Full-length Transcriptome Sequences and Identification of Putative Genes for Carotenoids Accumulation in Polymorphic Noble Scallop *Chlamys Nobilis*

Hongkuan Zhang

Shantou University <https://orcid.org/0000-0001-8307-1565>

Helu Liu

Shantou University

Karsoon Tan

Shantou University

Hongyu Ma

Shantou University

Shengkang Li

Shantou University

Huaiping Zheng (✉ hpzheng@stu.edu.cn)

<https://orcid.org/0000-0001-6542-2448>

Research article

Keywords: PacBio sequencing, *Chlamys nobilis*, Carotenoids, DEGs, Key genes

Posted Date: August 3rd, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-34743/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Carotenoids are ubiquitous in marine bivalves and the accumulation mechanism has attracted widespread interest, but still remains unclear. The enrichment carotenoids noble scallop *Chlamys nobilis* provide a model animal to explain the molecular mechanism of carotenoids in marine bivalves. For a better understanding of transcriptional data and putative genes involved in carotenoids accumulation in noble scallop, PacBio sequencing and RNA-seq data were jointed to analysis the differentially expressed genes in different tissues.

Results: In the present study, we obtained 26,237 non-redundant transcripts, and total 9263 DEGs were identified among the five tissues, and 3361 were up-regulated and 4980 genes were down-regulated in golden scallop (rich in carotenoids). Examination of the GO terms indicated that the DEGs mainly related to metabolic process, binding, catalytic activity and cellular process. KEGG pathway analysis showed that DEGs mainly related to phagosome, protein processing in endoplasmic reticulum and lysosome. Carotenoids-related genes, including CD36, Retinoid X Receptor (RXR), glutathione S-transferase (GST) and low density lipoprotein (LDL) showed significantly difference between golden and brown scallops, and several genes function were also identified in the early developmental stages.

Conclusions: This study revealed that there are large transcriptome differences between golden scallops and brown scallops, these results provide fundamental resources of large-scale full-length transcripts in *C. nobilis* and contribute to the understanding of carotenoids accumulation mechanism in mollusks.

Background

Carotenoids are ubiquitous in the natural environment, which synthesized by plants, some bacteria and fungi [1]. Animals must acquire carotenoids through their diet, and then modified carotenoids to self-specific [2, 3]. In recent years, the advance in sequencing methods has deepened our understanding of the genetics process of carotenoids in animals, for example, the genes and pathways involved in carotenoids absorption, transportation and metabolism [3].

As the first identified gene participating in carotenoid absorption and transportation in animals, scavenger receptor class B (SRB) first recognize lipoproteins (transport carotenoids to tissues through plasma/hemolymph) and then facilitate the movement of carotenoids into cells [4, 5]. In domesticated silkworm *Bombyx mori*, the SRB homologous of mammalian, SCRB15 and Cameo2, are involved in the selective transport movement of β -carotene and lutein, respectively [6], these genes with sequence similarity may diverge from duplication event of a gene [5]. Steroidogenic acute regulatory (StAR)/ metastatic lymph node 64 (MLN64), belonging to the StAR domain family, are responsible for the transport of lutein [7, 8]. The β -carotene oxygenases involved in cleavage of carotenoids, are the best characterized enzymes, and two genes have been identified in mammals and birds, including β , β -carotene-15,15'-monooxygenase (BCMO)/ β , β -carotene-9,9'-oxygenase (BCDO) [9]. BCMO, responsible for cleaving β -carotene, plays crucial roles in visual system, and BCDO can directly influence coloration for its diverse substrates from carotenoids [10]. Due to the important functions of BCMO, the function loss of BCMO from mutation would result in hypercarotenemia in human [9]. Moreover, yellow skin of domesticated chickens has a significantly association with the variation in BCDO [11]. More importantly, a candidate ketolase gene encoded by a cytochrome P450 enzyme was identified as a key gene responsible for the red plumage and bill coloration in birds, which can convert carotenoids from yellow to red [12, 13].

In recent years, the single-molecule real-time (SMRT) sequencing which can sequence the full length cDNA molecules, provides us a better understanding the transcriptomes. Comparing with the transcripts assembled from second-generation sequencing (NGS) platforms, SMRT sequencing increase the accuracy of transcriptome, such as higher completeness and better accuracy, whereas, quantifying gene expression cannot be use the SMRT sequencing. In Barley (*Hordeum vulgare* L.), DEGs were identified in GN121 and GN42 which involved in phosphorus metabolism by combining NGS and SMRT sequencing [14], and the combination of PacBio Iso-Seq and RNA-seq was also using in *Populus Nanlin895* to identify differentially expressed transcripts during the developmental gradient [15].

The noble scallop *Chlamys nobilis* is a deliciously edible mollusk that displays conspicuous polymorphism in shell and tissues color, and has been commonly cultured in the southern coast of China since 1980s [16, 17]. In our previous study, we found a scavenge receptor gene termed *SRB-like-3* was only expressed in noble scallop with higher total carotenoids content (golden scallops) but was absent noble scallop with lower total carotenoids content (brown scallops), and the results from RNAi experiment of this gene provides convincing evidence that SRB-like-3 is involved in carotenoids deposition in hemolymph [18]. At present, lack genome and few transcriptome information of noble scallop impede the study of the molecular mechanisms underlying the carotenoids accumulation in this species. In this study, PacBio Iso-Seq and RNA-seq analysis were combined to uncover candidate genes for carotenoids accumulation in golden scallops.

Methods

Ethics Statement

The noble scallop used in the current study was collected from Nan'ao Marine Biology Station of Shantou University, located at Nan'ao island of Shantou, Guangdong, China (23.4786°E, 117.1132°N). All experiments were conducted under the guidelines of institution and nation.

Sample Collection and Preparation

Tissues, including the gonad, adductor, intestine, hemolymph and mantle were collected from five female adult golden scallops (enriched carotenoids) and five female adult brown scallops (less carotenoids). The intestine (GI, BI), adductor (GA, BA), mantle (GM, BM), ovary (GO, BO) and hemolymph (GH, BH) of each color scallop were pooled together for next generation sequencing and Pacbio sequencing. All collected samples were frozen in liquid nitrogen

immediately for further use. Trizol Reagent was used to isolate the total RNA. The purity of RNA was detected by NanoPhotometre spectrophotometer (Implen, CA, USA), its degradation and contamination was monitored on 1% agarose gels. And the RNA's concentration was measured by Qubit RNA Assay Kit in Qubit 2.0 Fluorometer (Life Technologies, CA, USA). The integrity of RNA was evaluated by the RNA Nano 6000 Assay Kit. All the RIN (RNA integrity number) values of samples were ranging from 7.8 to 8.5.

Library preparation and sequencing

Total RNA was used to purify the mRNA by poly (T) oligo-attached magnetic beads. Full-length cDNA library were synthesized using the SMART PCR cDNA Synthesis Kit (Clontech, CA, USA), and NGS cDNA libraries were produced using First Strand Master Mix and Second Strand Master Mix. And then PCR amplification was carried out to enrich the cDNA fragments. Agilent Bioanalyzer 2100 system was used to evaluate the cDNA libraries' quality. The Pacific Biosciences' real-time sequencer and Illumina HiSeq 2000 System were used to SMAT sequencing and NGS, respectively.

Quality filtering and error correction of PacBio long reads

Low quality SMRT sequencing reads were removed according to the following parameter: length < 50 bp and quality < 0.75. After cleaning, the high quality reads were processed into reads of inserts (ROIs): full passes ≥ 0 and quality > 0.8. The ROIs with 5' and 3' primer sequences and a poly(A) tail were identified as full-length transcripts, and among them, ROIs were considered as full-length non-chimeric (FLNC) for possessing poly(A) and 5'tail. Consensus isoforms were clustered and corrected by approaching clustering using SMRT analysis (v2.3.0), and full-length consensus sequences were refined using Quiver [19]. High-quality full-length transcripts were obtained from the full length transcripts with high quality which corrected by proovread 2.13.841 and removed by the CD-HIT [20].

Alternative splicing, simple sequence repeat (SSR) detection and prediction of lncRNA

To identify alternative splicing events, unigenes were directly to run BLAST, which accord with the following conditons were considered as alternative splicing events: 1) in the alignment, two High-scoring Segment Pair must have the same forward/reverse direction; 2) and at least 100 bases at the 5' and 3' ends within their complete open reading frames [21].

MISA (<http://pgrc.ipk-gatersleben.de/misa/>) was used to identify the SSRs of these transcripts. Four computational approaches were used to predict the lncRNAs in this study, including CPC, CNCI, CPAT and Pfam.

CDS detection and functional annotation of transcripts

The coding regions of transcripts were identified by TransDecoder (<https://github.com/TransDecoder/TransDecoder/releases>) [22]. And the function of these transcripts was annotated by eight public databases, including NR, Pfam, KOG, COG, eggNOG, Swiss-Prot, KEGG and GO.

Differential expression analysis and functional enrichment analysis

To analyze the different expression genes (DEGs), the gene expression levels were used by FPKM, and EBSeq R package were used to performed the differential expression analysis between two samples, the threshold was $FDR < 0.05$ and $|\log_2 \text{foldchange}| \geq 1$ for significantly differential expression.

The GOseq R packages was used to analyze the GO enrichment of DEGs, and to analyze the KEGG enrichment of DEGs, the KOBAS software was used in this study [23].

Function identified of the carotenoids related DEGs in the scallop early development

To identify the carotenoids related genes function, samples were collected from early developmental stages of golden scallops and brown scallops, including fertilized egg, trochophore larva (T-larva), D-shape larva (D-larva), umbo larvae (U-larva), secondary shell stage (S-stage) and the juvenile stage (J-stage), and the expression levels of these related carotenoids genes were determined by qRT-PCR. The expression level of different genes was analyzed by $2^{-\Delta\Delta Ct}$ method.

Results

PacBio sequencing and error correction of long reads

To obtain comprehensive transcriptome profiles of golden and brown scallops between the same tissue, 10 libraries of five tissues were constructed from golden and brown scallops. Totally, 53.97 Gb NGS clean reads were produced with an average of 5.4 Gb for each sample (Additional file 1: Table S1), the sequencing data of this study was deposited in the NCBI (SRA: SRP171083).

Five tissues from golden scallops were used in extracting RNA, and the library was constructed by equally RNA from each tissues, and then sequencing. Totally 679,907 polymerase reads were obtained, and their full passes ≥ 0 , consensus accuracy > 0.8, the average length was 2,677 bp, and with quality of 0.89, and 12 passes (Additional file 1: Table S2). These polymerase reads were filtered using the standard protocol of SMRT Analysis software suite, and 182,010 ROIs were obtained, which included 90,727 FLNC and 80,556 non-full-length reads (Table 1). Duo to the high error of SMRT sequencing reads, it is indispensable to perform error correction using high-quality NGS short reads. After errors correction, low quality and redundant transcripts were removed. Finally, 26,237 of non-redundant transcripts were obtained from scallops.

Functional annotation and enrichment analysis

The obtained transcripts combining PacBio and NGS data increase the accuracy and efficiency of functional gene prediction and annotation, especially lacking reference genome information. Functional annotation of the non-redundant transcripts was searched against the public databases using the BLAST, such as NR, GO, Swissprot and KEGG databases. In the GO database, 4,227 transcripts were annotated, and 6,424 in COG; 9,977 in KEGG; 13,320 in KOG; 15,759 in Pfam; 11,729 in Swiss-Prot; 17,251 in eggNOG; 20,961 in NR. Total of 21,030 transcripts were annotated at least one of the eight databases (Fig. 1a). Based on the NR database, homologous species of *C. nobilis* were predicted using sequence alignment. Approximately 91.11% of sequences were aligned to *Mizuhopecten yessoensis*, followed by *Crassostrea gigas* (1.16%) (Fig. 1b).

Alternative splicing analysis and SSR detection

Total 227 alternative splicing events were identified (Additional file 1: Table S3). Due to no reference genome is available for this species, hence the types of alternative splicing events cannot be identify. And total of 26,135 transcripts (78,152,940 bp) were subjected to SSR analysis, including 23,758 SSRs and 12,442 SSR-containing sequences and found that most of them were with mono-, di-, or tri-nucleotide repeats (Fig. 2; Additional file 1: Table S4). Considering the high quality of transcriptome sequences, the detected SSRs would be useful for marker-assisted breeding and genetic analysis in the *C. nobilis*.

Long non-coding RNAs (lncRNAs)

As an emerging hot topic in biology, lncRNAs has been found to be functional as crucial regulators in variety of biological processes. In the present study, lncRNA transcripts were predicted by four methods, including CPC, CNCI, CPAT and pfam protein structure domain analysis, totally identified 6,032 lncRNAs in noble scallop (Fig. 3, Additional file 1: Table S5).

Identification of differentially expressed genes (DEGs)

Based on the reads from RNA-seq, FPKM values were used to investigate the gene expression patterns of different tissues of *C. nobilis*. Thus, the comparison of gene expression between golden scallops and brown scallops was performed. DEGs were analyzed using the edgeR software. To identify significantly differential expression genes, $FDR < 0.05$ & $|\log_2 \text{fold change}| \geq 1$ was set as the criteria. DEGs of five tissues from golden scallops and brown scallops was list in Table 2. And a total of 9263 DEGs were identified among the five tissues, and the up-regulated and down-regulated genes number were showed in Additional file 1: Table S6, respectively. Then, Venn diagrams showed the number of genes uniquely expressed in each tissue or genes shared between one or more tissues (Fig. 4). GO and KEGG pathway enrichment analysis of DEGs in the same tissue were shown in Fig. 5 and Fig. 6. And the top 20 KEGG pathways in each tissues were listed in Additional file 1: Table S7.

Functional annotation of DEGs and candidate genes involved in carotenoids accumulation

Total of 9263 DEGs between golden scallops and brown scallops, 3361 were up-regulated and 4980 genes were down-regulated. Of the 9263 DEGs, 8422 were annotated at least one of the following: Nr (8406), Swissprot (4766), KEGG (3702), KOG (4460) and GO (1697).

Based on the GO enrichment analysis, total of 47 significantly GO terms were observed (corrected P -value < 0.05). The top five GO terms were macromolecular complex, protein complex, oxoacid metabolic process, organic acid metabolic process and lyase activity (Fig. 7). And these include a number of terms related to carotenoids, such as lipid transport and lysosome. Total of 229 KEGG pathways were identified (Additional file 1, Table S8). These include a number of terms related to carotenoids accumulation, such as lysosome, fat digestion and absorption and ABC transporters (Fig. 8).

To explore the genetic mechanisms of the golden scallops, the differentially expressed genes of five tissues were filtered respectively for those believed to be involved in carotenoids accumulation. Several genes, including CD36, ABC transporter G family member 5 (ABCG5), beta, beta-carotene 15,15-dioxygenase (BCMO1), glutathione S-transferase (GSTs), intestine-specific homeobox (ISX), very low density lipoprotein receptor (VLDLR), low-density lipoprotein receptor (LDLR), Craotene oxygenase related to carotenoids accumulation were shown in Fig. 6, CD36, GAST-theta and very low density lipoprotein receptor (VLDLR) genes were significantly higher expressed in the hemolymph of golden scallops than brown scallops; ABCG5 gene was highly expressed in the mantle of golden scallops; BCMO1 gene was highly expressed in the intestine of golden and brown scallops than other tissues; GST-Mu gene was significantly higher expressed in the gonad of golden and brown scallops than other tissues; Craotene oxygenase gene was significantly higher expressed in the intestine of golden and brown scallops than other tissues; low-density lipoprotein receptor (LDLR) has a significant differences in mantle and adductor muscle. Overall, our results of the high expression levels for these genes related to carotenoids accumulation are consistent with higher carotenoids content in golden scallops, suggesting that these genes play important roles in carotenoids accumulation.

Function identified of the carotenoids related DEGs in the scallop early development

To explore the carotenoids related DEGs function in the scallop early development, we selected six genes to analyze their spatial and temporal expression characteristics (Fig. 9). The results indicate that all these genes showed a strongly expression level in fertilized egg, except for the BCMO1; and these genes also showed a higher level in golden scallops than that of brown scallops at S-stage and J-stage, except for BCMO1 and GST-Pi. The expression level of BCMO1 showed a decreasing trend, and brown scallops had higher levels than golden scallops at most stages.

Discussion

Transcriptome data generated by PacBio sequencing improved the unigenes from *C. nobilis*, and after the NGS data correcting, the full-length transcripts in the present study has a higher level completeness and gene annotation, providing us a better understanding of the transcriptome. Previous studies have

shown that the data produced by Illumina HiSeq platform has good sequencing depth and coverage in transcriptome, however, unigenes assembled by short reads would lead to a higher mis-assembly rate and unreliable gene annotation [24, 25], and resulting in the inability to identify the genes associated with traits. To avoid such limitations in *C. nobilis*, a full-length transcriptome data from five tissues (intestine, adductor, mantle, ovary and hemolymph), was produced using the PacBio SMRT sequencing approach, which maximizes transcript diversity. As expected, a great amount of transcriptome data was generated, 26,237 non-redundant transcripts and 21,030 transcripts were annotated which were much better than reported previously [18].

Carotenoids are fat-soluble pigments, showing red, orange or yellow which are commonly displayed in animals, however, except for a few arthropods, most animals are mainly obtained carotenoids from diets [26, 27]. In recent years, carotenoids accumulation in bivalves was documented in *Argopecten irradians* [28], *Hyriopsis cumingii* [30] and *Patinopecten yessoensis* [30]. Previous studies have shown that the mechanisms of carotenoid accumulation in bivalves involved carotenoids absorption, transportation, deposition, and metabolic processing [3]. Several genes involved in carotenoid accumulation have been identified, for example, in our previous study, a scavenger receptor (*SRB-like-3*) gene shows a significantly correlation with the concentration of carotenoids in the tissues of noble scallops [18]. In *P. yessoensis*, *PySCD* (stearoyl-CoA desaturase) as the crucial enzyme in the monounsaturated fatty acids biosynthesis process, has been shown to enhance carotenoid absorption [31], and *PyBCO-like 1* responsible for its muscle coloration [32]. However, the underlying mechanisms behind the carotenoid accumulation in marine bivalves remain unclear.

In the present study, NGS and SMRT sequencing approaches were combined to generate a more complete *C. chlamys* transcriptome. The average length of ROIs can represent the full-length transcripts (Table 1), and Illumina short reads were used to correct these SMRT long reads to generate high-quality transcripts, and can reduce mis-assemblies of genes to increase their sequence identity. Total of 21,030 transcripts were annotated in eight databases, and 227 alternative splicing events and 6,032 lncRNAs were identified. To explore the genes involved in carotenoids accumulation in noble scallop, a total of 9263 DEGs in different tissues between golden and brown scallops were identified, and 3361 genes were up-regulated and 4980 genes were down-regulated. According to the KEGG pathway analysis, the Lysosome pathway was identified to be highly positive to carotenoids accumulation in golden scallop, and many carotenoids related genes shown to exhibit a high transcriptome expression level in golden scallop tissues, for example, the CD36 has a highly expressed in the hemolymph of golden scallop. In the hemolymph of golden scallops, a higher expression level of CD36, GST-omega and VLDLR may be associated to carotenoids transportation. In silkworm, Cameo2, a member of the CD36 family, has been proved to transport lutein [33]; GST and VLDLR also reported that have a significantly associated to carotenoids transport in human [34, 35]. Higher expression level ABCG5 and LDL in golden scallop mantle suggested they play a role in carotenoid accumulation. Previous studies in human have shown that ABCG5 polymorphism plays major roles in plasma response to cholesterol and carotenoids, total LDL cholesterol levels has a correlations of serum carotene concentrations [36, 37]. In addition, we also found that many structural proteins, such as actin, fibronectin, tektin and tubulin, were significantly higher in golden scallops comparing to the brown ones, previous studies believed that carotenoids transported to the target tissues and then deposited in specific receptor cells, and then binding to the structural proteins [38–40]. Therefore, we putative that these genes are associated to the carotenoids accumulation in the golden scallop.

In the early development stages of scallops, the brown scallops have a significantly higher expression level of BCMO1 than that of golden scallops at most stages, suggesting BCMO1 maybe play a crucial role in the carotenoids accumulated process. In animals, BCMO1 gene is an essential enzyme in carotenoid metabolism, which can catalyze symmetrical cleavage of carotenoids, such as α -, β -carotene and β -cryptoxanthin [41]. For example, a *BCO-like 1* was identified responsible for depositing carotenoid and coloring the muscle in *Patinopecten yessoensis* [32]. CD36, ABCG5, GST-Pi and StAR-like-3 also play important roles the early development stages of golden scallops' carotenoids accumulated process, which shown significantly higher expression levels in golden scallops than that of brown scallops. These genes are also responsible for the carotenoids absorption and transportation in other animals, for example, CD36 play a crucial role in absorbing provitamin A carotenoids in cell [42]; and the polymorphisms of ABCG5 in human affect the lutein supplementation [43]; GST-P1, exhibiting a higher affinity and specificity for zeaxanthin, is widely expressed in human epithelial tissue, showing a specific ability of carotenoid binding protein for zeaxanthin [44]; StARD3 is a membrane protein of human retinal, which has been identified as the lutein binding protein involved in absorption of lutein [45]. All these results showed that these genes play crucial roles in the early developmental phases of carotenoids accumulated process in golden scallops.

Conclusion

Full-length transcriptome of golden and brown scallops was obtained by PacBio RS II sequencing and comparative analysis was conducted using NGS data between golden and brown scallops, these results provide a basis in understanding the molecular mechanism of carotenoids accumulation in golden scallops. In addition, the expression patterns of these carotenoids accumulated genes in golden and brown scallops were analysed, and these genes showed large differences in expression, and several genes' function were identified in the early developmental stages of golden scallops. AS and lncRNA are also analyzed in the present study. Our results also provide valuable resource for further study on carotenoids accumulation in animals.

Declarations

Ethics approval and consent to participate

All animals used in this study conducted under the guidelines of institution and nation.

Consent for publication

All authors agree to publish in this journal.

Availability of data and materials

All available data and materials have been shown in the manuscript and additional files.

Competing interests

All authors listed in the manuscript have no competing interests.

DET set

the first character B and G represent Brown and Golden scallop, respectively; the second character G, M, I, B and A represent gill, mantle, intestine, blood and adductor, respectively.

Funding

Fundings have been list in the section of Acknowledgements

Authors' contributions

K. and H.L. carried out the experiment; H.K. wrote the manuscript; K.S. revised the manuscript; H.Y, S.K. and H.P. designed the experiment; H.P. instructed the experiment.

Acknowledgments

Funding for this research was provided by National Key R&D Program of China (2018YFD0901400), National Natural Science Foundation of China (31872563), China Modern Agro-industry Technology Research System (CARS-49), Department of Education of Guangdong Province, China (2017KCXTD014).

References

1. Maoka T. Recent progress in structural studies of carotenoids in animals and plants. *Arch Biochem Biophys*. 2009;483(2):191–5.
2. Toews DPL, Hofmeister NR, Taylor SA. The evolution and genetics of carotenoid processing in animals. *Trends Genet*. 2017;33(3):171–82.
3. Ahi EP, Lecaudey LA, Ziegelbecker A, Steiner O, Glabonjat R, Goessler W, et al. Comparative transcriptomics reveals candidate carotenoid color genes in an East African cichlid fish. *BMC Genom*. 2020;21(1):54.
4. Kiefer C, Emerich S, Mathias FW, Johannes VL. A class B scavenger receptor mediates the cellular uptake of carotenoids in *Drosophila*. *Proc Natl Acad Sci USA*. 2002;99:10581–6.
5. Toomey MB, Lopes RJ, Araújo PM, Johnson JD, Gazda MA, Afonso S, et al. High-density lipoprotein receptor SCARB1 is required for carotenoid coloration in birds. *Proc Natl Acad Sci USA*. 2017;114(20):5219–24.
6. Sundvold H, Helgeland H, Baranski M, Omholt SW, Vage DI. Characterisation of a novel paralog of scavenger receptor class B member I (SCARB1) in Atlantic salmon (*Salmo salar*). *BMC Genet*. 2011;12:52.
7. Li B, Vachali P, Frederick JM, Bernstein PS. Identification of StARD3 as a lutein-binding protein in the macula of the primate retina. *Biochemistry*. 2011;50(13):2541–9.
8. Sakudoh T, Tsuchida K, Kataoka H. BmStart1, a novel carotenoid-binding protein isoform from *Bombyx mori*, is orthologous to MLN64, a mammalian cholesterol transporter. *Biochem Bioph Res Co*. 2005;336:1125–35.
9. Lietz G, Oxley A, Boesch-Saadatmandi C, Kobayashi D. Importance of β , β -carotene 15, 15'-monooxygenase 1 (BCMO1) and β , β -carotene 9', 10'-dioxygenase 2 (BCDO2) in nutrition and health. *Mol Nutr Food Res*. 2012;56(2):241–50.
10. Kiser PD, Golczak M, Maeda A, Palczewski K. Key enzymes of the retinoid (visual) cycle in vertebrate retina. *BBA-Mol Cell Biol Lipids*. 2012;1821(1):137–51.
11. Kong FL, Chen SY, Ran JS, Yang CW, Jiang XS, Lan D, et al. The identification of SNPs in BCDO2 gene for skin color in Chinese indigenous chicken. *Rev Bras Cienc Avic*. 2017;19(3):393–8.
12. Lopes RJ, Johnson JD, Toomey MB, Ferreira MS, Araujo PM, Melo-Ferreira J, et al. Genetic basis for red coloration in birds. *Curr Biol*. 2016;26:1427–34.
13. Mundy NI, Stapley J, Bennison C, Tucker R, Twyman H, Kim K, et al. Red carotenoid coloration in the zebra finch is controlled by a cytochrome P450 gene cluster. *Curr Biol*. 2016;2:1435–40.
14. Ren P, Meng Y, Li B, Ma X, Si E, Lai Y, et al. Molecular mechanisms of acclimatization to phosphorus starvation and recovery underlying full-length transcriptome profiling in barley (*Hordeum vulgare*L.). *Front Plant Sci*. 2018;9:500.
15. Chao Q, Gao ZF, Zhang D, Zhao BG, Dong FQ, Fu CX, et al. The developmental dynamics of the *Populus* stem transcriptome. *Plant Biotechnol J*. 2018;17:206–19.

16. Zheng H, Liu H, Zhang T, Wang S, Sun Z, Liu W, et al. Total carotenoid differences in scallop tissues of *Chlamys nobilis* (Bivalve: Pectinidae) with regard to gender and shell colour. *Food Chem.* 2010;122(4):1164–7.
17. Zheng H, Liu H, Liu W, Sun Z, Zhang Q. Changes of total carotenoid and lipid content in scallop tissues of *Chlamys nobilis* (Bivalve: Pectinidae) during gonad maturation. *Aquaculture*, 2012, 342–343: 7–12.
18. Liu H, Zheng H, Zhang H, Deng L, Liu W, Wang S, et al. A *de novo* transcriptome of the noble scallop, *Chlamys nobilis*, focusing on mining transcripts for carotenoid-based coloration. *BMC Genomics*, 2015, 16.
19. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. *Science*. 2009;323:133–8.
20. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22:1658–9.
21. Thompson JD, Higgins DG, Gibson TJ. Clustal W. Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl Acids Res.* 1994;22:4673–80.
22. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. *De novo* transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat Protoc.* 2013;8:1494–512.
23. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 2011;39(suppl):W316–22.
24. Rhoads A, Au KF. PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics*. 2015;13:27–89.
25. Dong L, Liu H, Zhang J, Yang S, Kong G, Chu JS, et al. Single-molecule realtime transcript sequencing facilitates common wheat genome annotation and grain transcriptome research. *BMC Genom.* 2015;16:1039.
26. Moran NA, Jarvik T. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science*. 2010;328:624–7.
27. Altincicek B, Kovacs JL. Horizontally transferred fungal carotenoid genes in the two-spotted spider mite *Tetranychus urticae*. *Biol Letters*. 2012;8:253–7.
28. Du X, Song K, Wang J, Cong R, Li L, Zhang G. Draft genome and SNPs associated with carotenoid accumulation in adductor muscles of bay scallop (*Argopecten irradians*). *J Genomics*. 2017;5:83–90.
29. Li X, Bai Z, Luo H, Liu Y, Wang G, Li J. Cloning, differential tissue expression of a novel hcApo gene, and its correlation with total carotenoid content in purple and white inner-shell color pearl mussel *Hyriopsis cumingii*. *Gene*. 2014;538(2):258–65.
30. Li N, Hu J, Wang S, Cheng J, Hu X, Lu Z, et al. Isolation and identification of the main carotenoid pigment from the rare orange muscle of the Yesso scallop. *Food Chem.* 2010;118(3):616–9.
31. Li X, Ning X, Dou J, Yu Q, Wang S, Zhang L, et al. An SCD gene from the Mollusca and its upregulation in carotenoid-enriched scallops. *Gene*. 2015;564(1):101–8.
32. Li X, Wang S, Xun X, Zhang M, Wang S, Li H, et al. A carotenoid oxygenase is responsible for muscle coloration in scallop. *BBA-Mol Cell Biol Lipids*. 2019;1864(7):966–75.
33. Sakudoh T, Iizuka T, Narukawa J, Sezutsu H, Kobayashi I, Kuwazaki S, et al. A CD36-related transmembrane protein is coordinated with an intracellular lipid-binding protein in selective carotenoid transport for cocoon coloration. *J Biol Chem*. 2010;285(10):7739–51.
34. Bhosale P, Larson AJ, Frederick JM, Southwick K, Thulin CD, et al. Identification and characterization of a Pi isoform of glutathione S-transferase (GSTP1) as a zeaxanthin-binding protein in the macula of the human eye. *J Biol Chem*. 2004;279(47):49447–54.
35. Traber MG, Diamond SR, Lane JC, Brody RI, Kayden HJ. β -Carotene transport in human lipoproteins. Comparisons with α -tocopherol. *Lipids*. 1994;29(10):665–9.
36. Herron KL, McGrane MM, Waters D, Lofgren IE, Clark RM, et al. The ABCG5 polymorphism contributes to individual responses to dietary cholesterol and carotenoids in eggs. *J Nutr*. 2006;136(5):1161–5.
37. Clevidence BA, Bieri JG. Association of carotenoids with human plasma lipoproteins. In *Methods in enzymology*, 1993, 214: 33–46. Academic Press.
38. Bernstein PS, Balashov NA, Tsong ED, Rando RR. Retinal tubulin binds macular carotenoids. *Invest Ophth vis Sci*. 1997;38(1):167–75.
39. Crabtree DV, Ojima I, Geng X, Adler AJ. Tubulins in the primate retina: evidence that xanthophylls may be endogenous ligands for the paclitaxel-binding site. *Bioorgan Med Chem*. 2001;9(8):1967–76.
40. Zhang Y, Zhang L, Sun J, Qiu J, Hu X, Hu J, et al. Proteomic analysis identifies proteins related to carotenoid accumulation in Yesso scallop (*Patinopecten yessoensis*). *Food Chem.* 2014;147:111–6.
41. Alvarez R, Melendez-Martinez AJ, Vicario IM, Alcalde MJ. Carotenoid and vitamin A contents in biological fluids and tissues of animals as an effect of the diet: a review. *Food Rev Int*. 2015;31(4):150213130216007.
42. Moussa M, Gouranton E, Gleize B, El Yazidi C, Niot I, et al. CD36 is involved in lycopene and lutein uptake by adipocytes and adipose tissue cultures. *Mol Nutr Food Res*. 2011;55:578–84.
43. Reboul E, Borel P. Proteins involved in uptake, intracellular transport and basolateral secretion of fat-soluble vitamins and carotenoids by mammalian enterocytes. *Prog Lipid Res*. 2011;50(4):388–402.
44. Bhosale P, Larson AJ, Frederick JM, Southwick K, Thulin CD, et al. Identification and characterization of a Pi isoform of glutathione S-transferase (GSTP1) as a zeaxanthin-binding protein in the macula of the human eye. *J Biol Chem*. 2004;279(47):49447–54.
45. Binxing L, Preejith V, Jeanne MF, Paul SB. Identification of StARD3 as a lutein-binding protein in the macula of the primate retina. *Biochemistry*. 2011;50(13):2541–9.
46. Declarations.

Tables

Table 1 Full-length sequences statistics.

Samples	cDNA Size	Reads of Insert	Number of five prime reads	Number of three prime reads	Number of poly-A reads	Number of filtered short reads	Number of non-full-length reads	Number of full-length reads	Number of full-length non-chimeric reads	Average full-length non-chimeric read length	Full-Length Percentage (FL%)	Artificial Concatemers %
F01	All	182,010	112,996	116,545	112,824	8,457	80,556	92,997	90,727	2,733	51.09%	2.44%

cDNA size: insert fragment size of cDNA libraries; reads of insert: the number of reads of insert (ROI) sequences; Number of five prime reads: the number of ROI sequences containing 5' primer; Number of three prime reads: the number of ROI sequences containing 3' primer; Number of poly-A reads: the number of ROI sequences containing poly-A; Number of filtered short reads: the number of filtered ROI of <300 bp; Number of non-full-length reads: the number of non-full-length ROI; Number of full-length non-chimeric reads: the number of full-length non-chimeric ROI; Average full-length non-chimeric read length: average length of full-length non-chimeric sequence; Full-length percentage (FL%): the percentage of full-length sequence in ROI sequence; Artificial concatemers (%): the percentage of full-length chimeric sequence in full-length sequence.

Table 2 DEGs in the same tissues between Golden scallop and Brown scallop.

DET set	DET number	Up	Down
GG vs BG	2665	1246	1419
GM vs BM	2941	1411	1530
GI vs BI	2150	1329	821
GB vs BB	2924	1115	1809
GA vs BA	4508	3495	1013

DET set: the first character B and G represent Brown and Golden scallop, respectively; the second character G, M, I, B and A represent gill, mantle, intestine, blood and adductor, respectively.

Figures

Figure 1 Functional annotation. Venn distribution of the non-redundant transcripts in different databases (a). Homologous species distribution of noble scallop *Chlamys nobilis* annotated in the NR database (b).

Figure 2

The SSR type distribution Note: p1, p2, p3, p4, p5, p6 and c represent perfect single-base repeat, double-base repeat, three-base repeat, four-base repeat, five-base repeat, six-base repeat and mixed SSR, respectively, (c contains at least two perfect SSRs, and the distance between them less than 100 bp).

Figure 3

Venn diagram of the number of lncRNAs predicted by Calculator (CPC), Coding-Non-Coding Index (CNCI), Coding Potential Assessment Tool (CPAT), and pfam protein structure domain analysis.

Figure 4

Venn diagrams of genes uniquely expressed in each tissue or genes shared between one or more tissues in golden and brown scallops.

Figure 5

GO enrichment analysis of DEGs in the same tissue.

Figure 6

KEGG pathway enrichment analysis of DEGs in the same tissue.

Level2 GO terms of *Melopsittacus undulatus*

Figure 7

GO enrichment analysis of DEGs in all tissue.

Figure 8

KEGG pathway enrichment analysis of DEGs in all tissue.

Figure 9

The expression levels of carotenoids related genes in the early developmental stages in scallops.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS8.xlsx](#)
- [TableS8.xlsx](#)
- [TableS7.xlsx](#)
- [TableS7.xlsx](#)
- [TableS6.xlsx](#)
- [TableS6.xlsx](#)
- [TableS5.xlsx](#)
- [TableS5.xlsx](#)
- [TableS4.docx](#)
- [TableS4.docx](#)
- [TableS3.xlsx](#)
- [TableS3.xlsx](#)
- [TableS2.docx](#)
- [TableS2.docx](#)
- [TableS1.docx](#)
- [TableS1.docx](#)