a reallocation of energy and resources away from investment activities such as growth (Wendelaar Bonga, 1997). The biological process 'response to stress' was increased in the HTT compared to the LTT suggesting increased levels of stress at the higher temperature. The HTT experienced a significant upregulation of a large number of heat shock proteins (HSP) and other chaperones. The role of chaperones is to refold and stabilize proteins to their correct conformations and are expressed under a range of normal and stress conditions (Hartl, 1996). HSPs are expressed in response to a wide range of biotic and abiotic stressors and are commonly referred to as stress proteins (Iwama, Thomas, Forsyth, & Vijayan, 1998). In particular, HTT fish upregulated the expression of *HSP70*, the most widely studied HSP which is commonly recognised as a biomarker for stress (Ryan & Hightower, 1996). Another well studied HSP upregulated in HTT fish is *HSP90* which is commonly expressed during heat shock experiments. It has also been shown to increase in killifish under constant high temperature stress (Podrabsky & Somero, 2004) and has been associated with the upper temperature tolerance in arctic charr (Quinn, McGowan, Cooper, Koop, & Davidson, 2011). Upregulation of a large number of HSPs and other chaperones, as in the HTT fish in this study, is related to a process called the unfolded protein response (UPR). The UPR is activated under ER stress under unfolded protein load (Hetz, 2012). The UPR is a combination of several complementary adaptive mechanisms to deal with the excess of unfolded proteins (Hetz, 2012). One of these mechanisms requires the action of activating transcription factors (ATF), particularly ATF 4 and 6 (Hetz, 2012; Ishikawa, Taniguchi, Okada, Takeda, & Mori, 2011). In the HTT, fish upregulated the expression of four ATF genes, *ATF3, ATF4, ATF5,* and *ATF6,* which have all been identified in UPR and stress signalling (Chen, Wolfgang, & Hai, 1996; Hetz, 2012; Wang, Lin, & Zhang, 2007). The upregulation of genes associated with the ER GO cellular component in the HTT may be explained by the upregulation of the UPR. Activation of the UPR usually results in the expression of UPR target genes such as chaperones to deal with reversible protein damage; however above a certain threshold of unresolved ER stress, the UPR can induce apoptosis

(Hetz, 2012). In the HTT, fish upregulated the expression of ChaC Glutathione Specific gamma-glutamylcyclotransferase 1 (*CHAC1*) and TNF receptor associated factor 2 (*TRAF2*). *CHAC1* and *TRAF2* are involved in proapoptotic signalling in response to the UPR (Hetz, 2012; Mungrue, Pagnon, Kohannim, Gargalovic, & Lusis, 2009). *TRAF2* is also an important protein in nuclear factor-κβ signalling which has been identified as a stress responsive pathway (Mercurio & Manning, 1999). Furthermore, downregulation of B-cell lymphoma 2 (*BCL2*), has been implicated in the induction of apoptosis in response to the UPR (Tabas & Ron, 2011) and this downregulation of *BCL2* was observed in the HTT fish. Together, these DEGs suggest the HTT fish are upregulating the UPR in response to high temperatures indicating a certain level of stress in these fish.

A number of genes upregulated in the HTT are known to respond to DNA damage including Sex Determining Region-box 4 (*SOX4*) (Pan et al., 2009) and cyclin dependent kinase inhibitor 1A (*CDN1A*) (Karimian, Ahmadi, & Yousefi, 2016) which have been identified in a similar study of fish under heat stress conditions (Tomalty et al., 2015). High temperatures can limit the dissolved oxygen in water and can eventually put stress on the oxygen transport system in fish (Pörtner, 2001). Sestrin 2 (*SESN2*), a gene commonly expressed in response to oxidative stress (Budanov, 2011), was upregulated in HTT fish. A number of haemoglobin subunit genes (*HBB, HBAA,* and *HBAB*) were also upregulated in HTT fish. The upregulation of haemoglobin genes is a common response to oxidative stress and works to maintain sufficient oxygen transport under oxidative stress (Alves et al., 2010; Feng, Liu, et al., 2014).

Although *C. auratus* exhibited an increased growth rate at 21 °C, it is clear that they were also experiencing some level of stress. The apparent negative regulation of growth shown through downregulation of IGF signalling pathways, and the potential detrimental effects of stress on growth in fish in the HTT suggest that although *C. auratus* growth rates increased at 21 °C, this is unlikely to be the optimal temperature for growth. Growth rates may increase at a lower temperature where heat stress and negative feedback on growth are less of an influence. The experimental design of this study did not aim to determine the optimal temperature for *C. auratus* growth rate; instead, it was to investigate how the transcriptomes change in response it high and low temperatures. However, the findings of this study also offer important information for identifying optimal growth conditions in the aquaculture of this species. To further investigate optimal growth temperatures in *C. auratus*, an experimental design with

several more temperature treatments between the 13 °C and 21 °C used in this study is required. Identification of optimal temperatures for growth of *C. auratus* in aquaculture would greatly increase the efficiency of the aquaculture improving the commercial interest in this species for aquaculture.

In this chapter, the regulation of growth was investigated using information on the expression of individual genes. However, genes do not work in isolation and an observed phenotype such as growth rate is likely to be the result of multiple gene-gene interactions in complex pathways or networks (Dixon, Costanzo, Baryshnikova, Andrews, & Boone, 2009). Furthermore, transcription of a gene does not necessarily translate to the expression or action of that protein. Therefore, in order to build on the research in this chapter and gain a more holistic understanding of the genetic basis of growth traits, future research should include pathway and network analyses.

4.4.1. Conclusions

This study used transcriptome datasets for high and low temperature treatments of *C. auratus* and aimed to test for differential gene expression that correlated with growth rate. The overall goal was to identify key processes and genes contributing to differential growth phenotypes in response to temperature. The difference in growth rates of *C. auratus* between high and low temperatures appears to be regulated by the opposing processes of biosynthesis and catabolism and, to some extent, by responses to temperature stress. Increased growth rates in the HTT may be influenced by indirect action of the GH-IGF axis. Transcriptome analysis of other tissues is required to confirm this. Evidence suggests that *C. auratus* in the HTT experienced some level of negative regulation of growth by a GH-IGF feedback loop as a result of heat stress. This study also identified a number of genes that are putatively associated with growth in this species. The findings of this study contribute to a better understanding of the regulation of growth in *C. auratus* and other teleost species. It also provides useful information for developing the optimal growth conditions in *C. auratus.*

Chapter Five: General discussion

5.1. Overview

Growth is a complex trait controlled primarily by the GH-IGF axis, though, it can also be influenced by a multitude of physiological and environmental factors (Fuentes et al., 2013; Moriyama et al., 2000). In particular, temperature is considered a highly influential factor affecting growth of poikilotherms such as fish (Hochachka & Somero, 2002). A primary objective for aquaculture selective breeding programmes is to improve growth rate, as a quicker growth rate will reduce the amount of time required to raise fish and achieve market size earlier, lowering the cost per unit (Gjedrem et al., 2012). While being important for decreasing the production time, improving growth rate has also been associated with better feed conversion rates (Thodesen et al., 1999) and disease resistance (Overturf et al., 2010; Rye et al., 1990) which are characteristics that also enhance production (Gjedrem et al., 2012). The application of genetic marker technologies to selective breeding programmes has been successful for improving production efficiency in terrestrial livestock species; however, adoption of these techniques has been much slower in aquaculture (Gjedrem & Robinson, 2014). This thesis investigated the genes that are typically associated with fish growth in a commercially important teleost species, *C. auratus*, which has the potential to become a successful aquaculture species in New Zealand.

5.2. Identification and characterisation of the *GH* gene in *C. auratus* and other Perciform species

The *caGH* was identified as approximately 5,577 bp in length and is comprised of six exons and five introns. The *GH* gene structure is shared with a large number of fish species (Almuly et al., 2000; Chen et al., 2004) and the similarity is conserved across all Perciform species that were analysed in Chapter Two. Two of the outgroup species, *P. olivaceus* and *O. keta*, also have this conserved this gene structure; however, the third outgroup, *C. batrachus*, has five exons and four introns. This is consistent with the current view that an intron was inserted into what was originally exon V leading to the creation of a fifth intron and sixth exon. It is suggested that this insertion took place only within the ray-finned fish, after the divergence of the Cypriniforms but before Salmoniformes, Perciforms, and Terradontiforms (Moriyama et al., 2006; Zhang et al., 2009), which is consistent with the findings of the current study. The evolutionary conservation of the *GH* gene made it useful for phylogenetic analysis of the

Perciformes order. The relationships between species and the monophyly of families are largely consistent with other studies (Betancur-R et al., 2017; Wei et al., 2014). However, as the Perciform order is the most diverse vertebrate order (Nelson, 2006), relationships among families are still difficult to define.

Within *C. auratus, GH* exon sequences are highly conserved while the non-coding regions vary considerably. This high level of intraspecific variation in non-coding regions of *GH* is also found in a range of other species (Almuly et al., 2008; Arango et al., 2014; Nie et al., 2005; Quéré et al., 2010; Zhang et al., 2009). This pattern also holds for interspecific variation. Despite high variability in intron sizes in the species used in the phylogenetic analysis, the exon lengths were largely conserved. In particular, the lengths of exons II-IV were identical for all Perciform species with the exception of *C. auratus* and *P. major* which had a 3 bp deletion in exon III. The assessment of amino acid sequences of fish in this study showed that while most Perciform species have 204 amino acids, *C. auratus* and *P. major* have 203, having lost an amino acid at this locus. This amino acid deletion appears to have been tolerated because of additional changes in surrounding amino acids such that the polarity, hydrophobicity, and consequently, the secondary structure at this locus, remain unchanged in *C. auratus* and *P. major* relative to other Perciforms. The overall patterns of polarity, hydrophobicity, and secondary structure for GH are highly conserved across all of the perciform fish analysed in this study and there was little difference from the outgroup species. This conservation of the *GH* structure throughout Perciform evolution suggests, not only common ancestry, but also selection for this architecture to ensure protein functionality (Pál et al., 2006; Taylor et al., 2004). This reinforces the importance of GH function in fish. Furthermore, this high level of sequence and structure conservation allows for the simplified identification of *GH* genes in non-model species through comparative genetic approaches.

5.3. Testing for associations between *caGH* polymorphisms and growth rate

Two large repeat regions were identified in the first and third introns of *caGH*. Repeat regions are commonly found within the promotor and 5' proximal introns of genes and have been shown to affect gene expression (Chorev & Carmel, 2012; Li et al., 2002) and the expression of phenotypic traits (Li et al., 2004). This has been demonstrated in the *GH* genes of multiple species, where variation in the number of repeats in an intronic repeat region has been associated with changes in growth traits (Almuly et al., 2008; Arango et al., 2014; Nie et al.,

2005; Sawayama & Takagi, 2015b). In particular, the intron 1 and 3 repeat regions identified in *C. auratus* in Chapter Two have been identified in other fish species as having a significant association with growth (Almuly et al., 2008; Li et al., 2017; Sawayama & Takagi, 2015b). For these reasons, the genetic variation within intron 1 and 3 repeat regions was assessed in *C. auratus.* To gain an overall representation of the variation at these loci, wild samples of *C. auratus* were used. Both intronic repeats were identified as polymorphic with variable repeat numbers. All individuals sampled were genotyped for intron 1 revealing six alleles at this locus. Intron 1 was in HWE in the wild populations. Investigation of the intron 3 repeat region was discontinued due to issues with amplification and genotyping at this locus. This locus was identified as a significant marker for growth in *P. major*, the closest relative of *C. auratus* (Sawayama & Takagi, 2015b). Long-read DNA sequencing of repetitive DNA, such as the method presented by De Bustos et al. (2016), would be a promising approach for precisely determining the level of repeat unit variation at this locus.

GH Intron 1 variation was then assessed in slow- and fast-growing groups of *C. auratus* to test for associations between this marker and growth rate. The slow-growing group showed deviation from HWE and had a significant heterozygote excess. No significant differences were detected in genotype or allele frequencies between slow- and fast-growing groups; however, results only marginally exceeded the 5% significance level. Although not statistically significant, the results from Chapter Three were consistent with the findings of other gene expression studies. For example, in this study, two alleles, 550 and 765, were common in all wild and cultured populations in this study. When compared to all other populations including the slow-growing group, the fast-growing *C. auratus* had the highest frequency of the 550 allele and the lowest frequency of the 765 allele. In comparison to the fast-growing group, the slow-growing *C. auratus* had a higher frequency of the 765 allele and a lower frequency of the 550 allele. Consequently, the slow-growing group had a high frequency of long alleles and the fast-growing group had a high frequency of short alleles. This phenomenon has also been observed in the first intron of *GH* in a close relative of *C. auratus*, *S. aurata.* Almuly et al. (2008) showed that longer repeat regions in the first intron were associated with repressed gene activity of *GH*. The reduced expression of gene variants with long introns has also been observed in the first intron of the other genes such as the Friedreich ataxia gene (Ohshima et al., 1998).

Though non-coding regions are known to be important for gene expression, the underlying mechanisms regulating expression are still largely unknown. One hypothesis predicts that due 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). It has also been suggested that repeat elements in non-coding regions 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). The characterisation of *caGH* in Chapter Two also identified putative transcription factor binding sites along this gene. Interestingly, consensus sequences for important transcription factors, TRE and GAGA factor, were identified within the core repeat unit of this intron 1 repeat region. 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). Similarly, GAGA-factors are known to regulate transcription at multiple levels (Lomaev et al., 2017) and are also likely to affect the expression of the *GH* gene. As these binding elements occur within the core repeat unit, the number of repeats in intron 1 (and therefore transcription factor binding sites) may alter the expression levels of this gene and help to explain how the length of this intron may affect growth. Though the results of this study are promising, a larger sample size would be required in order to produce more conclusive results. If significant associations were found between the intron 1 marker and growth rate of *C. auratus*, it could be used in marker-assisted selection for individuals with increased growth rates.

The assessment of the genetic variation of intron 1 in wild and cultured populations of *C. auratus* also revealed a high level of heterozygosity at this locus. Several possibilities for this finding were assessed, such as methodological limitations, negative assortative mating, outbreeding, sampling error, sample size, sex linkage, and selection. Selection against a detrimental and potentially lethal allele was considered the most likely cause of the high level of heterozygosity for several reasons. Most significantly, there was a complete absence of 485/485 homozygotes in wild and cultured populations in this study. The 485 allele was also found in relatively low frequencies in all populations. The lack of 485/485 homozygotes in this study may be due to a reduced chance of detecting a homozygote with an allele that occurs at low frequencies. Alternatively, the low frequency of this allele may suggest a low survival rate associated with this genotype. Lethal alleles can cause detrimental but non-lethal phenotypes in heterozygotes. For example, Achondroplasia, the most common form of short limb dwarfism in humans, is caused by a mutation in the *FGFR3* gene. In heterozygous form, the mutation

causes the detrimental growth phenotypes of the disease, whereas when in homozygous form, the mutation causes death very early in development (Horton et al., 2007). If the 485 allele from this study has detrimental effects in heterozygote form the allele may be in the process of being purged from these populations, which may also explain the low frequencies. In support of this theory, the frequency of the 485 allele is higher in the slow-growing group than the fast-growing group. The slow-growing group also had the largest excess of heterozygotes. This supports the hypothesis that the allele has detrimental effects (slow growth) when in heterozygous form. This allele was the shortest of all detected alleles in this study and may be the result of a deletion mutation. Large deletions in the first introns of other genes have been shown to cause detrimental consequences on gene function and health (Jansen et al., 2000; Rahkonen et al., 2004). There were no significant differences in the frequency of this allele between populations. More extensive sampling of wild and cultured *C. auratus* may provide more conclusive results. Sampling of individuals earlier in development may also assist in determining whether a lethal allele exists at this locus. If this is supported by additional data, it would provide important information for selective breeding programmes in *C. auratus* and other closely related species.

5.4. Gene regulation of *C. auratus* growth in response to temperature

In Chapter Four, transcriptomes of wild and domestic groups of *C. auratus* were compared under high and low temperature treatments (HTT and LTT respectively) to determine important cellular processes and genes involved in the regulation of growth in response to temperature changes. *C. auratus* kept at 21 °C grew faster than those kept at 13 °C. A large percentage of DEGs in HTT fish were associated with metabolic activity which is known to increase with temperature and body size (Brown et al., 2004). Consistent with the increase in metabolic activity, the HTT was characterised by upregulation of biosynthetic processes, genes encoding skeletal muscle components, and an increase in catalytic activity. Growth is experienced when anabolic processes exceed catabolic processes (Johnston et al., 2011; Schiaffino et al., 2013). In the HTT, the percentage of DEGs associated with biosynthetic processes largely exceeds the percentage associated with catabolic processes. This suggests that the upregulation of these anabolic processes is involved in the improved growth of HTT fish. In comparison, the LTT was characterised by an upregulation of processes often associated with responses to stimuli, specifically cold responses. In particular, the LTT fish experienced an upregulation of important genes associated with protein degradation pathways. In line with this, LTT fish had

a relatively equal balance between DEGs associated with catabolic and biosynthetic processes when compared to the HTT fish. Fish from the LTT experienced low growth rates and some lost body condition which is consistent with the observed percentages of DEGs associated with catabolic and anabolic processes in these fish. As stated, growth and atrophy are regulated by the opposing forces of anabolism and catabolism. The results of this study suggest that the differential growth rates of *C. auratus* under different temperature treatments are, in part, due to shifts in this balance.

Interestingly, fish in the HTT experienced an upregulation of several genes associated with the negative regulation of growth signalling pathways. This was accompanied by the downregulation of a number of important genes involved in the main IGF growth signalling pathways compared to the LTT. The GH-IGF axis is known to be subjected to multiple negative feedback loops (Backeljauw & Hwa, 2016), some of which appear to be occurring in these fish in response to chronic heat exposure. However, in order to determine whether growth is truly under negative feedback in the HTT and to further understand the mechanisms underlying the growth differences in these fish populations, the physiology of the whole organism must be considered. The GH-IGF axis is complex and controls growth through a number of pathways that affect target tissues in the body both directly and indirectly (Backeljauw & Hwa, 2016; Fuentes et al., 2013). This study assessed transcriptome data only from skeletal muscle in *C. auratus.* A more complete understanding of the mechanisms involved in this system would be gained by assessing gene expression in other tissues important to the GH-IGF axis such as the pituitary and the liver. Measuring the levels of GH and IGF in the bloodstream in response to different temperatures would also shed more light on the findings reported in this study. Furthermore, some of the results presented in this study may have been influenced by the underlying genetic variation within and between individuals in this study. This should be considered in future transcriptome studies and any effects of this variation may be limited by the use of genetic marker technologies for the selection of individuals for testing.

Growth rates can also be negatively affected by stress as energy and resources are reallocated away from investment activities such as growth (Wendelaar Bonga, 1997). The GO biological process, 'response to stress', was upregulated in the HTT and was accompanied by the upregulation of genes associated with responses to heat stress, DNA damage, and oxidative stress. These results imply that although fish kept at 21 °C experienced improved growth rates,

they also experienced a certain level of stress that may have limited their growth potential. It is likely that at a slightly lower temperature, the growth rates of *C. auratus* may improve further. These results provide useful information for determining the optimal growth temperature of *C. auratus.* To determine the optimal temperature for growth, a larger range of temperatures between 13 °C and 21 °C should be considered to establish at which temperature the highest growth rates are observed while simultaneously detecting little evidence of stress. The application of the optimal growth temperature to rearing programmes in *C. auratus* would likely improve growth rates and the overall production efficiency of this species.

The genetic regulation of growth rate in *C. auratus* was investigated using information on the expression of individual genes. However, genes do not work in isolation and an observed phenotype such as growth rate is likely to be the result of multiple gene-gene interactions in complex pathways or networks (Dixon, Costanzo, Baryshnikova, Andrews, & Boone, 2009). Moreover, transcription of a particular gene does not necessarily lead to the expression or action of the coded protein. Therefore, future research on the genetic regulation of growth traits in *C. auratus* should involve genetic pathway and network analyses. This will provide a more holistic understanding of the gene-gene interactions that affect growth traits.

5.5. Implications and future directions

Until now, little was known about the regulation of growth in *C. auratus*. Previous evidence suggested that growth of *C. auratus* varied with geographical distribution (Paul, 1976; Paul & Tarring, 1980) and appeared to be affected by environmental factors such as temperature (Francis, 1994; Murphy et al., 2013), photoperiod (Fielder, Bardsley, Allan, & Pankhurst, 2002), and nutrient availability (Murphy et al., 2013). However, a lack of specific growth studies in this species has prevented a deeper understanding of the mechanisms regulating growth in this species. This thesis presents the first study to characterise the gene of a member of the GH-IGF axis in *C. auratus* and test its association with growth rate. Here, a repeat region in the first intron of the *caGH* gene was identified as a potential future marker for growth studies. Several other repeat regions and SNPs were identified along this gene providing guidance for further studies investigating the effects of *GH* regulation on growth in this species. This study also showed that the *GH* gene structure is highly conserved within the fish order Perciformes hence the findings of studies on *GH* such as in this thesis may have implications for a large number of fish species. This study exemplifies the benefits of applying comparative

genetic techniques to gain genetic information on a non-model species from a well-studied one. *C. auratus* is fortunate as its sister species, *P. major*, is a well-established aquaculture species and has one of the longest histories of domestication in marine finfish (Murata et al., 1996). These two species have such a high level of DNA sequence similarity that there has been historical debate as to whether or not *P. major* and *C. auratus* should be classed as one species (Parsons et al., 2014; Tabata & Taniguchi, 2000). Therefore, there is large potential for using the genetic findings reported for *P. major* to inform the experiment designs and genomic targets to investigate in *C. auratus*.

The results presented in Chapter Four of this thesis and within the wider study carried out by PFR (M. Wellenreuther, unpublished data) represent the first study to investigate the specific growth response of *C. auratus* to different temperatures. Previous knowledge of the effects of temperature on growth in this species was largely based on wild populations. In the wild, changes in temperature are associated with corresponding changes in photoperiod and nutrient availability which are also known to affect growth in fish (Björnsson et al., 2002; Gabillard et al., 2005). Therefore, this study offers a more direct investigation of the effects of temperature on *C. auratus*. The findings presented in this thesis highlight several important cellular pathways and genes involved in the regulation of growth in response to temperature and highlight the potential effects of heat stress at high temperatures in this species. This study also provides a useful body of research from which aspects of growth control can be targeted for further investigation in this species. While these results provide useful information for rearing fast-growing *C. auratus* for aquaculture, they also have implications for climate change research in this species. This issue however, was not within the scope of this thesis and is an area for future research.

The research in this thesis has focused primarily on how *C. auratus* growth is influenced by the underlying genetic variation of the *GH* gene and the gene expression patterns of many other genes in response to temperature. However, it does not account for the possible effects of epigenetic and RNA interference mechanisms on growth traits. There is growing evidence to suggest an important role of epigenetics (Álvarez-Nava & Lanes, 2017; Zhong et al., 2014) and RNA interference mechanisms such as post-transcriptional gene repression by microRNAs (Huang et al., 2012; Lui, 2017) in regulating growth, and in particular, influencing the expression of genes in the GH-IGF axis (Álvarez-Nava & Lanes, 2017; Huang et al., 2012). This represents a somewhat complex area for future research into the regulation of growth in

C. auratus. Epigenetic mechanisms are often referred to as the mediators between the genome and the environment, so investigating mechanisms such as DNA methylation may shed light on the mechanisms that control how growth rate responds to environmental factors such as temperature.

This thesis research demonstrates the benefits of using genetic techniques in aquaculture programmes to improve our understanding of complex traits such as growth and, in doing so, provide valuable information for future breeding programmes. Despite its obvious importance in breeding programmes, growth remains poorly understood in fish. Its regulation is exceedingly complex and is influenced by a multitude of genetic and environmental factors. This thesis focused on a small subset of these factors. In order to unravel the complex mechanisms involved in growth regulation all of these genes and environmental factors must be considered. Significant advances in genetic technologies such as next generation sequencing, offer great potential for assessing complex, quantitative traits such as growth. Markers associated with growth can now be identified on a genome-wide scale by using methods such as GWAS and QTL and can be applied for the use of marker-assisted selection techniques (Yue, 2014). These methods allow the identification of large numbers of loci associated with a trait of interest. Genome-wide studies such as these have already been carried out successfully in economically important aquaculture species such as Atlantic salmon (Gutierrez et al., 2015), rainbow trout (Wringe et al., 2010), and big-head carp (Fu et al., 2016). With increasing knowledge of the *C. auratus* genome and the capacity of comparative genetic studies, there is considerable potential for future genome-wide studies in *C. auratus.* These types of studies will be invaluable in the future for understanding and selectively improving growth rates in *C. auratus* and other important commercial species.

5.6. Final conclusion

The overall aim of this thesis was to use genome sequence data and transcriptomics to investigate the loci that influence growth rate of *C. auratus.* The *GH* gene was characterised in *C. auratus* and a polymorphic repeat region was identified in the first intron with potential effects on growth in this species. With further investigation, this locus may turn out to be a significant marker of growth rate in *C. auratus*. Transcriptome data from *C. auratus* under high and low temperature treatments were also assessed to investigate the effects of temperature on the gene regulation of growth. This led to the identification of several important cellular

pathways and genes putatively involved in the regulation of growth in *C. auratus* and highlighted the potential effects of heat stress in this species. The research of this thesis represents one of the first studies to explore the genetic regulation of growth in *C. auratus* and significantly contributes to the current understanding of growth not only in this species, but in many other fish species. The findings of studies such as this may be applied to selective breeding programmes in *C. auratus* to increase the growth rate, and consequently, the production efficiency of this species improving its economic viability as a commercial aquaculture species. If applied to selective breeding programmes, studies such as those described in this thesis have the potential to greatly enhance the production and the efficiency of the aquaculture sector, relieve fishing pressure of wild-caught fisheries, and support the nutritional protein requirements of a growing human population.

References

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.

Álvarez-Nava, F., & Lanes, R. (2017). GH/IGF-1 signaling and current knowledge of epigenetics; a review and considerations on possible therapeutic options. *International Journal of Molecular Sciences, 18*(10), 1624.

Alves, R. N., Cordeiro, O., Silva, T. S., Richard, N., de Vareilles, M., Marino, G., et al. (2010). Metabolic molecular indicators of chronic stress in gilthead seabream (Sparus aurata) using comparative proteomics. *Aquaculture, 299*(1-4), 57-66.

Amores, A., Force, A., Yan, Y.-L., Joly, L., Amemiya, C., Fritz, A., et al. (1998). Zebrafish hox clusters and vertebrate genome evolution. *Science, 282*(5394), 1711-1714.

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.

Bazykin, G. A., & Kochetov, A. V. (2010). Alternative translation start sites are conserved in eukaryotic genomes. *Nucleic acids research, 39*(2), 567-577.

Bentsen, H., Eknath, A., Rye, M., Thodesen, J., & Gjerde, B. (2005). *Genetic improvement of farmed tilapias: response to selection for growth performance in the GIFT project.* Paper presented at the Aquaculture.

Ber, R., & Daniel, V. (1993). Sequence analysis suggests a recent duplication of the growth hormone encoding gene in Tilapia nilotica. *Gene, 125*(2), 143-150.

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.

Björnsson, B. T., Johansson, V., Benedet, S., Einarsdottir, I. E., Hildahl, J., Agustsson, T., et al. (2002). Growth hormone endocrinology of salmonids: regulatory mechanisms and mode of action. *Fish Physiology and Biochemistry, 27*(3-4), 227-242.

Bodine, S. C., Latres, E., Baumhueter, S., Lai, V. K.-M., Nunez, L., Clarke, B. A., et al. (2001). Identification of ubiquitin ligases required for skeletal muscle atrophy. *Science, 294*(5547), 1704-1708.

Bodine, S. C., Stitt, T. N., Gonzalez, M., Kline, W. O., Stover, G. L., Bauerlein, R., et al. (2001). Akt/mTOR pathway is a crucial regulator of skeletal muscle hypertrophy and can prevent muscle atrophy in vivo. *Nature Cell Biology, 3*(11), 1014.

Bostock, J., McAndrew, B., Richards, R., Jauncey, K., Telfer, T., Lorenzen, K., et al. (2010). Aquaculture: global status and trends. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1554), 2897-2912.

Bower, N. I., & Johnston, I. A. (2010). Paralogs of Atlantic salmon myoblast determination factor genes are distinctly regulated in proliferating and differentiating myogenic cells. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 298*(6), R1615-R1626.

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., & West, G. B. (2004). Toward a metabolic theory of ecology. *Ecology, 85*(7), 1771-1789.

Budanov, A. V. (2011). Stress-responsive sestrins link p53 with redox regulation and mammalian target of rapamycin signaling. *Antioxidants & Redox Signaling, 15*(6), 1679-1690.

Butler, A. A., & Roith, D. L. (2001). Control of growth by the somatropic axis: growth hormone and the insulin-like growth factors have related and independent roles. *Annual Review of Physiology, 63*(1), 141-164.

Calnan, D., & Brunet, A. (2008). The foxo code. *Oncogene, 27*(16), 2276.

Cao, H., Zhou, R., Wei, Q., Li, C., & Gui, J. (2011). Molecular characterization of the growth hormone in Chinese sturgeon and its expression during embryogenesis and early larval stages. *Journal of Applied Ichthyology, 27*(2), 501-504.

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, B., Wolfgang, C. D., & Hai, T. (1996). Analysis of ATF3, a transcription factor induced by physiological stresses and modulated by gadd153/Chop10. *Molecular and Cellular Biology, 16*(3), 1157-1168.

Chen, D., Liu, J., Chen, W., Shi, S., Zhang, W., & Zhang, L. (2015). Expression and ontogeny of growth hormone (Gh) in the protogynous hermaphroditic ricefield eel (Monopterus albus). *Fish Physiology and Biochemistry, 41*(6), 1515-1525.

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.

Consitt, L. A., Saneda, A., Saxena, G., List, E. O., & Kopchick, J. J. (2017). Mice overexpressing growth hormone exhibit increased skeletal muscle myostatin and MuRF1 with attenuation of muscle mass. *Skeletal Muscle, 7*(1), 17.

Coolican, S. A., Samuel, D. S., Ewton, D. Z., McWade, F. J., & Florini, J. R. (1997). The mitogenic and myogenic actions of insulin-like growth factors utilize distinct signaling pathways. *Journal of Biological Chemistry, 272*(10), 6653-6662.

Danzmann, R. G., Kocmarek, A. L., Norman, J. D., Rexroad, C. E., & Palti, Y. (2016). Transcriptome profiling in fast versus slow-growing rainbow trout across seasonal gradients. *BMC Genomics, 17*(1), 60.

De Bustos, A., Cuadrado, A., & Jouve, N. (2016). Sequencing of long stretches of repetitive DNA. *Scientific Reports, 6*, 36665.

De-Santis, C., & Jerry, D. R. (2007). Candidate growth genes in finfish—Where should we be looking? *Aquaculture, 272*(1), 22-38.

Dekkers, J. (2012). Application of genomics tools to animal breeding. *Current Genomics, 13*(3), 207-212.

Devlin, R. H., & Nagahama, Y. (2002). Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture, 208*(3), 191-364.

Diouf, J. (2009). Statement of Jacques Diouf, Director-General of the Food and Agriculture Organization of the United Nations (FAO), Opening of the High-Level Expert Forum on 'How to feed the world in 2050'. *Rome: FAO*.

Dixon, S. J., Costanzo, M., Baryshnikova, A., Andrews, B., & Boone, C. (2009). Systematic mapping of genetic interaction networks. *Annual review of genetics, 43*, 601-625.

Du, S. J., Li, H., Bian, Y., & Zhong, Y. (2008). Heat-shock protein 90α1 is required for organized myofibril assembly in skeletal muscles of zebrafish embryos. *Proceedings of the National Academy of Sciences, 105*(2), 554-559.

Edens, A., & Talamantes, F. (1998). Alternative processing of growth hormone receptor transcripts. *Endocrine Reviews, 19*(5), 559-582.

Evdokimova, V., Tognon, C. E., Benatar, T., Yang, W., Krutikov, K., Pollak, M., et al. (2012). IGFBP7 binds to the IGF-1 receptor and blocks its activation by insulin-like growth factors. *Science Signaling, 5*(255), ra92-ra92.

FAO. (2016). The State of World Fisheries and Aquaculture 2016. Contributing to Food Security and Nutrition for All: FAO Rome.

Feng, J., Liu, S., Wang, X., Wang, R., Zhang, J., Jiang, Y., et al. (2014). Channel catfish hemoglobin genes: identification, phylogenetic and syntenic analysis, and specific induction in response to heat stress. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 9*, 11-22.

Feng, X., Yu, X., & Tong, J. (2014). Novel single nucleotide polymorphisms of the insulin-like growth factor-I gene and their associations with growth traits in common carp (Cyprinus carpio L.). *International Journal of Molecular Sciences, 15*(12), 22471-22482.

Fielder, D. S., Bardsley, W. J., Allan, G. L., & Pankhurst, P. M. (2002). Effect of photoperiod on growth and survival of snapper Pagrus auratus larvae. *Aquaculture, 211*(1-4), 135-150.

Forbes, S. H., Knudsen, K. L., North, T. W., & Allendorf, F. W. (1994). One of two growth hormone genes in coho salmon is sex-linked. *Proceedings of the National Academy of Sciences, 91*(5), 1628-1631.

Francis, M. P. (1994). Growth of juvenile snapper, Pagrus auratus. *New Zealand Journal of Marine and Freshwater Research, 28*(2), 201-218.

Fu, B., Liu, H., Yu, X., & Tong, J. (2016). A high-density genetic map and growth related QTL mapping in bighead carp (Hypophthalmichthys nobilis). *Scientific Reports, 6*, 28679.

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.

Gabillard, J.-C., Weil, C., Rescan, P.-Y., Navarro, I., Gutierrez, J., & Le Bail, P.-Y. (2005). Does the GH/IGF system mediate the effect of water temperature on fish growth? A review. *Cybium, 29*(2), 107-117.

Gabillard, J.-C., Weil, C., Rescan, P.-Y., Navarro, I., Gutiérrez, J., & Le Bail, P.-Y. (2003a). Effects of environmental temperature on IGF1, IGF2, and IGF type I receptor expression in rainbow trout (Oncorhynchus mykiss). *General and Comparative Endocrinology, 133*(2), 233-242.

Gabillard, J.-C., Weil, C., Rescan, P.-Y., Navarro, I., Gutiérrez, J., & Le Bail, P.-Y. (2003b). Environmental temperature increases plasma GH levels independently of nutritional status in rainbow trout (Oncorhynchus mykiss). *General and Comparative Endocrinology, 133*(1), 17-26.

Garikipati, D. K., & Rodgers, B. D. (2012). Myostatin inhibits myosatellite cell proliferation and consequently activates differentiation: evidence for endocrine-regulated transcript processing. *Journal of Endocrinology, 215*(1), 177-187.

Gjedrem, T., & Baranski, M. (2010). *Selective breeding in aquaculture: an introduction* (Vol. 10): Springer Science & Business Media.

Gjedrem, T., & Robinson, N. (2014). Advances by selective breeding for aquatic species: a review. *Agricultural Sciences, 5*(12), 1152.

Gjedrem, T., Robinson, N., & Rye, M. (2012). The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. *Aquaculture, 350*, 117-129.

Glasauer, S. M., & Neuhauss, S. C. (2014). Whole-genome duplication in teleost fishes and its evolutionary consequences. *Molecular Genetics and Genomics, 289*(6), 1045-1060.

Glass, D. J. (2003). Molecular mechanisms modulating muscle mass. *Trends in Molecular Medicine, 9*(8), 344-350.

Glickman, M. H., & Ciechanover, A. (2002). The ubiquitin-proteasome proteolytic pathway: destruction for the sake of construction. *Physiological Reviews, 82*(2), 373-428.

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/softwares/fstat. html*.

Gracey, A. Y., Fraser, E. J., Li, W., Fang, Y., Taylor, R. R., Rogers, J., et al. (2004). Coping with cold: an integrative, multitissue analysis of the transcriptome of a poikilothermic vertebrate. *Proceedings of the National Academy of Sciences of the United States of America, 101*(48), 16970-16975.

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., et al. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology, 31*(4), 337-350.

Gross, R., & Nilsson, J. (1999). Restriction fragment length polymorphism at the growth hormone 1 gene in Atlantic salmon (Salmo salar L.) and its association with weight among the offspring of a hatchery stock. *Aquaculture, 173*(1), 73-80.

Gutierrez, A. P., Yáñez, J. M., Fukui, S., Swift, B., & Davidson, W. S. (2015). Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (Salmo salar). *PLoS One, 10*(3), e0119730.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols, 8*(8), 1494.

Hartl, F. U. (1996). Molecular chaperones in cellular protein folding. *Nature, 381*(6583), 571.

Harvey, S., Martínez-Moreno, C. G., Luna, M., & Arámburo, C. (2015). Autocrine/paracrine roles of extrapituitary growth hormone and prolactin in health and disease: An overview. *General and Comparative Endocrinology, 220*, 103-111.

Hawkins, T. A., Haramis, A.-P., Etard, C., Prodromou, C., Vaughan, C. K., Ashworth, R., et al. (2008). The ATPase-dependent chaperoning activity of Hsp90a regulates thick filament formation and integration during skeletal muscle myofibrillogenesis. *Development, 135*(6), 1147-1156.

Hetz, C. (2012). The unfolded protein response: controlling cell fate decisions under ER stress and beyond. *Nature Reviews Molecular Cell Biology, 13*(2), 89.

Hochachka, P. W., & Somero, G. N. (2002). *Biochemical Adaptation: Mechanism and Process in Physiological Evolution*: Oxford University Press.

Horton, W. A., Hall, J. G., & Hecht, J. T. (2007). Achondroplasia. *The Lancet, 370*(9582), 162-172.

Hu, Z.-L., Bao, J., & Reecy, J. M. (2008). CateGOrizer: a web-based program to batch analyze gene on-tology classification categories. *Online Journal of Bioinformatics, 9*, 108-112.

Huang, C., Li, Y., Hu, S., Chi, J., Lin, G., Lin, C., et al. (2012). Differential expression patterns of growth-related microRNAs in the skeletal muscle of Nile tilapia. *Journal of Animal Science, 90*(12), 4266-4279.

Hulata, G. (2001). Genetic manipulations in aquaculture: a review of stock improvement by classical and modern technologies. *Genetica, 111*(1-3), 155-173.

Ishikawa, T., Taniguchi, Y., Okada, T., Takeda, S., & Mori, K. (2011). Vertebrate unfolded protein response: mammalian signaling pathways are conserved in medaka fish. *Cell Structure and Function, 36*(2), 247-259.

Iwama, G. K., Thomas, P. T., Forsyth, R. B., & Vijayan, M. M. (1998). Heat shock protein expression in fish. *Reviews in Fish Biology and Fisheries, 8*(1), 35-56.

Jaillon, O., Aury, J.-M., Brunet, F., Petit, J.-L., Stange-Thomann, N., Mauceli, E., et al. (2004). Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. *Nature, 431*(7011), 946.

Jansen, M. S., Cook, G. A., Song, S., & Park, E. A. (2000). Thyroid hormone regulates carnitine palmitoyltransferase Iα gene expression through elements in the promoter and first intron. *Journal of Biological Chemistry, 275*(45), 34989-34997.

Jansson, M., Andersson, G., Uhlén, M., Nilsson, B., & Kördel, J. (1998). The insulin-like growth factor (IGF) binding protein 1 binding epitope on IGF-I probed by heteronuclear NMR spectroscopy and mutational analysis. *Journal of Biological Chemistry, 273*(38), 24701-24707.

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.

Jiménez-Amilburu, V., Salmerón, C., Codina, M., Navarro, I., Capilla, E., & Gutiérrez, J. (2013). Insulin-like growth factors effects on the expression of myogenic regulatory factors in gilthead sea bream muscle cells. *General and Comparative Endocrinology, 188*, 151-158.

Jing, J., Xiong, S., Li, Z., Wu, J., Zhou, L., Gui, J.-F., et al. (2015). A feedback regulatory loop involving p53/miR-200 and growth hormone endocrine axis controls embryo size of zebrafish. *Scientific Reports, 5*.

Johnston, I. A. (2006). Environment and plasticity of myogenesis in teleost fish. *Journal of Experimental Biology, 209*(12), 2249-2264.

Johnston, I. A., Bower, N. I., & Macqueen, D. J. (2011). Growth and the regulation of myotomal muscle mass in teleost fish. *Journal of Experimental Biology, 214*(10), 1617-1628.

Jones, D. D. (2005). Triplet nucleotide removal at random positions in a target gene: the tolerance of TEM-1 β-lactamase to an amino acid deletion. *Nucleic Acids Research, 33*(9), e80-e80.

Juhua, Y., Xuefeng, C., Jianlin, L., Yongkai, T., Hongxia, L., Pao, X., et al. (2010). Isolation of IGF2 and association of IGF2 polymorphism with growth trait in genetically improved farmed tilapias, Oreochromis niloticus L. *Aquaculture Research, 41*(11).

Kailola, P. J. (1993). *Australian fisheries resources*: Bureau of Resource Sciences, Dept. of Primary Industries and Energy; Fisheries Research and Development Corp.

Kajimura, S., & Duan, C. (2007). Insulin-like growth factor-binding protein-1: an evolutionarily conserved fine tuner of insulin-like growth factor action under catabolic and stressful conditions. *Journal of Fish Biology, 71*(sc), 309-325.

Kamei, Y., Miura, S., Suzuki, M., Kai, Y., Mizukami, J., Taniguchi, T., et al. (2004). Skeletal muscle FOXO1 (FKHR) transgenic mice have less skeletal muscle mass, down-regulated Type I (slow twitch/red muscle) fiber genes, and impaired glycemic control. *Journal of Biological Chemistry, 279*(39), 41114-41123.

Karimian, A., Ahmadi, Y., & Yousefi, B. (2016). Multiple functions of p21 in cell cycle, apoptosis and transcriptional regulation after DNA damage. *DNA Repair, 42*, 63-71.

Kassahn, K. S., Crozier, R. H., Ward, A. C., Stone, G., & Caley, M. J. (2007). From transcriptome to biological function: environmental stress in an ectothermic vertebrate, the coral reef fish Pomacentrus moluccensis. *BMC Genomics, 8*(1), 358.

Kawase, T., Ohki, R., Shibata, T., Tsutsumi, S., Kamimura, N., Inazawa, J., et al. (2009). PH domain-only protein PHLDA3 is a p53-regulated repressor of Akt. *Cell, 136*(3), 535-550.

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.

Lappin, F. M., Shaw, R. L., & Macqueen, D. J. (2016). Targeted sequencing for high-resolution evolutionary analyses following genome duplication in salmonid fish: Proof of concept for key components of the insulin-like growth factor axis. *Marine Genomics, 30*, 15-26.

Lee, S.-J., & McPherron, A. C. (2001). Regulation of myostatin activity and muscle growth. *Proceedings of the National Academy of Sciences, 98*(16), 9306-9311.

Leggatt, R., Raven, P., Mommsen, T., Sakhrani, D., Higgs, D., & Devlin, R. (2009). Growth hormone transgenesis influences carbohydrate, lipid and protein metabolism capacity for energy production in coho salmon (Oncorhynchus kisutch). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 154*(1), 121-133.

Li, J., & Johnson, S. E. (2006). ERK2 is required for efficient terminal differentiation of skeletal myoblasts. *Biochemical and Biophysical Research Communications, 345*(4), 1425-1433.

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, X., Bai, J., Hu, Y., Ye, X., Li, S., & Yu, L. (2012). Genotypes, haplotypes and diplotypes of IGF-II SNPs and their association with growth traits in largemouth bass (Micropterus salmoides). *Molecular Biology Reports, 39*(4), 4359-4365.

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.

Liang, L., Chang, Y., He, X., & Tang, R. (2015). Transcriptome analysis to identify cold-responsive genes in amur carp (Cyprinus carpio haematopterus). *PLoS One, 10*(6), e0130526.

Liu, S., Wang, X., Sun, F., Zhang, J., Feng, J., Liu, H., et al. (2013). RNA-Seq reveals expression signatures of genes involved in oxygen transport, protein synthesis, folding, and degradation in response to heat stress in catfish. *Physiological genomics, 45*(12), 462-476.

Liu, Z. J., & Cordes, J. (2004). DNA marker technologies and their applications in aquaculture genetics. *Aquaculture, 238*(1-4), 1-37.

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.

Long, Y., Song, G., Yan, J., He, X., Li, Q., & Cui, Z. (2013). Transcriptomic characterization of cold acclimation in larval zebrafish. *BMC Genomics, 14*(1), 612.

Lui, J. C. (2017). Regulation of body growth by microRNAs. *Molecular and Cellular Endocrinology, 456*, 2-8.

Mackay, T. F. (2001). The genetic architecture of quantitative traits. *Annual Review of Genetics, 35*(1), 303-339.

McAndrew, B., & Napier, J. (2011). Application of genetics and genomics to aquaculture development: current and future directions. *The Journal of Agricultural Science, 149*(S1), 143-151.

McCormick, S. D., Moriyama, S., & Björnsson, B. T. (2000). Low temperature limits photoperiod control of smolting in Atlantic salmon through endocrine mechanisms. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 278*(5), R1352-R1361.

McKay, S., Trautner, J., Smith, M., Koop, B., & Devlin, R. (2004). Evolution of duplicated growth hormone genes in autotetraploid salmonid fishes. *Genome, 47*(4), 714-723.

Mercurio, F., & Manning, A. M. (1999). NF-κB as a primary regulator of the stress response. *Oncogene, 18*(45), 6163.

Mizushima, N., & Komatsu, M. (2011). Autophagy: renovation of cells and tissues. *Cell, 147*(4), 728-741.

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.

Mullis, P. E. (2005). Genetic control of growth. *European Journal of Endocrinology, 152*(1), 11-31.

Mungrue, I. N., Pagnon, J., Kohannim, O., Gargalovic, P. S., & Lusis, A. J. (2009). CHAC1/MGC4504 is a novel proapoptotic component of the unfolded protein response, downstream of the ATF4-ATF3-CHOP cascade. *The Journal of Immunology, 182*(1), 466-476.

Murata, O., Harada, T., Miyashita, S., Izumi, K.-i., Maeda, S., Kato, K., et al. (1996). Selective breeding for growth in red sea bream. *Fisheries Science, 62*(6), 845-849.

Murphy, H. M., Jenkins, G. P., Hamer, P. A., & Swearer, S. E. (2013). Interannual variation in larval abundance and growth in snapper Chrysophrys auratus (Sparidae) is related to prey availability and temperature. *Marine Ecology Progress Series, 487*, 151-162.

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 Replicationin Vivo. *Journal of Biological Chemistry, 273*(23), 14588-14595.

Overturf, K., LaPatra, S., Towner, R., Campbell, N., & Narum, S. (2010). Relationships between growth and disease resistance in rainbow trout, Oncorhynchus mykiss (Walbaum). *Journal of Fish Diseases, 33*(4), 321-329.

Pál, C., Papp, B., & Lercher, M. J. (2006). An integrated view of protein evolution. *Nature Reviews Genetics, 7*(5), 337-348.

Pan, X., Zhao, J., Zhang, W.-N., Li, H.-Y., Mu, R., Zhou, T., et al. (2009). Induction of SOX4 by DNA damage is critical for p53 stabilization and function. *Proceedings of the National Academy of Sciences, 106*(10), 3788-3793.

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.

Parsons, P., & Bodmer, W. (1961). The evolution of overdominance: Natural selection and heterozygote advantage. *Nature, 190*(4770), 7.

Paul, L. (1976). *A study on age, growth and population structure of the snapper, Chrysophrys auratus (Forster), in the Hauraki Gulf, New Zealand*: Fisheries Research Division, New Zealand Ministry of Agriculture and Fisheries.

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.

Paul, L., & Tarring, S. (1980). Growth rate and population structure of snapper, Chrysophrys auratus, in the East Cape region, New Zealand. *New Zealand Journal of Marine and Freshwater Research, 14*(3), 237-247.

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.

Pinheiro, C. H., & Guimarães-Ferreira, L. (2015). Frontiers in skeletal muscle wasting, regeneration and stem cells. *Frontiers in Physiology, 6*, 141.

Plateroti, M., Kress, E., Mori, J. I., & Samarut, J. (2006). Thyroid hormone receptor α1 directly controls transcription of the β-catenin gene in intestinal epithelial cells. *Molecular and Cellular Biology, 26*(8), 3204-3214.

Podrabsky, J. E., & Somero, G. N. (2004). Changes in gene expression associated with acclimation to constant temperatures and fluctuating daily temperatures in an annual killifish Austrofundulus limnaeus. *Journal of Experimental Biology, 207*(13), 2237-2254.

Pörtner, H. (2001). Climate change and temperature-dependent biogeography: oxygen limitation of thermal tolerance in animals. *Naturwissenschaften, 88*(4), 137-146.

Quéré, 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.

Quinn, N. L., McGowan, C. R., Cooper, G. A., Koop, B. F., & Davidson, W. S. (2011). Identification of genes associated with heat tolerance in Arctic charr exposed to acute thermal stress. *Physiological Genomics, 43*(11), 685-696.

Rahkonen, O., Su, M., Hakovirta, H., Koskivirta, I., Hormuzdi, S. G., Vuorio, E., et al. (2004). Mice with a deletion in the first intron of the Col1a1 gene develop age-dependent aortic dissection and rupture. *Circulation Research, 94*(1), 83-90.

Raymond, M. (1995). GENEPOP: Population genetics software for exact tests and ecumenism. Vers. 1.2. *Journal of Heredity, 86*, 248-249.

Reed, K. M., Mendoza, K. M., Abrahante, J. E., Barnes, N. E., Velleman, S. G., & Strasburg, G. M. (2017). Response of turkey muscle satellite cells to thermal challenge. I. transcriptome effects in proliferating cells. *BMC Genomics, 18*(1), 352.

Reinecke, M. (2010). Influences of the environment on the endocrine and paracrine fish growth hormone–insulin-like growth factor-I system. *Journal of Fish Biology, 76*(6), 1233-1254.

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.

Ricordel, M.-J., Smal, J., & Le Bail, P.-Y. (1995). Application of a recombinant cichlid growth hormone radioimmunoassay to measure native GH in tilapia (Oreochromis niloticus) bred at different temperatures. *Aquatic Living Resources, 8*(2), 153-160.

Robertson, A. (1965). The interpretation of genotypic ratios in domestic animal populations. *Animal Science, 7*(3), 319-324.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics, 26*(1), 139-140.

Rommel, C., Bodine, S. C., Clarke, B. A., Rossman, R., Nunez, L., Stitt, T. N., et al. (2001). Mediation of IGF-1-induced skeletal myotube hypertrophy by PI (3) K/Akt/mTOR and PI (3) K/Akt/GSK3 pathways. *Nature Cell Biology, 3*(11), 1009.

Rowe, R., & Goldspink, G. (1969). Muscle fibre growth in five different muscles in both sexes of mice. *Journal of Anatomy, 104*(Pt 3), 519.

Ryan, J., & Hightower, L. (1996). Stress proteins as molecular biomarkers for environmental toxicology *Stress-Inducible Cellular Responses* (pp. 411-424): Springer.

Rye, M., LilleviK, K. M., & Gjerde, B. (1990). Survival in early life of Atlantic salmon and rainbow trout: estimates of heritabilities and genetic correlations. *Aquaculture, 89*(3-4), 209-216.

Şahin-Çevik, M. (2013). Identification and expression analysis of early cold-induced genes from cold-hardy Citrus relative Poncirus trifoliata (L.) Raf. *Gene, 512*(2), 536-545.

Sandri, M., Sandri, C., Gilbert, A., Skurk, C., Calabria, E., Picard, A., et al. (2004). Foxo transcription factors induce the atrophy-related ubiquitin ligase atrogin-1 and cause skeletal muscle atrophy. *Cell, 117*(3), 399-412.

Sarropoulou, E., & Fernandes, J. M. (2011). Comparative genomics in teleost species: Knowledge transfer by linking the genomes of model and non-model fish species. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 6*(1), 92-102.

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.

Schiaffino, S., Dyar, K. A., Ciciliot, S., Blaauw, B., & Sandri, M. (2013). Mechanisms regulating skeletal muscle growth and atrophy. *The FEBS journal, 280*(17), 4294-4314.

Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology, 18*(2), 233-234.

Scott, G. R., & Johnston, I. A. (2012). Temperature during embryonic development has persistent effects on thermal acclimation capacity in zebrafish. *Proceedings of the National Academy of Sciences, 109*(35), 14247-14252.

Shimizu, N., Yoshikawa, N., Ito, N., Maruyama, T., Suzuki, Y., Takeda, S.-i., et al. (2011). Crosstalk between glucocorticoid receptor and nutritional sensor mTOR in skeletal muscle. *Cell Metabolism, 13*(2), 170-182.

Sjogren, K., Leung, K.-C., Kaplan, W., Gardiner-Garden, M., Gibney, J., & Ho, K. K. (2007). Growth hormone regulation of metabolic gene expression in muscle: a microarray study in hypopituitary men. *American Journal of Physiology-Endocrinology and Metabolism, 293*(1), E364-E371.

Smith, R. L. (1992). Elements of ecology. 3rd Edit: Harper Collins Publishers Inc.

Smith, S., Bernatchez, L., & Beheregaray, L. B. (2013). RNA-seq analysis reveals extensive transcriptional plasticity to temperature stress in a freshwater fish species. *BMC Genomics, 14*(1), 375.

Stevens, L., Salomon, B., & Sun, G. (2007). Microsatellite variability and heterozygote excess in Elymus trachycaulus populations from British Columbia in Canada. *Biochemical Systematics and Ecology, 35*(11), 725-736.

Tabas, I., & Ron, D. (2011). Integrating the mechanisms of apoptosis induced by endoplasmic reticulum stress. *Nature Cell Biology, 13*(3), 184.

Tabata, K., & Taniguchi, N. (2000). Differences between Pagrus major and Pagrus auratus through mainly mtDNA control region analysis. *Fisheries Science, 66*(1), 9-18.

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.

Tan, E., Wongwarangkana, C., Kinoshita, S., Suzuki, Y., Oshima, K., Hattori, M., et al. (2012). Global gene expression analysis of gill tissues from normal and thermally selected strains of rainbow trout. *Fisheries Science, 78*(5), 1041-1049.

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.

Thodesen, J., Grisdale-Helland, B., Helland, S. J., & Gjerde, B. (1999). Feed intake, growth and feed utilization of offspring from wild and selected Atlantic salmon (Salmo salar). *Aquaculture, 180*(3-4), 237-246.

Tomalty, K. M., Meek, M. H., Stephens, M. R., Rincón, G., Fangue, N. A., May, B. P., et al. (2015). Transcriptional response to acute thermal exposure in juvenile chinook salmon determined by RNAseq. *G3: Genes, Genomes, Genetics*, g3. 115.017699.

Valdés, J. A., Flores, S., Fuentes, E. N., Osorio-Fuentealba, C., Jaimovich, E., & Molina, A. (2013). IGF-1 induces IP3-dependent calcium signal involved in the regulation of myostatin gene expression mediated by NFAT during myoblast differentiation. *Journal of Cellular Physiology, 228*(7), 1452-1463.

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.

Wang, H., Lin, G., & Zhang, Z. (2007). ATF5 promotes cell survival through transcriptional activation of Hsp27 in H9c2 cells. *Cell Biology International, 31*(11), 1309-1315.

Wang, Y., Wang, X., Meng, X., Wang, H., Jiang, Z., & Qiu, X. (2014). Identification of two SNPs in myostatin (MSTN) gene of Takifugu rubripes and their association with growth traits. *Molecular and Cellular Probes, 28*(4), 200-203.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics, 10*(1), 57.

Waples, R. S. (2014). Testing for Hardy–Weinberg proportions: have we lost the plot? *Journal of Heredity, 106*(1), 1-19.

Watabe, S. (1999). Myogenic regulatory factors and muscle differentiation during ontogeny in fish. *Journal of Fish Biology, 55*(sA), 1-18.

Wei, T., Sun, Y., Zhang, B., Wang, R., & Xu, T. (2014). A mitogenomic perspective on the phylogenetic position of the Hapalogenys 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.

Wellenreuther, M., & Hansson, B. (2016). Detecting polygenic evolution: problems, pitfalls, and promises. *Trends in Genetics, 32*(3), 155-164.

Wendelaar Bonga, S. E. (1997). The stress response in fish. *Physiological Reviews, 77*(3), 591-625.

Williams, J. (2005). The use of marker-assisted selection in animal breeding and biotechnology. *Revue Scientifique et Technique-Office International des Epizooties, 24*(1), 379.

Windisch, H. S., Frickenhaus, S., John, U., Knust, R., Pörtner, H. O., & Lucassen, M. (2014). Stress response or beneficial temperature acclimation: transcriptomic signatures in Antarctic fish (Pachycara brachycephalum). *Molecular Ecology, 23*(14), 3469-3482.

Wringe, B. F., Devlin, R. H., Ferguson, M. M., Moghadam, H. K., Sakhrani, D., & Danzmann, R. G. (2010). Growth-related quantitative trait loci in domestic and wild rainbow trout (Oncorhynchus mykiss). *BMC Genetics, 11*(1), 63.

Yue, G. H. (2014). Recent advances of genome mapping and marker-assisted selection in aquaculture. *Fish and Fisheries, 15*(3), 376-396.

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

Zhang, X., Tang, N., Hadden, T. J., & Rishi, A. K. (2011). Akt, FoxO and regulation of apoptosis. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 1813*(11), 1978-1986.

Zhao, J., Brault, J. J., Schild, A., Cao, P., Sandri, M., Schiaffino, S., et al. (2007). FoxO3 coordinately activates protein degradation by the autophagic/lysosomal and proteasomal pathways in atrophying muscle cells. *Cell Metabolism, 6*(6), 472-483.

Zhong, H., Xiao, J., Chen, W., Zhou, Y., Tang, Z., Guo, Z., et al. (2014). DNA methylation of pituitary growth hormone is involved in male growth superiority of Nile tilapia (Oreochromis niloticus). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 171*, 42-48.

Appendix

Table A.1. Exon and intron sizes (bp) in the *GH* genes of Perciform and outgroup species used in the phylogenetic analysis in this study. I-VI = exon numbers; 1-5 = intron numbers; (?) = Genbank sequence does not have annotation for full length of exon; (-) = not present in species; bold = highly conserved exon sizes.

Species	**I**	**1**	**II**	**2**	**III**	**3**	**IV**	**4**	**V**	**5**	**VI**
P. major	71	563	**134**	696	114	969	**144**	85	**147**	79	296
C. auratus	71	410	**134**	713	114	3384	**144**	85	**147**	79	296
S. aurata	65	663	**134**	1723	**117**	449	**144**	82	**147**	75	296
N. coibor	?	378	**134**	416	**117**	324	**144**	85	**147**	81	?
P. crocea	?	372	**134**	440	**117**	236	**144**	74	**147**	86	?
D. labrax	?	483	**134**	96	**117**	929	**144**	126	**147**	76	?
S. scherzeri	?	445	**134**	3258	**117**	1006	**144**	86	**147**	94	?
S. chuatsi	63	406	**134**	2914	**117**	1148	**144**	86	**147**	94	296
S. kneri	63	406	**134**	2848	**117**	1149	**144**	86	**147**	94	295
E. coioides	?	107	**134**	93	**117**	719	**144**	99	**147**	1964	?
C. zillii	?	92	**134**	436	**117**	102	**144**	72	**147**	79	?
T. nilotica	75	96	**134**	420	**117**	123	**144**	72	**147**	79	259
S. galilaeus	?	84	**134**	403	**117**	123	**144**	72	**147**	79	?
T. mossambica	76	87	**134**	403	**117**	123	**144**	72	**147**	79	?
L. calcarifer	80	187	**134**	86	**117**	342	**144**	76	**147**	67	235
S. quinqueradiata	105	1247	**134**	98	**117**	1886	**144**	90	**147**	101	231
C. diplogramma	83	513	**134**	99	**117**	405	**144**	74	**147**	85	219
C. marulius	83	682	**134**	98	**117**	393	**144**	75	**147**	78	221
C. gachua	84	661	**134**	101	**117**	321	**144**	80	**147**	78	222
C. punctata	84	661	**134**	101	**117**	321	**144**	80	**147**	78	222
C. striata	84	661	**134**	101	**117**	321	**144**	80	**147**	78	222
P. olivaceus	92	874	**134**	120	**117**	121	**144**	72	105	89	243
C. batrarchus	65	265	140	116	**117**	371	132	299	425	-	-
O. keta	75	453	140	138	**117**	508	156	861	**147**	218	549

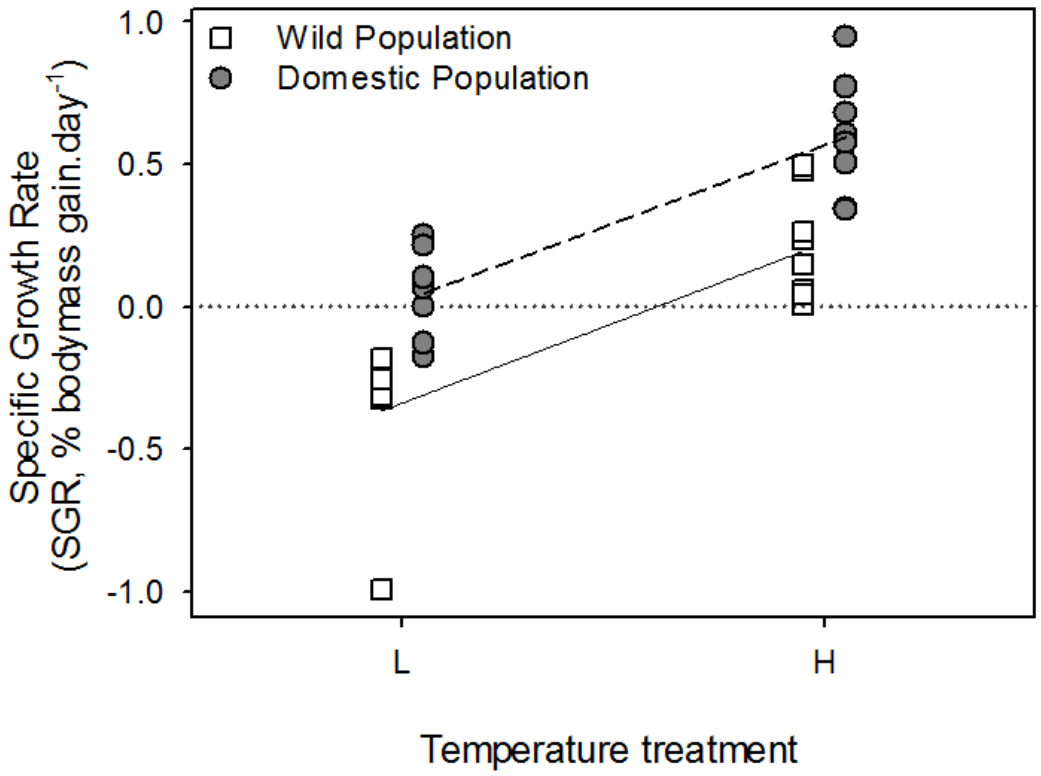

Figure A.1. Specific growth rates of wild and domestic populations of *C. auratus* under different temperature treatments. L = low temperature (13 °C); H = high temperature (21 °C); open boxes = wild population; grey circles = domestic population. This figure was supplied by M. Wellenreuther at PFR from the wider transcriptomic study that this thesis research extended from.