A transcriptomic investigation of digestive processes in orange-spotted grouper, Epinephelus coioides, before, during, and after metamorphic development

Kelli Anderson, Chia-Yu Kuo, Ming-Wei Lu, Ido Bar, Abigail Elizur

PII: S0378-1119(18)30317-2
DOI: doi:10.1016/j.gene.2018.03.073
Reference: GENE 42695
To appear in: Gene
Received date: 8 March 2018
Accepted date: 21 March 2018

Please cite this article as: Kelli Anderson, Chia-Yu Kuo, Ming-Wei Lu, Ido Bar, Abigail Elizur, A transcriptomic investigation of digestive processes in orange-spotted grouper, Epinephelus coioides, before, during, and after metamorphic development. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Gene(2017), doi:10.1016/j.gene.2018.03.073

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
A transcriptomic investigation of digestive processes in orange-spotted grouper, *Epinephelus coioides*, before, during, and after metamorphic development

Authors
Kelli Anderson¹#, Chia-Yu Kuo²,³,⁴#, Ming-Wei Lu²*, Ido Bar¹, Abigail Elizur¹*

Affiliations
¹ Genecology Research Centre, Faculty of Science, Health, Education and Engineering, University of the Sunshine Coast, Locked Bag 4 Maroochydore DC, Qld 4558, Australia.
² Department of Aquaculture, The College of Life Sciences, National Taiwan Ocean University, No.2, Beining Road, Jhongjheng District, Keelung 20224, Taiwan
³ Doctoral Degree Program in Marine Biotechnology, The College of Life Sciences, National Taiwan Ocean University, No.2, Beining Road, Jhongjheng District, Keelung 20224, Taiwan
⁴ Doctoral Degree Program in Marine Biotechnology, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 11529, Taiwan

# Equal first author

* Author to whom correspondence should be addressed, Phone: +61 7 5459 4813, Fax: +61 7 5430 2881, E-mail: AElizur@usc.edu.au

Email addresses: Kelli Anderson, ando.kelli@gmail.com; Abigail Elizur, AEElizur@usc.edu.au; Ido Bar, Ido.Bar@research.usc.edu.au; Chia-Yu Kuo, key3340@gmail.com; Ming-Wei Lu, mingwei@ntou.edu.tw

Abstract
The orange-spotted grouper (OG), *Epinephelus coioides*, is an ecologically and economically important species with strong market demand. However, larval rearing for this species is especially difficult, with mass mortality occurring at multiple stages including the period coinciding with metamorphic development. The aim of the present study was to characterise the molecular ontogenesis of genes that influence appetite, feeding, and digestion in OG larvae head and body tissue at 12, 18, and 50 days post hatch (dph), which coincides with the beginning and end of metamorphic development. The sequences of many transcripts involved in the regulation of appetite, feeding and digestive processes were detected from 12 dph in OG larvae, including those that were differentially expressed in body tissue in fish at different stages of development such as cholecystokinin, peptide Y, and meprin A. Of the transcripts encoding digestive enzymes, only the expression level of bile salt-activated lipase decreased as development progressed. In contrast, a dramatic increase in expression for other body-expressed transcripts encoding digestive enzymes and a proton pump subunit was observed at 50 dph, which is indicative of an increase in digestive capacity. In addition, we have provided evidence suggesting that various trypsinogen isoforms are present, and have differing expression patterns throughout larval development in whole body tissue. We also report on the presence of a prey-specific transcript encoding α-amylase that was present in
the body-transcriptome. Taken together, these results give insight into the processes underpinning attainment of digestive capacity, and form the basis of a new transcriptomic database that will aid further study into the digestive development and dietary requirements of orange-spotted grouper larvae.

**Keywords**

Digestive enzymes; larval development; next-generation sequencing; grouper; trypsinogen; molecular ontogenesis

**Declarations**

*Ethics approval and consent to participate*

Maintenance and handling of animals was conducted in accordance with the Institutional Animal Care and Use Committee (IACUC) of College of Life Sciences, National Taiwan Ocean University, Taiwan (approval number: 102062).

*Consent for publication*

All authors consent to the publication of this work, and are aware of its submission here.

*Availability of data and material*

All raw transcriptomic data used in the current study is publically available for download at https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP115865

*Competing interests*

Authors of the current study have no competing interests to declare.

*Funding*

This study was funded by an Australian Centre for International Agricultural Research grant (FIS-2012-101) awarded to Prof. Abigail Elizur at University of the Sunshine Coast. The study was also supported by the Center of Excellence for the Oceans, National Taiwan Ocean University (Prof. Ming-Wei Lu), at Department of Aquaculture, National Taiwan Ocean University, and partial support was provided by the Ministry of Science and Technology, Taiwan (MOST 104-2321-B-019-004-MY3; MOST 106-3114-B-019-001; MOST 106-3114-B-020-001) to MW Lu.

*Authors’ contributions*

All listed authors contributed to the following study. Profs. Elizur and Lu oversaw all research, provided continual guidance, intellectual input, and provided feedback on the manuscript. Dr. Kelli
Anderson performed the transcriptomic data analysis, qPCR data analysis, and drafted the manuscript. Chia-Yu Kuo collected the larval samples, did the RNA extractions, arranged the next-generation sequencing, and performed the qPCRs. Dr. Ido Bar assembled the transcriptome, provided bioinformatics and IT support, and provided feedback on the manuscript. All authors read and approved the final manuscript.

1.0 Introduction

Orange-spotted grouper (OG), *Epinephelus coioides*, has significant market value throughout Asia, making their aquaculture highly desirable. However, there are significant issues associated with the rearing of OG larvae, which makes large-scale and reliable culture difficult. Difficulties in larval rearing are underpinned by a general lack of knowledge surrounding larval biology, physiology and ecology (Kohno et al. 1997; Spedicato & Boglione 2000; Yamaoka et al. 2000). For example, high mortality is typically observed during weaning and metamorphosis in the orange-spotted grouper (Feng et al. 2008a), and there are currently no effective management strategies that adequately address this issue.

Metamorphosis is a critical stage in larval development, and among other physiological and morphological changes, it is characterised by significant changes to the digestive system, including the formation of a functional stomach and an associated increase in digestive capacity (Govoni et al. 1986; Quinitio et al. 2004a). Poor larval fitness and survival during this stage of development in captive reared OG are thought to cause a subsequent reduction in juvenile growth rate (Feng et al. 2008b). Due to OG’s commercial importance, several studies have been undertaken to characterise the expression of transcripts encoding digestive enzymes (Feng et al. 2008a; Feng et al. 2008b; Liu et al. 2013), and appetite regulators (Yan et al. 2011; Zhang et al. 2013) in fish of various life stages, although there is still relatively little information available for the period of development coinciding with metamorphic development. Furthermore, information regarding the digestive physiology of OG larvae is lacking (Feng et al. 2008b), which hinders efforts to improve larval husbandry.

The availability and activity of digestive enzymes is crucial in determining digestive capacity, which ultimately sets physiological limits and affects growth (Perez-Casanova et al. 2006; Torrissen 1991). In addition, an understanding of molecular processes underpinning digestive capacity is crucial in determining what macronutrients are suited to a particular species throughout development, and in turn this aids in the development of formulated feeds (Cahu & Zambonino Infante 1995). In the context of OG, more information regarding the regulation of feeding, and molecular ontogenesis of the digestive system is required to understand development, fitness, and suitability of feed. To date, there have been no transcriptomic studies in OG characterising the development of the digestive system in terms of appetite, feeding, and digestive regulation, or the appearance of digestive enzymes during metamorphic development. In other species, a transcriptomic approach has been beneficial, where the effect of diet on the transcriptome, and/or changes in the transcriptome over developmental stages were investigated (Li et al. 2015; Morais et al. 2012; Tacchi et al. 2011). These studies were used to characterise transcripts related to digestion, and study transcriptional responses to various diets. For OG larvae, a transcriptomic approach will provide an initial overview of the enzymes, neuropeptides, and hormones that are expressed, and how their expression changes with development. Thus, the aim of the current study was to gain an initial overview of genes that influence appetite, feeding, and digestion in OG larvae, and initiate a database of transcriptomic resources that will aid targeted studies in the future.
2.0 Methodology

2.1 Animal husbandry and sampling

Maintenance and handling of animals was conducted in accordance with the Institutional Animal Care and Use Committee (IACUC) of College of Life Sciences, National Taiwan Ocean University, Taiwan (approval number: 102062).

OG utilised in the present study were of unknown parentage, and may be the progeny of either wild caught or artificially propagated broodstock. One gram of fertilised eggs was collected following natural spawning of OG broodstock, and placed into a 200 L tank with gentle aeration. The tank containing newly hatched larvae was stocked with green algae (*Chlorella vulgaris*, Figure 1), under ambient temperature (25 - 26 °C) and photoperiod, at the National Taiwan Ocean University, Keelung, Taiwan. The tank was cleaned using a siphon, and water quality parameters were monitored daily.

![Graph showing feeding regime](image)

**Figure 1.** Feeding regime for pre and post-metamorphosis OG.

OG larvae were fed live rotifers (*Brachionus plicatilis*) from 3 – 18 dph, copepod nauplii (48 – 180 µm, *Schmackeria dubia* and *Apocyclops royi*) from 12 – 50 dph, and artemia nauplii from 35 dph (further details are provided in Figure 1).

OG were sampled at 12, 18, and 50 dph, beginning on 30/08/2014, 3 – 4 hours after feeding. Twelve dph coincides with the period of development just prior to the beginning of metamorphosis, at 18 dph larvae are metamorphosing (second dorsal and pelvic fin spines reach their maximal length), and metamorphic development is complete prior to 50 dph. Head and body segments were dissected separately, placed into TRIzol (Life Technologies, USA), and then stored at -80 °C.

2.2 RNA extraction

Total RNA was extracted from separate pools of homogenised head and body tissue from 150, 130, and 4 animals for 12, 18 and 50 dph respectively (total six samples) using TRIzol according to the
manufacturer’s specifications. RNA integrity, quantity and quality was assessed by Genomics BioSci & Tech Co., Ltd by analysing samples on an Agilent 2100 Bioanalyzer, RNA 6000 nano chip. Results ranged between RIN: 9.0 – 9.9 and 260/280: 1.92 – 2.0. Head and body tissues were separated to aid in detecting transcripts involved in the central (hypothalamic) regulation of appetite that may have low expression, versus highly expressed enzymes in the digestive system that are involved in the breakdown of nutrients. Pooling tissues or RNA has been a common approach in RNAseq studies for a range of marine species including fish (Shi et al. 2014), shrimp (Guo et al. 2013), copepods (Ning et al. 2013), and snails (Gleason & Burton 2015), where difficult or small tissues have been utilised.

2.3 RNA-sequencing and assembly

Library preparation and sequencing was performed by Genomics BioSci & Tech Co., Ltd according to protocols specific to the Illumina NextSeq platform with multiplexed, 150 bp, paired-end chemistry to achieve a minimum of 7.5 Gb per sample. Minor adaptor trimming of the reads from each tissue type was performed using BBduk (from BBMap tool v36.02), and average read quality was found to be Phred 34 for both tissue types using FastQC v0.11.5.

De novo assembly and quality assessment of the two tissue-specific transcriptomes was performed with Trinity v2.2.0 (and the scripts available at https://github.com/trinityrnaseq/trinityrnaseq/wiki) using the default parameters, which assembles transcripts at the isoform level (Haas et al. 2013b). Assessment of the assembled transcripts was performed by calculating and plotting an ExN50 value against a fraction of the most highly expressed transcripts (Ex). This plot enabled identification of the assembly saturation point, at which the maximum length of N50 was obtained, after removal of the transcripts with minor contribution to the total expression, which are often associated with assembly errors (Haas et al. 2013b).

2.4 Assembly annotation

The head and body assemblies were annotated using selected sections of the Trinotate pipeline (v3.0.0, http://trinotate.github.io/) as follows. Transdecoder was initially used to predict open reading frames (ORFs)/coding regions of transcripts, and obtain putative ORF peptide sequences for subsequent homology searches (v3.0.0, https://transdecoder.github.io/). The most likely ORF protein domains were searched against a locally stored Pfam-A database (v28, Finn et al. 2014) using HMMER (v3.1b1, http://hmmer.org). Specific gene sequences were annotated against the Swiss-Prot database, the manually annotated subsection of the UniProtKB knowledgebase, using BLASTP (Altschul et al. 1990; Camacho et al. 2016). Similarly, the whole nucleotide sequence for each transcript was searched against Swiss-Prot database using BLASTX (Altschul et al. 1990; Camacho et al. 2016), and the results from all three homology searches were uploaded to a preconfigured SQLite3 database as per Trinotate guidelines for each tissue type. Gene Ontology (GO) terms were automatically assigned during this stage based on matching proteins in Swiss-Prot, and protein family domains from Pfam databases.

To provide additional annotation, all assembled sequences were searched against the National Center for Biotechnology Information’s (NCBI) non-redundant nucleotide database (nt) using BLASTN via an in-house parallel computing method (https://github.com/IdoBar/par-spanner), with the BLAST parameters ‘e-value 1e-5, -max_target_seqs 1, and -max_hsps 1’. The output from this step was
manually uploaded to the SQLite3 database for each tissue type, and the single most reliable hit for each transcript was retained based on maximum bit-score.

2.5 Differential expression

To determine which transcripts were differentially expressed (DE), reads were mapped back to the assembled transcriptome in a tissue-specific manner using Bowtie2 short sequence aligner, v2.2.9 (Langmead & Salzberg 2012), and transcript abundance was estimated using RSEM, v1.2.28 (Li & Dewey 2011). The estimated counts were normalised using the Trimmed Mean of M-values (TMM, with dispersion set to 0.2), a normalisation method implemented in the edgeR Bioconductor package, to account for differences in library size between samples (Dillies et al. 2013; Robinson et al. 2010). The normalised count matrix was then used by edgeR to identify statistically significant DE transcripts over time for each sample type. During this process, raw P-values were adjusted for multiple comparisons by the Benjamini–Hochberg procedure (Benjamini & Hochberg 1995), which controls the false discovery rate (FDR). DE transcripts were then extracted using the analyse_diff_expr.pl script available in the Trinity suite, with an adjusted P-value (FDR cut-off) of 1e-3, and a C value of 2 (minimum 4 fold change). DE transcripts were also extracted at P 1e-2 and C 1 (minimum 2 fold change) which is considered a less stringent threshold. As a single pooled sample was utilised at each time point in the current study, the results from this analysis are intended to provide an initial overview of genes involved in digestion and the regulation of feeding, and serve as a starting point for further research.

Due to the large range of fold change among DE transcripts, the z-score of Fragments Per Kilobase of transcript per Million mapped reads (FPKM) values for transcripts of interest whose expression was >1 TMM FPKM was used for graphical purposes. The z-score for each transcript in each sample was calculated using the formula $z = \frac{(X - \mu)}{\sigma}$, where $z$ = z-score, $X$ = TMM FPKM, $\mu$ = mean TMM FPKM across time points, and $\sigma$ = standard deviation. Heatmaps were generated using GraphPad Prism, v7. When multiple transcripts with identical annotations were present and had similar expression patterns, a single representative transcript was chosen for visualisation purposes.

GO terms of interest were then used to extract data for DE transcripts involved in digestive processes from each tissue-specific SQLite3 database. For example, GO term GO:00077586 ‘digestion’ was used to extract DE transcripts involved in that process. To extract data for additional DE transcripts of interest missed during the process above, each SQLite3 database was also queried using a custom list of digestion-related gene names obtained via extensive literature review.

2.6 cDNA synthesis for qPCR analysis of target genes

qPCR analysis was undertaken to determine the reliability of DE results from the RNA-seq pipeline. The same pooled RNA samples that were used for library preparation were also used for qPCR analysis. Five hundred nanograms of RNA was used to synthesise cDNA for each tissue type, at each time point, for use in quantitative real-time PCR (qPCR) using HiScript I Reverse Transcriptase (BIONOVAS, Canada). Reverse transcription was conducted in accordance with the manufacturer’s instructions for cDNA synthesis with random primers. Following synthesis, cDNAs were stored at −20 °C.
2.7 qPCR primer design

Gene specific primers (GSPs) for nucleobindin 2 (Nucb2), leptin b (Lepb), calcitonin receptor (Calcr), trypsinogen 2 (Try2), bile salt-activated lipase 2 (Bal), α-amylase (α-Amy), trypsinogen 3 (Try3), meprin A B (MepAβ), apolipoprotein A-IV (ApoA4), and candidate reference genes: β-actin, elongation factor 1-α (Ef1α), ADP ribosylation factor guanine nucleotide exchange factor 1 (Arfgef1), neuron navigator 3 (Nav3) and beta-2 microglobulin (B2m) were designed from species specific mRNA sequences using Primer3 software (http://frodo.wi.mit.edu/primer3/, Table 1). All primers were designed to have an optimum annealing temperature of 60 °C and were supplied by Tri-I Biotech, Inc. (Taiwan).

Table 1. qPCR primers

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Sequence (5’ → 3’)</th>
<th>Prod. size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucb2</td>
<td>F GAA GGT GAA CCA GCA CCA GT</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>R GGC CTG CTA AGA GTG TCC TG</td>
<td></td>
</tr>
<tr>
<td>Lepb</td>
<td>F TCC TGG GTG CTC AAG TCT TC</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>R TGG ACC AGG GTT ATC TGA GC</td>
<td></td>
</tr>
<tr>
<td>Calcr</td>
<td>F GGA AGA ACC AAC CCC TCC TA</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td>R GTG CTC AGC TGG ACA TGG TA</td>
<td></td>
</tr>
<tr>
<td>Try2</td>
<td>F CCC GCT ACC CTC AAC AAC TA</td>
<td>110</td>
</tr>
<tr>
<td></td>
<td>R CTT CCA GAG CTG CTG GTG TT</td>
<td></td>
</tr>
<tr>
<td>Try3</td>
<td>F TGc TGG CAG AAC CCT TAT TC</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td>R ATG GCA TCA ACT GGC ATG TA</td>
<td></td>
</tr>
<tr>
<td>Bal</td>
<td>F GGT TCT ATG GGG CCA AAC TT</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>R CAC AGG GTA TCC CAC TGA CA</td>
<td></td>
</tr>
<tr>
<td>α-Amy</td>
<td>F CAT CAA CCC TGA CCA GAC CT</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>R CGT TCC ATT GAC CAC ATT GC</td>
<td></td>
</tr>
<tr>
<td>MepAβ</td>
<td>F CGC AGG GAG Ctg CTA ATA AC</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>R GCG TTC CCT ACA ATC ACG AT</td>
<td></td>
</tr>
<tr>
<td>ApoA4</td>
<td>F CCA GCT GAT GTC CCT CTA CC</td>
<td>93</td>
</tr>
<tr>
<td></td>
<td>R CAA ACA AGG CCA GCA TCT TT</td>
<td></td>
</tr>
<tr>
<td>β-actin*</td>
<td>F TCC ACC GCA AAT GCT TCT AA</td>
<td>121</td>
</tr>
<tr>
<td></td>
<td>R TGC GCC TGA GTG TGT ATG A</td>
<td></td>
</tr>
<tr>
<td>Ef1α*</td>
<td>F ACG CAG GGG AGC TCT TGA TAC T</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>R ATC TTT TCC TTT CCC ATT TCG</td>
<td></td>
</tr>
<tr>
<td>Arfgef1*</td>
<td>F CAG CAC TTT ACC GCC AAT CAA</td>
<td>126</td>
</tr>
<tr>
<td></td>
<td>R TGT AAA CAG TCG AGC GAG GT</td>
<td></td>
</tr>
<tr>
<td>Nav3*</td>
<td>F AGG GAA GGA GTG GTT GAG GT</td>
<td>128</td>
</tr>
<tr>
<td></td>
<td>R GGC TCA GGA GTG TGG AGT AG</td>
<td></td>
</tr>
<tr>
<td>B2m*</td>
<td>F GGA CAG CAT GGC AAA GAC AAC</td>
<td>157</td>
</tr>
<tr>
<td></td>
<td>R GCT TGG TCA GAT GGA AGT G</td>
<td></td>
</tr>
</tbody>
</table>

bp = base pairs, *candidate reference gene

2.8 qPCR protocol

qPCRs were conducted on a Toptical Thermocycler (Biometra, Germany) using qPCRBIO SyGreen Mix (PCR Biosystems, UK) as the fluorescent label according to the manufacturer’s instructions. Validation curves were carried out in triplicate using serially diluted cDNA as the template and the following cycling conditions: 95 °C for 2 min; 40 cycles of 95 °C for 5 s; 60 °C for 30 s (acquiring). At the end of cycle 40, all primers were tested for specificity via melt curve analysis which consisted of
a 90 s preconditioning step at 72 °C, followed by a temperature gradient up to 95 °C at 1 °C per 5 s. Reaction efficiencies were automatically calculated by the qPCR Efficiency Calculator (Thermo Fisher Scientific) using the equation: \( E = \left[ 10^{-1/M} \right]^{-1} \), where \( E \) is equal to efficiency and \( M \) is equal to slope.

To determine the relative expression of target genes, each qPCR reaction contained 10 \( \mu \)L SYBR, 10 \( \mu \)M each primer, 1 \( \mu \)L of 10-fold diluted cDNA template, and ddH2O to a final volume of 20 \( \mu \)L. Cycling conditions and melt curve analysis were performed as described above. The suitability of using the five candidate reference genes for normalisation was assessed using RefFinder (Xie et al. 2012), which utilises the GeNorm (Vandesompele et al. 2002), BestKeeper (Pfaffl et al. 2004), and NormFinder (Andersen et al. 2004) algorithms to assess gene expression stability. For pooled body tissue, BestKeeper predicted that Arfgf1 and \( \beta \)-actin were the most stable genes (stability values 0.149 and 0.213 respectively). Similarly, for pooled head tissue BestKeeper also predicted that Arfgf1 and \( \beta \)-actin were the most stable genes (stability values 0.316 and 0.438 respectively). BestKeeper rankings appeared to be the most reasonable, and were supported by the low level of variability observed between samples for those genes in both head and body tissue (Figure 2).

Consistent with our previous work (Anderson & Elizur 2012), all algorithms were in agreement when ranking the least stable genes which were Nav3 and B2m in body tissue, and B2m in head tissue. Thus, the expression of target genes for head and body tissue were normalised using Arfgf1 and \( \beta \)-actin, as the stability of these genes was stable across time points. Relative expression of target genes for each sample was calculated using the Pfaffl method (Pfaffl et al. 2002).

2.9 Phylogenetic analysis, and tentative classification of trypsinogens

As BLASTN annotation for putative trypsinogens (Try) was somewhat ambiguous, phylogenetic analyses of DE trypsinogen transcripts were carried out in an attempt to better classify those
transcripts. To ensure that coverage was adequate before phylogenetic analysis, read support for all DE sequences was manually assessed using Integrative Genomics Viewer v2.3 (Robinson et al. 2011). Samtools v0.1.18 (Li et al. 2009) was used to prepare files for upload as required. Then, in order to classify each putative Try transcript, reference Try sequences from fish species were downloaded from GenBank, and aligned with the nucleotide sequences of DE Try transcripts using ClustalW (Thompson et al. 1994). Using Mega7 (Kumar et al. 2016), phylogenetic trees were then constructed using the neighbour joining method (Saitou & Nei 1987) with gaps deleted, and the reliability of each tree was assessed via bootstrapping with 1000 replicates (Felsenstein 1985). Phylogenetic trees were constructed concurrently only when putative Try sequences were of a very similar length and contained homologous regions. If those criteria were not met, transcripts were assessed in separate trees.

3.0 Results and discussion

3.1 De novo transcriptome assembly and annotation

There is currently no publicly available reference genome for E. colioides, thus a pool of 169,888,860 and 156,501,272 trimmed reads were used to assemble de novo transcriptomes for head and body tissue respectively (all raw data is publicly available). Descriptive data for each assembly are presented in Table 2. In the absence of a well annotated reference genome, it is possible that some ‘isoforms’ may in fact originate from paralogous genes or misassemblies. Thus, many studies focus on differential expression at the gene level and do not include information on the presence of potential isoforms (e.g. Sun et al. 2016). However, the importance of differential splicing events on organism development and function has been demonstrated by many studies. For example, growth rates varied significantly depending on which Try variants were present in the pyloric caeca of Atlantic salmon (Salmo salar) (Torrissen 1991). In the context of the present study, such information could help to shed light on the specific needs of OG during larval development, or provide areas for further research. Therefore, in a similar fashion to Schunter et al. (2014), subsequent analyses took isoform predictions into account.

Table 2. Descriptive statistics for the head and body reference transcriptomes

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Head</th>
<th>Body</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isoforms</td>
<td>461,615</td>
<td>239,710</td>
</tr>
<tr>
<td>Genes</td>
<td>400,629</td>
<td>209,681</td>
</tr>
<tr>
<td>Longest contig</td>
<td>17,018 bp</td>
<td>35,370 bp</td>
</tr>
<tr>
<td>N20</td>
<td>2,963 bp</td>
<td>3,126 bp</td>
</tr>
<tr>
<td>N50</td>
<td>1,011 bp</td>
<td>1,270 bp</td>
</tr>
<tr>
<td>Ex90N50</td>
<td>1,957 bp</td>
<td>1,916 bp</td>
</tr>
</tbody>
</table>

BLAST annotation of the head and body reference assembly against NCBI’s nucleotide database, resulted in 75.56 and 89.11% of transcripts having significant hits, respectively. The percent of transcripts assigned a GO term through the Trinotate pipeline was low, with GO terms assigned to just 18.12 and 27.91% of head and body transcripts, respectively. Thus, for transcript annotation at the gene name level, blasting to the NCBI nucleotide database was far more effective than the other means tested. As suggested by Schunter et al. (2014), it is possible that the presence of orphan proteins, un-translated regions, and misassemblies may account for the low homology with known
proteins. In addition, working with non-model species is notoriously difficult due to the lack of data from closely related species (Matz 2017). Hence, the percent of contigs assigned GO terms was similar to other studies in fish (Chen et al. 2014; Patel et al. 2016), and the low hit rate would exclude a large percentage of contigs from downstream data mining and analyses that utilise GO terms. The work presented here is the first to take a transcriptomic approach to studying aspects of OG physiology before, during, and after metamorphic development, and represents an important first step to understanding transcriptome-wide gene expression in the stages studied.

3.2 Differential gene expression

Using the threshold: minimum 4-fold difference in expression, adjusted P-value 1e-3, and dispersion 0.2, 2,316 head-expressed isoforms were identified as differentially expressed over time. While at the less stringent threshold (minimum 2-fold change), this number increased to 5,293 DE isoforms. Using the former criteria, 1,669 body-expressed isoforms were differentially expressed over time. While at the latter, this number increased to 3,858.

The reliability of DE results from the RNA-seq pipeline were determined via qPCR, and was found to be acceptable for the body transcriptome. For example, transcript expression was maximal at 50 dph for mepAβ, try2, try3, and nuc2, decreased over time for bal, and was highest at 18 dph for α-amy according to both methods of quantification (Figure 3).

**Figure 3.** Z-score of expression levels for pooled samples as determined by next generation sequencing (RNAseq, TMM FPKM) and qPCR (relative expression) for body-expressed transcripts at 12, 18 and 50 dph. For
qPCR, the expression of target transcripts was normalised against β-actin and Arfgef1. Higher and lower z-scores are reflective of higher and lower gene expression levels respectively.

For head tissue, many factors thought to play a role in appetite and the regulation and feeding appeared to be DE. However, the results for several of these genes could not be confirmed via qPCR (data not shown). This is likely due to the low expression of neurotransmitters of interest relative to the expression of transcripts involved in other more dominating processes (for example, eye development), which resulted in low expression levels (<1 TMM FPKM) for many transcripts of interest. This is likely, since the expression of neurotransmitters in the brain is typically low (van den Pol 2012), and as highlighted in a review by Bräutigam and Gowik (2010), quantifying the expression of transcripts involved in signalling and regulation using RNA-seq pipelines can be difficult. In a similar study on Atlantic cod (Gadus morhua) larvae, Kortner et al. (2011) noted that the expression level of appetite regulating neurotransmitters (orexin and neuropeptide Y) and a peptide hormone (cholecystokinin) may have been influenced by the proportion of brain tissue to other tissue types when whole larvae were utilised. For these reasons, the following sections refer to head-expressed transcripts for genes of interest in a qualitative sense only.

3.3 Phylogenetic analysis, and tentative classification of DE trypsinogens

To better annotate putative Try transcripts that were DE in the body transcriptome, phylogenetic analyses were undertaken. Based on this analysis (for example, Figure 4), three, two, and two transcripts were partitioned into the groups Try 1, 2, and 3 respectively. Two transcripts did not group with the reference nucleotide sequences included in the analysis. These transcripts were reanalysed using protein sequences in the place of nucleotide sequences, which allowed the inclusion of additional reference sequences for species where Try nucleotide sequences were not available. Using this method, one putative Try protein sequence clustered convincingly with ‘trypsin-1’ from large yellow croaker, Larimichthys crocea (Figure 5), and has been tentatively annotated as Try1, bringing the total number of putative Try1 transcripts to four. For the other putative Try, reanalysis did not help in elucidating its classification, and this transcript has been annotated as ‘trypsinogen A’. Labelling for graphical purposes was based on this analysis, although further experimental analysis is required to confirm the sequence of each transcript, as artefacts can be created during the RNA-Seq pipeline (González & Joly 2013; Haas et al. 2013a).
**Figure 4.** Example of a phylogenetic relationship for a body-expressed trypsinogen transcript constructed using the neighbour joining method and 1000x bootstrapping. The Latin name for each reference species, the associated GenBank reference number, and bootstrap values over 50% are displayed. The transcript marked with a black circle has been annotated as trypsinogen 3 for visualisation purposes in subsequent figures. Note that trypsinogen Y is a member of the trypsinogen 3 group.
Figure 5. Example of a phylogenetic relationship for a body-expressed trypsinogen transcript where using protein instead nucleotide sequences in combination with additional reference sequences resulted in grouping with an acceptable level of confidence. The transcript marked with a black square has been tentatively annotated as trypsinogen 1 for visualisation purposes in subsequent figures. Other details as per Figure 4.

3.4 Regulation of appetite, feeding and digestion

Transcripts detected in head tissue in the current study with well supported or putative roles in the regulation of appetite and feeding behaviour are listed in Table 3, and DE transcripts from the body transcriptome appear in Figure 6D. For head-tissue, all transcripts listed were present from 12 dph, albeit at low levels (data not shown).
There are many internal and external variables that affect appetite and feeding in fish, including genetic makeup, temperature, stress, photoperiod, life stage and reproductive status, energy balance, food availability and gut contents (Rønnestad et al. 2017). These factors influence a complex network of regulating neuropeptides, neurotransmitters, hormones, and their receptors in brain and peripheral tissues, with the end result being an increase or decrease in appetite and feeding (Martínez-Álvarez et al. 2009; Volkoff 2016). In the current study, peptides involved in the central (hypothalamic) regulation of appetite were detected in head tissue including, neuropeptide Y (Npy), agouti-related protein (Agrp), cocaine- and amphetamine-regulated transcript (Cart), orexin (Ox), Galanin (Gal), corticotropin-releasing hormone (Crh), and members of the melanocortin system (such as pro-opiomelanocortin precursor). Npy has a powerful orexigenic effect in teleosts, and is able to drive changes in feeding behaviour (Volkoff 2006; Yokobori et al. 2012). In addition, the activity of Npy is mediated through multiple receptors that were also detected in the current study (Narnaware & Peter 2001). In the context of OG larvae feeding, survival, and growth, this particular transcript is of interest as Npy increases growth hormone levels, weight gain, and feed conversion efficiency in E. coioides (Wu et al. 2012), and elevated npy expression can be indicative of food deprivation in many species (Rønnestad et al. 2017). Similarly, Ox is an orexigenic neuropeptide that is able to stimulate the expression of npy in E. coioides (Yan et al. 2011), and whose expression changes throughout larval development in some species such as G. morhua (Kortner et al. 2011). Like npy, ox levels tend to increase around feeding, and following fasting (Rønnestad et al. 2017), and may therefore be useful for assessing appetite, feeding, and development of OG larvae.

Peripheral signals originating in the gastrointestinal tract (GIT), thyroid, pancreas, and adipose tissue locally regulate digestion, and provide feedback that influences the regulation of appetite and feeding. For example, cholecystokinin (Cck) and pancreatic peptide Y (Py has been renamed as Pyyb, (Sundström et al. 2008)) levels are influenced by dietary nutrients, and are thought to have antagonistic roles in the regulation of pancreatic enzyme secretion (Murashita et al. 2008). This regulatory mechanism has been observed in juvenile yellowtail, Seriola quinqueradiata, where inverse expression patterns were observed for cck and pyyb in response to dietary protein and/or fat (Murashita et al. 2008). Furthermore, Cck level appears to be modulated by gut content in flatfish (Solea senegalensis) larvae (Navarro-Guillén et al. 2017). For OG body tissue, the expression of cck decreased from 12 and 50 dph, and an opposing pattern was observed for pyyb expression (Figure 6D). Whether the inverse expression pattern observed for these transcripts is a reflection of larval gut contents, or presumed antagonistic relationship to one another requires further investigation.

Table 3. Transcripts detected in OG larvae head tissue, their effect on feeding, other functions, and primary site of synthesis

<table>
<thead>
<tr>
<th>Orexigenic</th>
<th>Other functions, primary site of synthesis (ligand)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agouti-related protein</td>
<td>Brain</td>
<td>(Volkoff et al. 2005)</td>
</tr>
<tr>
<td>Galanin, receptors 1, 2</td>
<td>Brain, central nervous system (CNS), gastrointestinal tract (GIT)</td>
<td>(Li et al. 2013; Volkoff &amp; Peter 2001)</td>
</tr>
<tr>
<td>Ghrelin</td>
<td>Growth hormone secretagogue, some reports of anorexigenic activity, GIT</td>
<td>(Kojima et al. 1999; Miura et al. 2007; Rønnestad et al. 2017)</td>
</tr>
<tr>
<td>Growth hormone</td>
<td>Swimming activity, immune function, pituitary</td>
<td>(Johansson et al. 2005; Shved et al. 2011)</td>
</tr>
<tr>
<td>Neuromedin S</td>
<td>Brain, ovary, pituitary</td>
<td>(Chen et al. 2016)</td>
</tr>
<tr>
<td>Neuropeptide Y, receptors 1, 2, 4, 8</td>
<td>Regulation of metabolism, body temperature, circadian rhythm, CNS, GIT</td>
<td>(Narnaware &amp; Peter 2001; Volkoff et al. 2010; Zhou et al. 2013)</td>
</tr>
</tbody>
</table>
### Orexin (hypocretin), receptor 2

Influence diurnal behaviour, brain  
(Kortner et al. 2011; Panula 2010; Volkoff et al. 1999)

### Anorexigenic

<table>
<thead>
<tr>
<th>Peptide/Receptor</th>
<th>Function</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apolipoprotein A-IV</td>
<td>Satiation factor (rat), intestine (rat)</td>
<td>Fujimoto et al. 1993</td>
</tr>
<tr>
<td>Calcitonin gene-related peptide, and receptor</td>
<td>Glucose homeostasis, brain</td>
<td>Martinez-Álvarez et al. 2009; Murphy &amp; Bloom 2006</td>
</tr>
<tr>
<td>Cholecystokinin and receptor</td>
<td>Release of pancreatic enzymes, gut peristalsis, GIT, brain</td>
<td>Murphy &amp; Bloom 2006; Volkoff 2006</td>
</tr>
<tr>
<td>Cocaine- and amphetamine-regulated transcript Corticotropin-releasing hormone and receptor</td>
<td>Brain</td>
<td>Volkoff et al. 2016</td>
</tr>
<tr>
<td>Gastrin-releasing peptide/bombesin and receptor</td>
<td>Stimulate release of adrenocorticotropic brain, gastric secretion and motility, GIT</td>
<td>Himick &amp; Peter 1994; Rønnestad et al. 2017</td>
</tr>
<tr>
<td>Glucagon-like peptide and receptor</td>
<td>Gastric evacuation, GIT</td>
<td>Silverstein et al. 2001</td>
</tr>
<tr>
<td>Melanocyte-stimulating hormone receptor/melanocortin receptor</td>
<td>Energy homeostasis, brain and peripheral tissues (receptors)</td>
<td>Matsuda et al. 2008a; Volkoff et al. 2005</td>
</tr>
<tr>
<td>Melatonin receptors</td>
<td>Reproduction, antioxidant, HPG-axis, peripheral</td>
<td>Maitra &amp; Hasan 2016; Pinillos et al. 2001</td>
</tr>
<tr>
<td>Neuropeptide B</td>
<td>Possible regulator of feeding in teleosts, growth, energy homeostasis, brain</td>
<td>Yang et al. 2014</td>
</tr>
<tr>
<td>Pituitary adenylate cyclase-activating polypeptide receptor</td>
<td>Locomotor activity, brain</td>
<td>Matsuda et al. 2005</td>
</tr>
<tr>
<td>Prolactin-releasing peptide and receptor</td>
<td>Hydromineral balance, brain</td>
<td>Kelly &amp; Peter 2006</td>
</tr>
<tr>
<td>Pro-opiomelanocortin precursor</td>
<td>Precursor for adrenocorticotropic and melanocyte-stimulating hormones, pituitary</td>
<td>Volkoff et al. 2005</td>
</tr>
<tr>
<td>Tachykinin</td>
<td>Contraction in smooth gut muscle, brain</td>
<td>He et al. 2013; Olsson &amp; Holmgren 2001; Volkoff et al. 2005</td>
</tr>
<tr>
<td>Thyrotropin-releasing hormone</td>
<td>Increase feeding and locomotion, hypothalamus</td>
<td>(Abbott &amp; Volkoff 2011)</td>
</tr>
</tbody>
</table>

### Other

<table>
<thead>
<tr>
<th>Peptide/Receptor</th>
<th>Function</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-hydroxytryptamine (serotonin) receptors</td>
<td>Foraging behaviour, aggression, appetite regulation, CNS</td>
<td>Rønnestad et al. 2017</td>
</tr>
<tr>
<td>Gastrotropin</td>
<td>Intracellular fatty acid trafficking and metabolism, gut</td>
<td>Caldurch-Giner et al. 2016; Her et al. 2004</td>
</tr>
<tr>
<td>Meprin A</td>
<td>Degradation of Pyy, possible indicator of phosphorus deficiency, intestine</td>
<td>Addison et al. 2011; Kirchner et al. 2007</td>
</tr>
<tr>
<td>Neuropeptide B</td>
<td>Possible regulator of feeding in teleosts, growth, energy homeostasis, brain</td>
<td>Yang et al. 2014</td>
</tr>
</tbody>
</table>

In higher vertebrates, and some teleosts (Gonzalez & Unniappan 2010), peptide YY (Pyy) is a ‘gut-brain’ hormone that has been shown to reduce appetite by acting through the Y family of receptors.
In rodent models MEPRIN is able to degrade PYY (Addison et al. 2011), and given that intestinal transcript levels change in response to dietary phosphorus in salmon (Li et al. 2015), it is reasonable to hypothesise that meprin may play a role in the regulation of appetite, or is responsive to diet in fish. mepAβ was differentially expressed in body tissue, with the highest expression observed at 50 dph (Figure 6D). As this is the first study to report on the abundance of this transcript during larval development in grouper, and since the physiological role of meprin is yet to be established experimentally in fish, this result highlights a need for further research into the processes regulating appetite, feeding and digestion.

Various studies report on the differential expression of peripheral regulating factors throughout larval development that were not differentially expressed in the current study. For example, ghrelin mRNA levels increase significantly in metamorphosing Atlantic halibut (*Hippoglossus hippoglossus*) larvae (Manning et al. 2008), while expression in the current study was stable at all time points (data not shown). Similarly, the expression of other important regulating factors appeared to be stable in the body transcriptome, including leptin and leptin receptor, and galanin and its receptors (data not shown). However, this data is difficult to interpret due to a lack of information concerning these transcripts during digestive ontogenesis in fish larvae, especially grouper.

The transcripts encoding genes typically associated with peripheral tissues were detected in head tissue (Table 3), albeit at low levels. This finding was not unexpected, as gene expression is often observed in multiple tissues in lower levels than what is observed for the primary site of synthesis. For example, in adult *G. morhua*, ghrelin is highly expressed in the stomach, and weakly expressed in brain tissue (Xu & Volkoff 2009). As such, the current study has provided a basis for further targeted analyses of appetite regulation in OG, and this data will be useful in distinguishing between isoforms.

![Figure 6. Body tissue. Z-scores of TMM normalised FPKM values for DE transcripts thought to be involved in A: protein digestion, B: carbohydrate digestion, C: lipid digestion, and D: regulation of digestion or other processes. Asterisk: where multiple DE isoforms are present with the same annotation and similar gene expression pattern, only one representative is displayed. The number next to the asterisk is the total number of DE transcripts with the same annotation and similar expression pattern. Red represents a lower, and green represents a higher level of gene expression.](image-url)
that are differentially expressed on the basis of tissue type, developmental stage or nutritional status.

3.5 Protein digestion

In higher vertebrates, the process of digestive enzyme activation is thought to begin with the activation of proenteropeptidase by duodenase, the resultant enteropeptidase then cleaves the N-terminal activation peptide from trypsinogen, yielding trypsin which then goes on to activate other pancreatic zymogens (Zamolodchikova et al. 2000). In body tissue, the expression of duodenase-1 was higher at 50 dph relative to other time points, which is broadly consistent with the pattern observed for other proteases such as enteropeptidase (Figure 6A). Bovine duodenase displays dual specificity, with trypsin and chymotrypsin-like activity (Zamolodchikova et al. 1995), and likely plays an important role in the activation cascade of digestive proteases (Zamolodchikova et al. 2000). Whether duodenase-1 makes the ‘first cut’ in the zymogen activation cascade, and is a crucial step in the activation of digestive enzymes in grouper (or other teleosts) remains to be explored.

For marine fish larvae that lack a functional stomach at hatching, such as groupers (Eusebio et al. 2004; Wu et al. 2011), pancreatic enzymes are typically the main source of proteolytic action (Zambonino Infante & Cahu 2001). Pancreatic trypsin is generally considered to be the most important proteolytic enzyme during early development in teleost larvae, as it plays a key role in hydrolysing proteins and activating other digestivezymogens (Rønnestad et al. 2013). Trypsinogens can be classified into at least three groups (1, 2, and 3), according to their amino acid sequences and biochemical properties, and multiple isoforms have been isolated (Gudmundsdóttir & Pálsdóttir 2005; Liu et al. 2013; Liu et al. 2012; Manchado et al. 2008) and predicted (Li et al. 2015) in teleosts. In the current study, DE Try isoforms were tentatively sorted into group 1, 2, or 3, where possible, as a result of phylogenetic analysis (for example, Figure 4). For body-tissue, four putative Try1 isoforms were isolated, and their expression was maximal at either 18 or 50 dph depending on the transcript (Figure 6A). Similarly, a study by Liu et al. (2013) sequenced and characterised the expression of two group 1 Try isoforms in E. coioides larvae from hatching to 50 dph. While the authors for that study found that the expression patterns of both isoforms were similar, and increased sharply from 30 or 35 dph, they noted that one isoform was widely expressed in many tissue types, and the other was expressed only in digestive tissues. Thus, the authors hypothesised on the potentially varying roles of group 1 isoforms.

In addition to Try1, the expression of two putative OG Try2 isoforms were quantified, and their expression was maximal at either 18 dph or 50 dph depending on the isoform. As the current study utilised pooled samples, it’s not currently clear whether each larvae expressed multiple Try1 and 2 isoforms, or whether isoforms were larvae-specific due to genetic variation between individuals. In other teleost species, the synthesis of several Try isozymes has been noted, with differences in catalytic efficiency, and the degree of adaptation to cold climate varying between the G. morhua trypsinogens (Gudmundsdóttir & Pálsdóttir 2005). In addition, differences in substrate specificity been noted between Try groups originating from distinct loci (Gudmundsdóttir & Pálsdóttir 2005), and changes in growth rate have been linked to the presence of specific Try isoforms in S. salar fry (Torrissen 1987; Torrissen 1991). Therefore, the divergent expression pattern among group 1 and 2 isoforms in the present study adds to the body of evidence suggesting different roles and/or functionality of various isoforms, and further elucidation of their structure, sites of expression, and importance is warranted.
The two body-expressed DE isoforms that were classified as group 3 Try (for example, Figure 4) had significantly higher expression at 50 relative to 12 and 18 dph (Figure 6A). In *S. senegalensis*, it has been suggested that group 1 and 2 Try variants are important for larval digestive processes, while group 3 Try is more prominent during the juvenile phase (Manchado et al. 2008). This may also be the case for OG, as expression was highest at the end of metamorphosis for both putative isoforms. However, this is the first description of putative group 3 isoforms and their expression in grouper, and generally speaking, data detailing the expression of group 3 Try in larvae is sparse, and does not take into account the possible presence of multiple isoforms. Thus, this data provides a starting point for a more in-depth analysis to confirm the presence of multiple group 3 isoforms, and elucidate their role in digestive processes, growth and development in larvae. Overall, the expression data for Try reported here is generally consistent with the trajectory for total trypsinogen activity in *E. coioides*, where activity increases dramatically during late metamorphic development (Eusebio et al. 2004).

In the current study, expression of the novel body-expressed transcript CUB and zona pellucida-like domain-containing protein 1 (alternate name: integral membrane-associated protein 1) was significantly elevated at 50 dph relative to earlier time points (Figure 6A). While this protein may perform a range of physiological functions, it is tightly associated with pancreatic zymogen granule membranes, and has been strongly implicated in the activation of TRY in mice (Imamura et al. 2002). Chymotrypsinogen is produced in the pancreas, and released into the lumen of the intestine where the active form digests proteins, though its substrate specificity is broader than that of trypsin (Klomklao 2008; Rønnestad et al. 2013). In OG body tissue, chymotrypsinogen C transcript abundance was significantly elevated at 50 dph relative to earlier time points (Figure 6A). This pattern of expression is comparable to observations for chymotrypsin enzyme activity in *E. coioides* larvae, where activity began to increase from 30 dph, then increased dramatically at 55 dph (Eusebio et al. 2004). In other species, such as Asian seabass (*Lates calcarifer*), mRNA levels approximately mirror enzyme activity, and levels tend to decrease after attainment of stomach functionality (Srichanun et al. 2013). Thus, it appears as though digestive capacity is distinctly different in grouper species, with chymotrypsin contributing more to digestion as development progresses. Similarly, chymotrypsin-like elastase family member 3B (Cel3b) expression was highest at 50 dph (Figure 6A). While data concerning Cel3b is scarce, its abundance is sensitive to changes in type of dietary oil in *C. idellus* (Tian et al. 2015), has been linked to changes in digestive function in Eurasian perch (*Perca fluviatilis*) (Chen et al. 2017), and is expressed at high levels in Japanese eel (*Anguilla japonica*) larvae (Hsu et al. 2015).

Pepsinogens are mainly synthesised and secreted by gastric cells in the stomach, and their presence and activity is often used as a marker for both attainment of stomach function, and the transition from larval to juvenile stage (Eusebio et al. 2004; Srichanun et al. 2013). Pepsinogens are important for the breakdown of complex proteins (Gildberg 2004), and it has even been suggested that their role in determining growth rate in *E. coioides* exceeds that of Try (Liu et al. 2012). In the present study, the expression of pepsinogen C and A1 was undetectable at 12 and 18 dph, and was elevated at 50 dph in body tissue (Figure 6A). Furthermore, the expression of pepsinogen A2 followed the same expression pattern, except that low levels of the transcript were detected at 12 and 18 dph relative to 50 dph. This is consistent with observations for the grouper *E. septemfasciatus* where pepsinogen A expression was first detected at 38 dph (Wu et al. 2011), and *E. coioides* (Feng et al. 2008a), where pepsinogen C transcripts were present from 41 dph, and increased significantly by 52 dph. Similarly, in *E. coioides* pepsinogen A1 was expressed from 29 dph, however A2 was consistently detected from hatching through to adult life stages, most likely due to A2’s expression mechanism.
in tissues other than the stomach (Feng et al. 2008b). Thus, the pepsinogen A2 expression detected at 12 and 18 dph in the current study is likely due to the contribution of non-stomach tissues, such as the mid-intestine (Feng et al. 2008b). This may indicate that the pepsinogen A1 and A2 have differing physiological roles in grouper development as suggested by Feng et al. (2008b), and the importance and function of early A2 expression remains to be explored.

Digestive capacity in the context of protein degradation in the stomach is attained through the actions of pepsin and the effects of low pH (Rønnestad et al. 2013). The protein encoded by the ‘proton pump’ gene is essential for the production of hydrochloric acid and its secretion from the gastric glands, which is essential for the conversion of pepsinogen to the active form, pepsin. In winter flounder, *Pleuronectes americanus*, the expression of pepsinogen and (α and β subunits of) proton pump transcripts coincided with formation of gastric glands and an increase in pepsin enzyme activity (Douglas et al. 1999). In OG larvae, a sharp increase in the expression of proton pump β was observed at 50 dph relative to earlier time points (Figure 6A), and while expression of pump α tended to increase (>2 fold, between 12 and 50 dph) over time, changes were not significant (data not shown). Thus, the expression of proton pump subunits may be coupled to some extent with the expression of pepsinogen A1 and C, and may also be a useful marker for attainment of stomach functionality in OG.

Enzymes associated with the intestinal brush border membrane, such as aminopeptidase and alkaline phosphatase, and pancreatic carboxypeptidases, break down the partially digested products of primary proteases, cleaving them to smaller peptides and free amino acids (FAA) that can then be absorbed (Rønnestad et al. 2013; Rungruangsk Torrisen & Male 2000). For aminopeptidase N, alkaline phosphatase, and carboxypeptidase A and B, expression was significantly higher at 50 dph relative to 12 and 18 dph (Figure 6A), likely indicating an associated increase in digestive capacity at the end of metamorphic development. However, an increase in alkaline phosphatase transcript level was evident between 12 and 18 dph, suggesting that the capacity to absorb small peptides and FAAs increases steadily throughout development. Similarly, aminopeptidase and alkaline phosphatase specific activities increased during larval development in the leopard grouper *Mycteroperca rosacea*, and the authors suggested that this, in conjunction with a decrease in leucine–alanine peptidase activity, represented a switch from intracellular digestion to digestion by brush border enzymes in a mature digestive system (Martínez-Lagos et al. 2014).

### 3.6 Carbohydrate digestion

The gastric glands of the stomach produce multiple forms of chitinase (Chi), which are thought to aid in the digestion of crustaceans (including copepods) by hydrolysing chitin, the main constituent of the exoskeleton (Kurokawa et al. 2004). In the current study, *chi1* and *chi2* were undetectable in body tissue at 12 and 18 dph, then expression was significantly elevated at 50 dph (Figure 6B). While the expression of *chi3* was also significantly higher at 50 dph relative to 12 and 18 dph (Figure 6B), transcripts were detected at the earlier time points, before the timing of gastric gland formation. Similarly, *chi1* and *chi2* were expressed with pepsinogen from the time of gastric gland development in Japanese flounder (*Paralichthys olivaceus*) larvae, and *chi3* was expressed from an earlier stage (Kurokawa et al. 2004). Furthermore, *chi3* was expressed in multiple tissues in adult flounder, and the authors of that study suggested that *chi1* and *chi2* primarily digest dietary chitin, while *chi3* may be a macrophage specific chitinase involved in biodefence. While further investigation into the function and site of expression for chitinases is required, these preliminary results suggest that some similarities can be drawn between OG and other species.
Pancreatic α-Amy is crucial for complex carbohydrate digestion in larvae, and can be detected early in development from the yolk sac stage (Rønnestad et al. 2013). In the current study, a transcript for amylase was DE in body tissue, with expression decreasing from 12 dph then remaining stable (Figure 6B). However, upon closer inspection, this transcript had one significant match (e-value 3e-92) to amylase 2 from the rotifer Brachionus plicatilis, an important species in the context of global aquaculture, and the rotifer used as live feed in the present study. Thus, live feed is the likely source of this transcript, with the number of ingested rotifers, and therefore expression levels, dropping as larvae (presumably) began to feed preferentially on copepods from 12 dph. In a study on haddock (Melanogrammus aeglefinus) and G. morhua larvae, α-Amy activity was only detected when large numbers of live prey were present in the gut (Perez-Casanova et al. 2006). The authors of that study estimated suggest that the contribution of live prey to α-Amy activity was 100%. As such, the potential contribution of endogenous versus live prey α-Amy to OG larval digestion is a suggested area for future research. Furthermore, expression of OG’s endogenous α-amy was maximal at 18 dph, though differences in expression over time were not statistically significant. While specific (when measured) and total enzyme activities do not mirror each other during larval development, maximal activity is typically observed after metamorphic development in OG and leopard grouper (Eusebio et al. 2004; Martínez-Lagos et al. 2014). Thus, gene expression levels and enzyme activities are not necessarily grouped in OG.

Lactase is traditionally thought of as a mammal-specific enzyme, although its activity has been detected in a range of fish species (Gawlicka et al. 1995; Krogdahl et al. 2004; Prasad & Suneesha 2013). Lactase activity is affected by dietary carbohydrate levels in some species (Krogdahl et al. 2004), and increases significantly in the intestine of larval white sturgeon, Acipecerus transmontanus, from the 24th day of feeding (Gawlicka et al. 1995). In the current study, body-expressed lactase-phlorizin hydrolase (lactase) expression was significantly higher at 50 dph relative to earlier time points (Figure 6B). While it has been suggested that the presence of lactase is important for the digestion of lactose-like materials present in diatoms and other algae consumed by teleosts (Olatunde et al. 1991), its importance in carnivorous OG remains to be elucidated. Similarly, maltase-glucosamylase expression was 2.1 fold higher at 50 than 12 dph, though this difference was not significant (data not shown). Maltase is commonly used as a marker for maturation of intestinal cells (Zambonino Infante & Cahu 2001), and it has been suggested that maltase is more important in herbivorous than carnivorous species (Gioda et al. 2017) such as grouper. However, maltase activity steadily increases from ~20 dph in the grouper M. rosacea (Martínez-Lagos et al. 2014), suggesting that this enzyme may be important for grouper larvae fed live prey, and digestive capacity in the context of maltase may be determined at the post-transcriptional level.

### 3.7 Lipid digestion

Lipids are an important source of metabolic energy throughout the life cycle of marine fish (Izquierdo et al. 2000), and Bal (also referred to as carboxylester lipase or cholesterol esterase) is an important pancreatic neutral lipase that is secreted into the intestine in teleosts (Sæle et al. 2010). In the present study, the expression of bal decreased from 12 to 50 dph in body tissue (Figure 6C), a similar pattern to what was observed for L. calcarifer, where bal expression declined from 15 dph (Srichanun et al. 2013), and in G. morhua where expression peaked at 25 dph, then decreased as metamorphosis approached (Kortner et al. 2011). In contrast, general lipase activity increased from 25 dph in E. coioides, and from 30 dph in M rosacea (Martínez-Lagos et al. 2014; Quinitio et al. 2004b), though this may be explained by the expression of phospholipase A2 (Pla2, see below). For M. rosacea, it was suggested that diet change and the higher lipid content of food during weaning at
30 dph contributed to this trend (Martínez-Lagos et al. 2014). It is also likely related to the stage of development (Martínez-Lagos et al. 2014), and the increased production of bile salts which is essential for enzyme activity (Gjellesvik et al. 1992). Furthermore, Sæle et al. (2010) reported a somewhat uncoupled relationship between the expression of bal and neutral lipase activity in G. morhua, and suggested this may be due to the un especific nature of enzyme activity testing, which is also highlighted in a review by Rønnestad et al. (2013). To our knowledge, this is the first report on the expression levels of bal during early development in a grouper species.

While Bal is a multi-substrate lipase that is able to hydrolyse phospholipids, it has been suggested that pancreatic Pla2 is even more important in the context of phospholipid digestion (Rønnestad et al. 2013; Zambonino Infante & Cahu 2007). In the present study, the expression of secretory pla2 increased slightly from 12 to 18 dph in body tissue, with a significant increase in transcript level by 50 dph (Figure 6C). This is similar to the findings of Kortner et al. (2011), who showed that in G. morhua larvae, pla2 levels increased sharply from 20 to 60 dph, and hypothesised that Pla2 becomes increasingly important as development progresses. In fact, Pla2 has been singled out as a key player during epithelial restructuring of digestive tissues during organogenesis (Neas & Hazel 1985; Ozkizilcik et al. 1996). Interestingly, the expression profile of bal in that study was non-parallel with that of pla2, which is similar to what was observed in the current study as the two mRNAs had inverse expression patterns. In addition, the release of Pla2 can be stimulated by an increase in the level of its substrate (Zambonino Infante & Cahu 2001), and therefore may have been influenced by the introduction of artemia nauplii at 35 dph. Thus, transition to the juvenile stage in OG may be supported by a shift towards the production of Pla2, although data on the activity of the enzyme itself throughout development would help to shed light on its importance during metamorphosis.

In mammals, pancreatic lipase (Pl) is the most important enzyme for the digestion of triglycerides, and is dependent on the cofactor colipase (Lowe 2002). In contrast, Bal, which is dependent on bile salts, can hydrolyse polyunsaturated fatty acids from triacylglycerol better than Pl (Chen et al. 1990), and may be better suited than Pl for lipid digestion in teleosts (Sæle et al. 2010). In fact, the importance of Pl in teleosts has been the subject of much debate (Kurtovic et al. 2009), especially since lipases purified from fish seem to require bile salts for activation (Tocher 2003), and feeding and fasting did not alter pl levels in the hepatopancreas of red sea bream (Pagrus major) (Oku et al. 2006). In the present study, the expression of a body-expressed isoform with the annotation ‘colipase-dependent pancreatic lipase’ was significantly higher at 50 dph relative to earlier time points (Figure 6C). The top three hits to the BLASTN database for this transcript were colipase-dependent pancreatic lipase from E. coioides (EU683731) and Sparus aurata (JX975717), and pancreatic lipase from P. major (AB252856). Interestingly, a transcript corresponding to colipase was not found in the OG transcriptome. Similarly, colipase was not found in the genomes of G. morhua, Japanese rice fish (Oryzias latipes), three-spined stickleback (Gasterosteus aculeatus), or zebrafish (Danio rerio) (Sæle et al. 2010), although evidence of colipase activity has been reported for several species (Kurtovic et al. 2009). Due to ambiguity in the literature and the potential for species-specific differences, it is difficult to predict the physiological role of this transcript, though due to its differential expression during early development in OG, a more in depth investigation is warranted.

Acyl-CoA dehydrogenase (Acadm) catalyses the first step in mitochondrial fatty acid β-oxidation, and is therefore important for the metabolism of dietary fatty acids. Expression of acadm was significantly higher at 50 dph relative to earlier time points, which is broadly consistent with published work in G. morhua, where expression increased throughout larval development (Kortner et al. 2011). The increased expression of this transcript at 50 dph is likely a reflection of an increased...
ability to acquire energy for metabolism and growth, as suggested by Kortner et al, (2011) and in agreement with the increased digestive capacity of post-metamorphic juveniles.

3.8 Conclusion

In the current study, the sequences of many transcripts involved in the regulation of appetite, feeding and digestive processes were detected at 12, 18, and 50 dph in OG larvae. Analysis of head tissue revealed many transcripts that form part of the complex regulatory network controlling appetite and feeding. In body tissue, the dramatic increase in expression transcripts encoding digestive enzymes at 50 dph is broadly consistent with changes in digestive enzyme activity during larval development in other grouper species, where the greatest variations were observed at the onset, or end of metamorphic development. In addition, we have provided evidence suggesting that various trypsinogen isoforms are present, and have differing expression patterns throughout larval development, and report on the presence of a prey-specific transcript encoding α-amylase. Taken together, these results give insight into the processes underpinning attainment of digestive capacity, and form the basis of a new transcriptomic database that will aid further study into the digestive development and dietary requirements of orange-spotted grouper larvae.

Acknowledgements

This study was funded by an Australian Centre for International Agricultural Research grant (FIS-2012-101) awarded to Prof. Abigail Elizur at University of the Sunshine Coast. The study was also supported by the Center of Excellence for the Oceans, National Taiwan Ocean University (Prof. Ming-Wei Lu), at Department of Aquaculture, National Taiwan Ocean University, and partial support was provided by the Ministry of Science and Technology, Taiwan (MOST 104-2321-B-019-004-MY3; MOST 106-3114-B-019-001; MOST 106-3114-B-020-001) to MW Lu.

References


Krogdahl Å, Sundby A, and Olli JJ. 2004. Atlantic salmon (Salmo salar) and rainbow trout (Oncorhynchus mykiss) digest and metabolize nutrients differently. Effects of water salinity and dietary starch level. Aquaculture 229:335-360.


Maitra SK, and Hasan KN. 2016. The role of melatonin as a hormone and an antioxidant in the control of fish reproduction. *Frontiers in endocrinology* 7.


Quinitio G, Sa-an AC, Toledo JD, and Tan-Fermin JD. 2004b. Localisation of enzymes in the digestive system during early development of the grouper (Epinephelus coioides). ADVANCES IN GROUPER AQUACULTURE:30-34.


Spedicato MT, and Boglione C. 2000. Main constraints in the artificial propogation of the dusky grouper Epinephelus marginatus (Lowe, 1834): Three years of experimental trials on induced spawning and larval rearing. CIHEAM: Options Méditerranéennes 47:227-234.


Abbreviations

TMM - Trimmed Mean of M-values
qPCR – quantitative PCR
FDR – false discovery rate
FPKM - Fragments Per Kilobase of transcript per Million mapped reads
GSP – Gene Specific Primer
Nucb2 - nucleobindin 2
Lepb – Leptin b
Calcr - calcitonin receptor
Try2 - trypsinogen 2
Bal - bile salt-activated lipase 2
α-Amy - α-amylase
Try3 - trypsinogen 3
Try1 – trypsinogen 1
MepAβ - meprin A β
ApoA4 - apolipoprotein A-IV
β-actin – beta actin
Ef1α - elongation factor 1-α
Arfgef1 - ADP ribosylation factor guanine nucleotide exchange factor 1
Nav3 - neuron navigator 3
B2m - beta-2 microglobulin
Bp – base pairs
Try – trypsinogen
DE – differentially expressed
dph – days post hatch
GO – Gene Ontology
OG – Orange grouper
Npy - neuropeptide Y
Agrp - agouti-related protein
Cart - cocaine- and amphetamine-regulated transcript
Ox - orexin
Gal - Galanin
Crh - corticotropin-releasing hormone
GIT - Gastrointestinal tract
Cck – cholecystokinin
Cel3b - chymotrypsin-like elastase family member 3B
Chi - Chitinase
Pla2 - phospholipase A2
Pl – pancreatic lipase
Acadm - Acyl-CoA dehydrogenase
Py/Pyyb - pancreatic peptide Y