Transcriptome analysis of differentially expressed genes in the fore- and hind-intestine of ovate pompano *Trachinotus ovatus*

Ke-Tao Lin\textsuperscript{a,b}, Wen-Xiong Wang\textsuperscript{b,c}, Hui-Ting Ruan\textsuperscript{a,b}, Jia-Ge Dai\textsuperscript{a,b}, Ji-Jia Sun\textsuperscript{a,b}, Li Liu\textsuperscript{a,b,⁎}, Xian-De Huang\textsuperscript{a,b,⁎}

\textsuperscript{a} College of Marine Sciences, South China Agricultural University, 483 Wushan Road, Guangzhou 510642, China
\textsuperscript{b} Joint Laboratory of Guangdong Province and Hong Kong Region on Marine Bioresource Conservation and Exploitation, College of Marine Sciences, South China Agricultural University, Guangzhou 510642, China
\textsuperscript{c} Department of Ocean Science, Hong Kong University of Science and Technology, Clearwater Bay, Kowloon, Hong Kong

**A R T I C L E   I N F O**

**Keywords:**
Ovate pompano
Intestine
Transcriptome
Differentially expressed genes

**A B S T R A C T**

The fish intestine is a highly complex organ and plays an important role in fish immunity and nutrition. To characterize the molecular functions of the intestine of ovate pompano, we compared the gene expression differences and explored the molecular functions of the fore- and hind-intestine (FI and HI) using transcriptomic analysis. Our analysis yielded a total of approximately 21 Gb (gigabytes) of clean bases and 144 M of clean reads (mean of 24 M for each sample). Most of the 61,572 genes in the FI and HI were not differentially expressed, and only 105 genes were identified as the differentially expressed genes (DEGs). Among the 61,572 genes, the genes that exhibited the highest expression levels were mostly related to various digestive enzymes. Among the 105 DEGs, 86 were annotated as immune- and absorption-related genes in ovate pompano. Detailed analysis of these DEGs in FI and HI indicated that the FI required more abundant genes to handle the attacks and digestion, whereas the HI recruited more genes to deal with the absorption, suggesting that FI focused mainly on immunity and digestion, and the HI focused on absorption. Our results provided an insight on the molecular mechanisms of the function of intestine in ovate pompano.

1. Introduction

The intestine is a multifunctional organ critical for nutrient absorption, pathogen recognition, and intestinal microbiome regulation. Many fish diseases in the farming sector are caused by intestinal damage, which leads to serious economic loss (Groschwitz and Hogan, 2009). Intestinal health and normal function are thus critical for fish health for several reasons. First, animals farmed in aquaculture are susceptible to intestinal infections due to water environment and high stocking densities common in many farming practices. Second, intestinal health is usually manageable to some degrees by adding various additives and drugs to commercial pellet feeds. Finally, the intestinal micro-ecosystem allows for microbial colonization by symbionts and responds to dietary manipulations (Martin et al., 2016).

Recent advances in high-throughput technologies such as RNA sequencing (RNA-seq) have revolutionized the study of fish intestine at the level of whole transcriptome, which benefits both model and non-model species of fish at ever-decreasing costs. It also provides a powerful tool for exploring the complex molecular mechanisms of intestinal health (Li et al., 2012; Xia et al., 2013; Qian et al., 2014). To date, the transcriptomic studies of fish intestine mainly focus on the following four areas: 1) intestinal immune function and responses to disease challenge, such as parasitic infection, viral response, bacterial response, immunostimulants, and vaccination; 2) effects of alternative plant protein on fish intestinal health; 3) role of intestine in osmoregulatory function and responses to environmental stress; and 4) complex interactions between fish intestine and its microbial inhabitants (Martin et al., 2016). Furthermore, intestinal immunity and nutrition are closely related to feeding habits and intestinal microbiota (Buddington et al., 1997; Wang et al., 2018), and different intestinal segments have different functions in fish (Rombout et al., 2011). For example, most of the macromolecules and foreign antigens were absorbed in fish hind intestine (Lokka et al., 2014; Brugman, 2016), where about 95–99% of short-chain fatty acids (SCFA) were rapidly absorbed (Titus and Ahearn, 1988). Oral vaccines could be degraded in the fish foregut (Rombout et al., 2011; Lovmo et al., 2017). In European sea bass, different intestinal segments were compared using high-quality sequencing to generate a precise functional map of the intestine (Calduch-
The ovate pompano, *Trachinotus ovatus*, a carnivorous fish, belongs to the genus *Trachinotus* (Perciformes: Carangidae) and is one of the most economically important warm-water cultured marine fish in the world due to its fast growth, good flesh quality and suitability for cage culture. In recent years, ovate pompano farming has been developed rapidly and widely along the southern coast of China, including Guangdong, Guangxi, Fujian and Hainan Provinces. Previous studies on ovate pompano mainly focused on nutrition and disease (Zhang et al., 2014; Yu et al., 2016; Tan et al., 2017). In a recent study, a transcriptome analysis was performed on various tissues in *T. ovatus*, which provided a valuable genomic resource for novel gene discovery, gene expression and regulation studies, as well as the identification of genetic markers (Xie et al., 2014). In spite of its economic importance, very few studies examined the intestine of ovate pompano. The elucidation of intestinal molecular function such as difference in different parts of intestine could help to better understand fish intestinal immunity and nutrition.

In the present study, we therefore compared the gene expression differences and explored the molecular functions in the fore- and hind-intestine (FI and HI) (three replicates each) using transcriptome analysis to characterize the intestinal molecular functions in the ovate pompano, which may provide insight into the intestinal molecular biology of the ovate pompano and other fish.

2. Materials and methods

2.1. Sample collection

Nine ovate pompano (body weight: 500 ± 50 g) were purchased from the Wushan farmer’s market (Guangzhou, Guangdong, China). After fish were randomly divided into three groups and sacrificed, the intestinal contents were scraped off with tweezers and discarded, and the fore- and hind-intestinal segments were removed by surgical scissors and rinsed in sterile 1 × phosphate-buffered saline solution. For each replicate, the FI segments were collected from three fish as mentioned above and pooled as one sample, and the HI segments were also prepared using similar method. Finally, a total of six samples (three FI samples and three HI samples) were frozen immediately in liquid nitrogen and then stored at −80 °C. All procedures about the handling of experimental fish met the China Law for Animal Health Protection and Instructions (Ethics approval No. SCXK (YU2005-0001)).

2.2. Library construction and sequencing

Total RNA was extracted with a Trizol RNA extraction kit (Invitrogen, USA) following the manufacturer’s instructions, and eukaryotic mRNA was enriched by Oligo (dT) beads. Then, the enriched mRNA was fragmented and reverse transcribed into cDNA with random primers. After the second-strand cDNA was synthesized by DNA polymerase I, RNase H, dNTP and buffer, purified with QiaQuick PCR extraction kit (Qiagen, GER), and end repaired, then the poly(A) was added and ligated to Illumina sequencing adapters. Following size selection by agarose gel electrophoresis, the ligation products were amplified by PCR and sequenced using Illumina HiSeqTM 4000 by Gene Denovo Biotechnology Co. (Guangzhou, China).

2.3. De novo assembly

To acquire high-quality clean reads, reads were further filtered to remove low quality reads with > 50% of low quality (Q-value ≤ 10) bases and reads containing adapters or > 10% of unknown nucleotides. Transcriptome de novo assembly was carried out using Trinity, the short-reads assembling program (Haas et al., 2013) with default parameters. First, Inchworm assembled reads by a greedy k-mer based approach to collect linear contigs. Next, Chrysalis clusters above related contigs and builds a de Bruijn graphs. Then, Butterfly analyzes the paths taken by reads and read pairings based on de Bruijn graph, and outputs alternatively spliced isoform and transcripts derived from paralogous genes. Finally, the longest transcripts were combined together and assembled into unigenes. Meanwhile, BUSCO (Benchmarking Universal Single-Copy Orthologs) (version 3.0.1) was used to evaluate the quality of transcriptomic assembly with a single copy of an orthologous gene from vertebrata_odb9.

2.4. Unigene expression analysis and annotation

The unigene expression was calculated and normalized to the reads per kb per million reads (RPKM) method (Mortazavi et al., 2008), which eliminated the bias of different gene lengths and abundance on the calculation of gene expression. RPKM was calculated as: RPKM = (1,000,000 × C)/(N × L/1000), where C is the number of reads that are uniquely mapped to one unigene, L is the length (base) of the unigene, and N is the total number of reads that are uniquely mapped to all unigenes.

Unigenes were annotated by the BLASTx program (http://www.ncbi.nlm.nih.gov/BLAST/) with an E-value threshold of 10^{−5} to the NCBI non-redundant protein (Nr) database (http://www.ncbi.nlm.nih.gov), the Swiss-Prot protein database (http://www.expasy.ch/sprot), Gene Ontology (GO), and the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database (http://www.genome.jp/kegg). Protein functional annotations were based on the best alignment results (lowest E-value), which were chosen to determine the sequence direction of genes. ESTScan program (Iseli et al., 1999) was used to determine the sequence direction and predict the coding regions. Gene Ontology (GO) annotation of unigenes was analyzed by Blast2GO software (Conesa et al., 2005), then functional classification of unigenes was performed using WEGO software (Ye et al., 2006; Ye et al., 2018). The Kyoto Encyclopaedia of Genes and Genomes (KEGG) annotation of unigenes was used by KEGG Automatic Annotation Server (http://www.genome.jp/kaas/).

2.5. Analysis of differentially expressed genes (DEGs)

To remove the rRNA mapped reads, short reads alignment tool Bowtie2 (2.2.8) (Langmead and Salzberg, 2012) was used for mapping reads to ribosome RNA (rRNA) database, and the remaining reads were mapped to the reference transcriptome by default parameters. Then the mapping ratio (Unique Mapped Reads number + Multiple Mapped Reads number)/All Reads number) and the number of expressive gene were calculated.

After the gene expression levels of the three replicates were normalized and equalized by RPKM, analysis of differentially expressed genes (DEGs) across the FI and HI was performed using the DESeq R package (Wang et al., 2010a) and adjusted using q-value (< 0.01). To identify DEGs across samples or groups, the edgeR package (http://www.r-project.org/) was used.

Genes were regarded as significantly differentially expressed when a fold change > 2 (|log2FC| > 1, namely |log2(Hi-gene_RPKM/FI-gene_RPKM)| > 1) and a false discovery rate (FDR) < 0.05. To determine the classifications of DEGs, enrichment analysis of GO functions and KEGG pathways were performed. After mapping the DEGs to GO terms or pathways, gene numbers for every term or pathway were calculated.

3. Results

3.1. De novo transcriptome assembly and quality evaluation

Six samples (three fore-intestine and three for hind-intestine samples) were sequenced using the Illumina HiSeq 4000 platform, which yielded a total of approximately 21 Gb (gigabytes) of clean bases and
approximately 144 M clean reads (mean of 24 M for each sample). The detailed statistics of each sample are listed in Table S1, which includes the total bases, Q20, Q30, read numbers, GC percentage, adapter number, and low-quality number before and after filtration. The unique mapped ratio (unique mapped reads/total reads) was approximately 87.7%. Raw read data were uploaded to the NCBI Sequence Read Archive (SRA) under the accession number SRP150471.

After using the Trinity software, de novo sequence assembly resulted in 61,575 unigenes (a total of 72,577,876 assembled bases) with 2491 bp N50 and an average length of 1178 ranging from 201 bps to 26,003 bps (Table 1). The unigene number and their length distribution indicated that 20,388 unigenes were > 1000 bp in length (Fig. 1). These results showed good quality of the assembled unigenes.

To further evaluate completeness of the unigenes, comparison of *T. ovatus* unigenes to the orthologs of closely related species (*Larimichthys crocea*, *Oreochromis niloticus* and *Stegastes partitus*) was performed. As shown in Table S2 and Fig. S1, there were > 10,000 unigenes with the ratio ≥0.9 and coverage > 80%, and most of *T. ovatus* unigenes could effectively cover the orthologs of closely related species. The ratio close to 1 or the percentage close to 100 indicated the good completeness of the transcriptome assembly. Meanwhile, to check the possible contamination, the list of species found in the NR blast results was analyzed and counted as in Table S3. Furthermore, BUSCO was performed and the result as in Fig. S2 showed that among total 2586 BUSCOs, complete BUSCOs was 2237 (86.5%) (complete and single-copy BUSCOs: 2152 (83.2%), complete and duplicated BUSCOs: 85 (3.3%)), fragmented BUSCOs was 205 (7.9%) and missing BUSCOs was 144 (5.6%).

### 3.2. Statistics of expressed genes

A total of 61,575 unigenes were obtained in the intestine. Among the 61,572 genes, 49,855, 45,912 and 54,971 genes were obtained from the three FI samples and represented 59,837 genes, and 53,203, 52,714 and 54,773 genes were obtained from the three HI samples and represented 60,124 genes. Of the 61,572 genes, 31,493 genes were not annotated, and 30,079 genes were annotated using the Nr databases. The detailed statistics are listed in Tables 1 and 2.

Furthermore, no difference in expression was observed for most of the 61,572 genes in the FI and HI. However, a significant difference was observed for 105 genes as indicated by the RPKM value (FI_rpkm and HI_rpkm values) (Table S4 and S5). Among the 30,079 sample genes annotated by the Nr Databases, 12,961 FI and 12,347 HI genes had RPKM values that were < 1, and 746 FI and 767 HI had RPKM values that were 0, which were represented by 0.001 to calculate the gene expression. The median RPKM values were 1.58 and 1.81 for the FI and HI, respectively, and the mean RPKM values were 18.4 and 16.96 for the FI and HI, respectively. Among the 31,493 genes not annotated by the Nr Databases, RPKM values were < 1 for 25,610 FI and 23,928 HI genes and 0 (0.001) for 989 FI and 681 HI genes. The median values were 0.43 and 0.56, and the mean values were 1.51 and 1.84 for FI and HI genes, respectively. The median and mean values for the annotated genes were about 4 times and 10 times higher than the non-annotated genes, respectively.

Furthermore, for 105 genes that were differentially expressed in the FI and HI, 86 genes were annotated, and 19 genes were not annotated by the Nr Databases (Fig. 2). In addition, 58 genes were significantly up-regulated, and 47 genes were down-regulated. For the annotated 86 genes, 46 were significantly up-regulated, and 40 were down-regulated. For the non-annotated 19 genes, 12 were significantly up-regulated,
Considering the lack of genomic and related ncRNA data for annotated DEGs were further clustered and represented as a heatmap threshold. To gain a global view of the DEGs in the FI and HI, 61 of 86 values. To make the results more reliable, these DEGs with the lowly expression may be caused by error due to the lower original FI and HI RPKM values determined using the GO functions and KEGG pathways. The GO classification of DEGs in the FI and HI was determined and shown in Fig. 3. 107 and 68 DEGs were grouped into biological process category, 80 and 29 DEGs into the molecular function category, and the 40 and 36 into the cellular component category. Most of the up-regulated GO-terms were greater than the down-regulated GO-terms. In sequential order, the top five up-regulated GO-terms were ‘binding’, ‘cell part’, ‘cell’, ‘biological regulation’, and ‘single-organism process’. To determine the most affected pathways in the FI compared to the HI, 21 DEGs were mapped to 24 KEGG pathways and listed in Table S6. Then, KEGG enrichment analysis (corrected Rich Factor > 0.05) was performed, and ‘neuroactive ligand-receptor interaction’ (Rich Factor: 5/395) and ‘PPAR signaling pathway’ (Rich Factor: 2/99) were identified as the top two pathway enrichments. Other pathway terms were only one as the numerator and some of the DEGs.

3.2. Detailed DEGs in the fore- and hind-intestine

Despite the significant difference for some DEGs especially based on the formula of $|\log_2(\text{HI-gene_RPKM}/\text{FI-gene_RPKM})| > 1$, such difference may be caused by error due to the lower original FI and HI RPKM values. To make the results more reliable, these DEGs with the lowly original FI and HI RPKM values were further removed using the sum (3.39) of the FI and HI median RPKM values (1.58 and 1.81) as a threshold. To gain a global view of the DEGs in the FI and HI, 61 of 86 annotated DEGs were further clustered and represented as a heatmap (Fig. 4). Considering the lack of genomic and related ncRNA data for the ovate pompano, the non-annotated 19 genes are listed in Table S5 but not shown in the heatmap.

Based on Fig. 4 and Table S5, the down-regulated DEGs were classified into three categories; 1) the immune effector factors, such as acidic mammalian chitinase (AMCase)-like (RPKM value: 648.32 in the FI Vs 0.85 in the HI), natterin-3-like, natterin-3-like isoform X1, Lectin, and eosinophil peroxidase-like, 2) the digestive and absorption related genes in the intestine, such as trypsinogen-like protein 3, alkaline phosphatase-like, meprin A subunit beta-like, peptide YYa, ammonium transporter Rh type B-like, aquaporin-10-like, carboxypeptidase O-like, and insulin-like peptide INSLS; and 3) other genes, such as sodium-dependent phosphate transport protein 2B, and fetuin-B-like. Similarly, the up-regulated DEGs also fell into three categories: 1) the immune gene collectrin 2) the digestive and absorption related genes, including ileal fatty acid-binding protein (fFABP) (RPKM value: 6.17 in the FI Vs 4430.41 in the HI), inosine-uridine preferring nucleoside hydrolase-like, NAPDH oxidase organizer 1-like, and cathepsin B; and 3) other genes, such as asialal sodium/bile acid cotransporter and telomerase reverse transcriptase. The details are listed in Table S5. These results indicated that the FI required more abundant genes to handle the attacks (the immune related effector factors) and digestion (the digestive related genes), and the HI recruited more genes to deal with the absorption (the absorption related genes), suggesting that FI focused on immunity and digestion, and the HI concentrated on absorption in the ovate.

In addition, many genes were not differentially expressed at significant levels between the FI and HI, but they displayed the high expression level (RPKM value) ranging from thousands to tens of thousands both in the FI and HI, and mainly annotated as various enzymes such as tryptase-2 and tryptase-2-like, chymotrypsin A- and B-like, chymotrypsin-like elastase family member 2A, elastase-1-like, pancreatic elastase, carboxypeptidase B, ubiquitin-like, elongation factor 1 alpha, trypsin-3, 40S ribosomal protein S2, cytochrome c oxidase subunit III (mitochondrion), apolipoprotein A-I, collagen alpha-1(I) chain-like, actin (aortic smooth muscle), and selenoprotein P isoform X1 (Table S4). To some degrees, the high expression levels of these genes (most of which were protease enzymes) indicated that they had multiple functions, such as functions related to digestion and absorption in both the FI and HI.

4. Discussion

The intestine is a highly complex organ that plays a key role in fish immunity and nutrition. In recent years, intestinal microbiome also attracted increasingly attention. Exploring the intestinal molecular mechanisms is necessary to better understand the intestine’s multiple functions, including interaction of the intestine and its microbiome. Several high-throughput profiling (transcriptome) studies have been performed to study the intestinal responses to various situations, such as disease challenge, feeds, environmental stress and developmental factors (Martin et al., 2016). In this study, we focused on the intestine of the ovate pompano, T. ovatus, and sequenced six transcriptomes of the FI and HI factors (Martin et al., 2016). In this study, we focused on the intestine of the ovate pompano, T. ovatus, and sequenced six transcriptomes of the FI and HI (each tissue was divided into three replicates), which yielded approximately 21 Gb of clean bases. Finally, a total number of 61,575 unigenes and 61,572 genes were obtained. Fewer than half (30079) of the genes were functionally annotated by various databases, such as Nr databases, and approximately half of genes (31493) did not completely match the known genes, which was reflected in the gene number in the
intestine of the ovate pompano. Some of the non-annotated genes may have been assembly artifacts, unknown genes, or long non-coding RNAs (lncRNAs) (Mattick, 2011) that required for further verification.

Non-annotated genes included unknown genes, error cluster sequences, and ncRNA such as lncRNA, the expression of which was associated with lower than normal mRNA (Derrien et al., 2012). Furthermore, there was almost no significant difference in expression between the FI and HI with only 105 DEGs among the 61,572 genes. The genes that exhibited the highest expression coexisted in the FI and HI with no significant difference and were mainly associated with various enzymes such as trypsin-2, trypsin-2-like, chymotrypsin A- and B-like, and chymotrypsin-like elastase family member 2A. The median and mean RPKM values of the FI and HI samples also confirmed this result.

Although intestinal gene expression under different conditions was reported previously (Martin et al., 2016), there are few DEGs studies on the different intestinal segments (Wang et al., 2010b; Calduch-Giner et al., 2016). The results of this study suggested that the functions of the DEGs in the FI were primarily related to immunity and digestion, and the functions of DEGs in the HI were primarily related to absorption in the ovate pompano. Compared to the HI, 45 annotated genes in the FI such as AMCase-like, natterin-3-like and trypsinogen-like protein 3, were down-regulated. Similar result was observed in the European sea bass (Dicentrarchus labrax) where an AMCase-like was identified (Calduch-Giner et al., 2016). Meanwhile, AMCase has been implicated in the induction of an interleukin-13 (IL-13)-mediated pathway via a T helper-2 (Th2)-specific protein and may be an important mediator of IL-13–induced responses. In addition, it may function as a major protease-resistant glycosidase to constitutively degrade the chitin substrates as part of the host’s defense against chitin-containing pathogens in the mammalian digestive system (Zhu et al., 2004; Ohno et al., 2016).

Some genes were also up-regulated. For example, a sharp increase in the expression of $> 700\times$ was observed for iFABP. iFABP belongs to the members of FABP family, which binds long-chain fatty acids with high affinity and function as central regulators of whole-body metabolic control. They are involved in reversibly binding intracellular hydrophobic ligands and trafficking them throughout cellular compartments and specific proteins, and they directly regulate the cognate nuclear transcription factor activity (Storch and Corsico, 2008; Thumser et al., 2014). Additionally, there were also some DEGs, such as Hox and vitamin K, whose expressions were not detected in the FI and HI. To some degrees, these DEGs indicated that the gene levels were different in the FI and HI of the ovate pompano, which sheds light on the different functions and mechanisms of action in the intestine, and contributes to identify and study the key indicators of intestinal functions.

Furthermore, the intestinal microbiota had been revealed to

![Fig. 3. The GO classification of the DEGs in the fore- and hind-intestine.](image-url)
regulate multiple host metabolic pathways, and interact with intestinal immune responses during health and disease (Round and Mazmanian, 2009; Nicholson et al., 2012; Kamada et al., 2013). As the reciprocal mechanisms between the host and intestinal microbiota, studies of the intestinal microbiota would promote to further understand the function of DEGs in the intestinal different segments. Therefore, in our other works, 16S rRNA gene sequencing was also applied to investigate the microbial composition and distribution in the FI and HI of the ovate pompano (data not shown). Combined with the transcriptional and 16S rRNA sequencing studies, it may be better to identify key indicators of intestinal gene functions and vital bacterial population of the intestinal microbiota.

In conclusion, we obtained the fore- and hind-intestinal (three replicates each) transcriptomes of the ovate pompano and determined both the similarities and differences in fore- and hind-intestinal gene expression. Of the 61,572 genes in the FI and HI, no difference in expression was observed with the exception of 105 DEGs. The genes that demonstrated the highest expression levels were mostly related to various digestive enzymes. Among the 105 DEGs, 86 were annotated as genes relating to immunity and absorption in the ovate pompano. The results of Detailed DEGs indicated that the FI required more abundant genes to handle the attacks (the immune related effecter factors) and digestion (the digestive related genes), and the HI recruited more genes to deal with the absorption (the absorption related genes). Our results may provide insight on the intestinal molecular biology of the ovate pompano.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.aquaculture.2019.04.078.

Acknowledgments

We thank the anonymous reviewers for their comments on this work. This study was supported by the National Infrastructure of Fishery Germplasm Resources (2018DKA30470).

References

K.-T. Lin, et al.

Aquaculture 508 (2019) 76-82


