KOREAN JOURNAL OF ICHTHYOLOGY, Vol. 32, No. 3, 117-129, September 2020
ISSN: 1225-8598 (Print), 2288-3371 (Online). DOI: https://doi.org/10.35399/ISK.32.3.1

Received: August 8, 2020
Revised: September 2, 2020
Accepted: September 7, 2020

체색 패턴이 다른 개볼락(*Sebastes pachycephalus*) 피부 전사체 프로파일링

*장요순**

한국해양과학기술원 동해환경연구센터

Skin Transcriptome Profiling of the Blass Bloched Rockfish (*Sebastes pachycephalus*) with Different Body Color Patterns ***by Yo-Soon Jang**** (East Sea Environment Research Center, Korea Institute of Ocean Science & Technology, Uljin 36315, Republic of Korea)

ABSTRACT The body color pattern in fish is a distinctive feature for species identification. The blass bloched rockfish *Sebastes pachycephalus* is a commercially important marine fish species, distributed in the central and southern parts of Korea and south Hokkaido of Japan. It has a morphological feature divided into four subspecies according to with or lacking distinct spots on the body surface, and to the location of markings on the body surface. However, the genetic basis of body color pattern of *S. pachycephalus* is still unknown. Thus we analyzed the transcriptome of *S. pachycephalus* skin samples using RNA-seq analysis to investigate functional genes related to body color patterns. The experimental skin samples were prepared by classified into 'Wild type' (lacking distinct spots and markings) and 'Color type' (with distinct spots and marking). Two skin sample transcriptomes were compared pairwise and the results revealed that were 164 differentially expressed unigenes in the skin samples of 'Wild type' and 'Color type'. Gene Ontology analysis of 164 differentially expressed unigenes showed that these genes were included in the functional group of molecular function (2 genes), biological process (46 genes), and cellular component (6 genes). There were several genes that body color type skin specific expression and the genes were CTL (Galactose-specific lectin nattectin), CUL1 (Cullin-1), CMAS (N-acylneuraminate cytidylyltransferase), NMRK2 (Nicotinamide riboside kinase 2), ALOXE3 (Hydroperoxide isomerase ALOXE3), SLC4A7 (sodium bicarbonate cotransporter 3). Our study is the first attempt to search for functional genes involved in the formation of body color patterns in *S. pachycephalus*. The differentially expressed unigenes obtained in this study can be used as candidate genes for functional gene study related to body coloration of fish.

Key words: Transcriptome, body color, RNA-seq, differentially expressed gene, *Sebastes pachycephalus*

서 론

체색과 체색 패턴은 척추동물에서 다양하게 나타나는 표현형질로, 포식자의 위협으로부터 회피하기 위하여 위장하거나, 빛을 차단 또는 감지하여 체온을 조절하는 기능요소에 해당한다. 체색은 생물이 지닌 고유특성으로 색상이나 체색 패턴을 종 식별지표로도 이용한다(Hubbard *et al*., 2010). 생물의 생체 색소는 형태학적 변화와 생리학적 변화에 의해 만들어진다. 먼저 형태학적 변화는 피부 색소의 형태, 농도, 밀도 및 색소포(chromatophore)의 분포에 따라 생기는 것으로 몇 주 동안에 걸쳐 비교적 천천히 진행되고, 지속적이며 장기적으로 영향을 끼친다. 반면 생리학적 변화는 색소세포 내의 구조나 pigment vesicle의 운동성에 따른 일시적인 반응으로, 빛과 같은 환경요인의 직접적인 영향으로 인한 1차 생리학적 변화와 부신피질자극 호르몬(adrenocorticotropic hormone, ACTH)이

저자 직위: 장요순(책임기술원)
*Corresponding author: Yo-Soon Jang Tel: 82-54-780-5326, Fax: 82-54-780-5349, E-mail: jangys@kiost.ac.kr

나 멜라닌생성세포자극 호르몬(alpha-melanocyte-stimulating hormone, α-MSH) 등에 의한 신경 및 내분비 조절로 인한 2차 생리학적 변화로 구분한다(Rodionov *et al*., 2003).

동물의 색소포는 특징적인 색상 범위에 따라 흑색소포(melanophore, 검정색~갈색), 적색소포(erythrophore, 적색~오렌지색), 황색소포(xanthophore, 황토색~노랑색), 홍색소포(iridophore, metallic iridescent / 은백색, 금속색, 녹청색), 백색소포(leucophore, 백색 / 희뿌연색), 청색소포(cyanophore, 푸른색)로 구별하고, 척추동물의 색과 체색 패턴에 관여하는 150개 이상의 유전자가 보고되었다(Leclercq *et al*., 2010). Hoekstra (2006)와 Protas and Patel (2008)은 사육집단과 야생집단을 조사하여 척추동물에서 색소침착의 다양성 및 체색 패턴 생성에 관한 유전적인 기초와 발달기작을 보고하였다. 최근에는 생물의 색소침착에 관여하는 유전자의 기능연구 대상이 야생의 개체군 수준까지 연구범위가 확대되었으며, 척추동물의 여러 분류군을 대상으로 한 연구에서 색소합성 관련 기능유전자가 체색의 유지 및 조정 역할을 담당하는 것으로 밝혀졌다(Hubbard *et al*., 2010).

경골어류의 체색은 생리학적 변화와 형태학적 변화가 크고, 생존, 성장, 번식 등에 중요한 요소로 작용하기 때문에, 진화압력(evolutionary pressure)이 강력하여 외부 환경의 급격한 변화에 저항하는 방향으로 진화가 이루어진다(Leclercq *et al*., 2010). 경골어류는 흑색소포, 적색소포, 황색소포, 홍색소포, 백색소포, 청색소포를 비롯하여 6개 이상의 색소세포를 가지고 있으며, *matp*, *oca4*, *sox10*, *kit*, *ednrb*1, *slc24a5* 등 많은 유전자들이 체색을 만드는 과정에 관여한다(Kelsh *et al*., 1996; Kelsh *et al*., 2004; Parichy, 2006). 말라위 호수의 오렌지 시클리드(orange-blotch, OB) 체색 패턴은 암컷의 성 결정 유전자와 밀접하게 관련된 Pax7 유전자의 돌연변이에 의한 성선택 형질로 나타나고(Roberts *et al*., 2009), 큰가시고기(Stickleback fish, *Gasterosteus aculeatus*)의 아가미 색상과 피부색 변화는 Kit 리간드(KITLG) 유전자의 발현량 차이로 나타나며(Miller *et al*., 2007), 제브라피쉬(*Danio rerio*)의 golden mutant는 멜라노좀(melanosomes)의 개수, 크기 및 밀도의 변화와 관련이 있는 것으로 보고되었다(Lamason *et al*., 2005).

전사체(transcriptome)는 생명정보가 DNA에서 RNA로 전달된 전사물(transcript)의 총합을 의미하고, 전사물은 mRNA, tRNA 및 non-coding RNA 등 RNA의 모든 것을 포함하며, 전사체 프로파일은 RNA의 염기서열을 모두 읽어내는 RNA-seq (RNA sequencing) 분석법을 이용하여 확보한다. RNA-seq은 high-throughput sequencing 방법을 통하여 전사물의 동정과 정량을 동시에 수행할 수 있고, 다양하게 적용할 수 있는 RNA-seq 실험방법과 분석법이 알려져 있다. 그러나 현재까지는 RNA-seq을 여러 샘플에 응용할 수 있는 최적의 분석 파이프라인이 없어, 연구대상 생물의 특성과 연구 목표에 따라 다양한 분석 전략을 적용해야 하는 실정이라 초심자가 쉽게 접근하기에는 여전히 어려움이 있다.

어류의 체색에 관여하는 핵심 유전자 탐색 연구가 전사체 수준에서 수행되었으며, Wang *et al*. (2014)은 잉어 Oujiang color 품종(RB; red with big black spots, WW; whole white, CC; gynogenic common carp)의 피부 전사체를 분석하여 color pattern-specific SNP marker를 확보하였다. Jiang *et al*. (2014)은 잉어 피부 색소생성 과정에 작용하는 유전자를 찾기 위하여 Xingguo red carp (XGC)와 Yellow River carp (YRC) strain 조직의 전사체를 비교 분석하여 차등발현유전자(Differentially expressed gene, DEG)를 조사하였다. 차등발현 양상을 나타낸 유전자의 일부는 pheomelanin 생성과 관련이 있었고, 특히 agouti signaling protein (ASIP) 유전자와 cysteine/glutamate transporter (*xCT*/*SLC7a11*) 유전자는 XGC (Xingguo red carp) strain에서 3배 이상 발현량이 증가한 것을 확인하였다.

개볼락은 쏨뱅이목 양볼락과 어류로 한국 중부 및 남부해역과 일본 홋카이도 남쪽 바다에 분포하는 정착성 어류로 근해 암초지역에 서식한다. 개볼락은 서식 깊이와 서식 장소에 따른 체색의 변이뿐 아니라, 형태형질의 복합적인 특성으로 인하여 분류학적으로 혼동이 많은 종이다(Kai *et al*., 2011). 체색 패턴이 복잡한 개볼락을 반점(spot)의 색상과 마킹(marking)의 분포 위치에 따라 *Sebastes pachycephalus pachycephalus* (복부에는 갈색 반점이 있으나, 등에는 뚜렷한 마킹이 없음), *Sebastes pachycephalus nigricans* (복부와 등에 뚜렷한 반점이나 마킹은 없음), *Sebastes pachycephalus nudus* (복부에는 뚜렷한 반점이나 마킹이 없지만, 등에 뚜렷한 노란색 마킹이 있음), *Sebastes pachycephalus chalcogrammus* (복부에 뚜렷한 반점이 없으나, 등에 적갈색 마킹이 있음)와 같이 4개의 아종으로 구분하였다(Matsubara, 1943; Kai *et al*., 2011에서 재인용). Kai *et al*. (2011)은 4개 아종 간의 유전적 차이를 확인하기 위하여 미토콘드리아 DNA의 D-loop 영역을 분석하였으나, 유의미한 차이는 발견하지 못하였다. Nakabo (2002)는 개볼락, *Sebastes pachycephalus*을 등(dorsum)에 있는 유색 마킹의 유무와 등 가시(dorsal spine) 아래에 있는 비늘의 유무를 근거로 Matsubara (1943)가 구분한 4개의 아종을 2개의 종으로 구분할 수 있다고 제안하였다. Nakabo (2002)에 따르면 *S. pachycephalus nudus*와 *S. pachycephalus chalcogrammus*는 등에 노란색이나 적갈색 마킹이 있고, 15번째 또는 그 뒤쪽의 등 가시 아래에 작은 비늘이 있으며, *S. p. pachycephalus*와 *S. p. nigricans*는 등에 마킹이 없지만, 등 지느러미의 전체 가시 부분 아래에 작은 비늘이 있다. Kai and Nakabo (2013)는 *Sebastes pachycephalus complex*를 분류학적으로 검토하여 *S. pachycephalus*와 *S. nudus*로 명명하였고, Yu and Kim (2014)은 이전 연구 기준에 따라 분류한 *Sebastes pachycephalus*를 다시 분석하여 *S. pachycephalus*와 *S. nudus*로 구분한 후, *S. nudus*

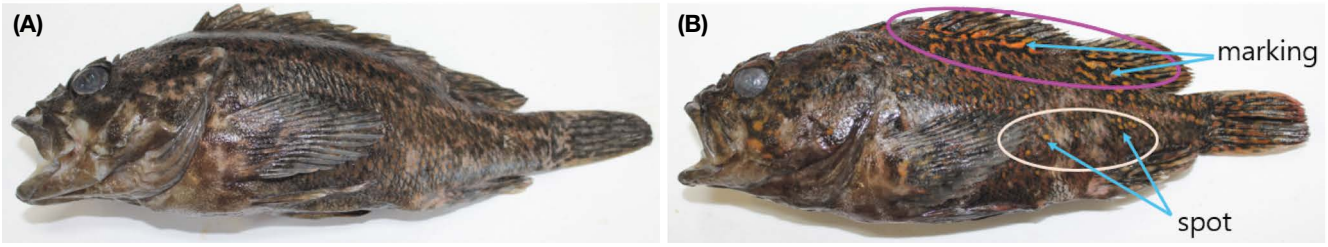

Fig. 1. Two skin color types of blass bloched rockfish (*Sebastes pachycephalus*) for RNA-seq analysis. (A) Wild type (Bbrf-Wt), No spots and No marking; (B) Color type (Bbrf-Ct), Above 2/3 scattered red spot of all skin.

의 한국명을 황점개볼락으로 제안하였다.

어류의 체색이나 마킹 패턴과 같은 뚜렷한 특징은 종 동정에 유용한 형태형질로 *Sebastes* 어류의 종 식별에도 이용하고 있으나, 반점 색상과 마킹 분포 특성으로 인한 복잡한 체색 패턴을 가진 개볼락은 형태지표를 이용한 종 동정 시 혼동이 빈번하였다. 개볼락의 복잡한 체색 패턴은 서식환경 요인뿐 아니라, 유전요인에 기인한 생명현상 때문인 것으로 판단된다. 이에 따라 본 연구에서는 기존의 분류 기준에 따른 개볼락의 복잡한 체색 패턴 형성과 관련된 후보유전자 발굴 및 체색 타입별 발현특성을 파악하기 위하여, 개볼락 몸 전체에 있는 반점과 마킹의 유무를 기준으로 Wild type과 Color type으로 각각 구분한 체색 타입별 시료의 피부 전사체를 프로파일링하였다.

재료 및 방법

1. 실험재료

본 연구에 사용한 개볼락(*Sebastes pachycephalus*)은 임원항(강원도 삼척시 원덕읍)에서 구입하였고, 체색 타입별 실험재료는 몸 전체에 있는 반점과 마킹의 유무를 기준으로 Wild type (반점과 마킹이 전혀 없음) 또는 Color type (몸 전체에 붉은 반점이 있고 배 부분에 마킹이 있음)으로 구분하여 (Fig. 1) 확보하였다. 전사체 분석을 위한 실험어는 체색 타입별로 각 1개체를 선정하였다. Wild type 샘플은 등지느러미 바로 아래 피부조직에서 취하였고, Color type 샘플은 등지느러미 아래 반점과 마킹이 뚜렷하게 나타난 피부조직에서 취하였다. 피부조직은 1.0 cm×1.0 cm×0.2 cm (가로×세로×두께) 크기로 절단하여 표피 및 진피층 일부를 포함하도록 샘플링하였으며, 체색 타입별로 각 5개 조각(100 mg)을 취하여 분석에 사용하였다.

2. Total RNA 분리 및 정제

개볼락 체색 타입별 피부조직 샘플을 100 mg씩 취하여 호모게나이저를 사용하여 파쇄하였다. 파쇄한 시료에 TRIzol® Reagent (Invitrogen, USA) 1 mL을 넣어 충분히 녹인 후 200 μL의 클로로포름을 첨가하여 Vortex Mixer (Vortex-Genie® 2)로 강하게 섞었다. 상온에서 5분 동안 반응시킨 후, 원심분리하여 회수한 상층액을 isopropyl alcohol로 침전시켜 total RNA를 분리하였다. 분리한 total RNA는 MiniBEST Universal RNA Extraction Kit (TaKaRa Bio Inc., Japan)의 RNA Spin Column을 사용하여 정제과정을 거친 후, Next Generation Sequencing (NGS) 플랫폼 사용을 위하여 RNA 품질을 조사하였다. 정제한 total RNA 품질은 BioAnalyzer 2100 system (Agilent Technologies, CA, USA)을 이용하여 농도(≥5 μg), 28S/18S 측정값(≥1.0) 및 RIN (RNA Integrity Number) 값(≥7)을 확인하고 RNA-seq 분석 가능여부를 결정하였다.

3. cDNA 라이브러리 제작 및 Sequencing

Total RNA 5 μg을 Covaris® system으로 절단하고(300~500 bp) Bioanalyzer High Sensitivity Kit (Agilent Technologies, Inc., USA)로 크기를 확인한 후, random hexamer로 제 1가닥(first strand) cDNA를 합성하였다. 제 1가닥의 cDNA를 주형으로 제 2가닥(second strand)의 cDNA를 합성한 후, TruSeq™ RNA Sample Preparation Kit (Illumina, Inc., USA)를 사용하여 평활 말단(blunt-end) 단편을 만들고, 염기 A 첨가(A-tailing) 및 index adapter의 부착 과정을 거친 최종산물을 증폭하여 염기서열 결정용 라이브러리를 각각 제작하였다. 염기서열 결정용 라이브러리의 농도는 qPCR 방법으로 측정하였고, cBot 기기를 이용하여 클러스터를 합성한 후(클러스터 수는 300~400 K 기준), HiSeq 2500 (Illumina, Inc., USA) 플랫폼을 사용하여 염기서열을 분석하였다(2×125 bp paired-end sequencing).

4. 염기서열 어셈블리 및 unigene 확보

염기서열 분석 후 FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/)로 raw data quality control을 진행하였고, paired-end (PE) sequencing 데이터는 N이 10% 이상

Table 1. Summary of the sequencing data from blass bloched rockfish (*Sebastes pachycephalus*)

Sample	Number of raw reads	Number of clean reads	Number of low quality reads
Bbrf-Wt*	78,199,118	73,392,934 (93.9%)	4,806,184 (6.1%)
Bbrf-Ct*	77,946,600	73,454,304 (94.2%)	4,492,296 (5.8%)
Total reads	156,145,718	146,847,238	9,298,480

*Color types of skin sample: Bbrf-Wt, Wild type of blass bloched rockfish; Bbrf-Ct, Color type of blass bloched rockfish

포함된 read 제거, Q20 미만의 염기가 20% 이상인 read 제거, 평균 quality가 Q20 미만인 read 제거, 서열의 양끝에 존재하는 Q20 미만의 염기를 제거하는 기준으로 필터링하였다.

염기서열 어셈블리는 각 샘플의 read 데이터를 한 개로 합쳐 Trinity (Grabherr *et al*., 2011; Hass *et al*., 2013) 프로그램을 이용하여 수행하였다. 조립된 전사체 서열의 집단화(clustering)는 TGICL (Pertea *et al*., 2003)을 사용했으며, 중복된 read와 키메라 전사물(transcript)을 제거하여 각 집단을 대표하는 서열을 확보하였다. 조립된 염기서열을 TransDecoder (Haas *et al*., 2013)와 GeneID (Blanco *et al*., 2007) 프로그램으로 coding sequence (CDS)를 예측하고 검증하였으며, log-likelihood 스코어가 0보다 큰 것을 선택하여 unigene으로 확보하였다.

5. Annotation 및 차등발현유전자 분석

Unigene의 annotation은 BLASTx 및 InterProScan (v5)을 이용하여 수행하였고, RSEM (RNA-Seq by Expectation Maximization / Li and Dewey, 2011) 프로그램으로 유전자의 발현량을 계산하였으며, FPKM (Fragment Per Kilobase of transcripts per Million mapped reads)으로 발현값을 측정하였다. 차등발현유전자 분석은 TCC (Sun *et al*., 2013) 프로그램의 DEGES/DESeq 방법을 활용하였으며, 정규화(normalization) 과정을 3번 반복하였고, 기준값은 q-value를 0.01 이하로 설정하여 DEG를 선정하였다.

차등발현유전자의 기능정보 파악을 위한 Gene Ontology (GO) 분석은 depth를 1로 지정하여 3개의 functional category, 즉 biological process, cellular component, molecular function으로 분류하였다.

결 과

1. 전사체 염기서열 해독 및 read data 전처리

개볼락 체색의 유전특성을 이해하고 반점 및 마킹 패턴과 관련된 기능유전자를 발굴하기 위하여 체색 타입별 전사체를 분석하였다. RNA-seq 분석법으로 개볼락 몸 전체에 붉은 반점이 있고, 배 부분에 마킹이 있는 Color type과 뚜렷한 반점이나 마킹이 없는 Wild type의 피부 전사체 염기서열을 해독하였다. 체색 타입별 피부 샘플의 total RNA로 제작한 cDNA 라이브러리를 시퀀싱하여 78,199,118 PE read와 77,946,600 PE read를 각각 확보한 후(Table 1), NCBI의 SRA database (https://www.ncbi.nlm.nih.gov/sra/)에 등록하였다 (PRJNA623997).

Table 2. Length distribution of assembled unigenes for use in comparative analysis of transcriptome

Length (bp)	Number of assembled unigenes
200~500	63,236
501~1,000	20,124
1,001~1,500	7,996
1,501~2,000	4,330
2,001~2,500	2,344
2,501~3,000	1,282
3,001~3,500	766
3,501~4,000	413
4,001~4,500	251
4,501~8,000	348
8,001~18,619	22
Total	101,112

개볼락 전사체를 어셈블리하기 위하여 서열 read 데이터를 제거하는 기준에 따라 필터링한 결과, 시퀀싱 품질은 Wild type 샘플과 Color type 샘플 모두 Q20 비율이 93% 이상이었다. 체색 타입별 전사체 비교 분석에 사용할 PE clean read를 73,392,934개 (Wild type)와 73,454,304개 (Color type)를 확보하였다.

2. 개볼락 체색 타입별 전사체의 *de novo* assembly

개볼락 피부 전사체의 어셈블리는 체색 타입 각각의 clean read를 한 개로 합쳐서 *de novo* assembly 방식으로 진행하였으며, 유사성 기준을 0.94로 정하여 염기서열을 클러스터링하여 코딩 서열을 예측하고 검증한 후 unigene을 선별하였다. 본 연구에서 확보한 unigene 개수는 101,112개로 전체 크기는 66,25 Mb (66,252,764 bp)이었고, 유전자의 평균 크기

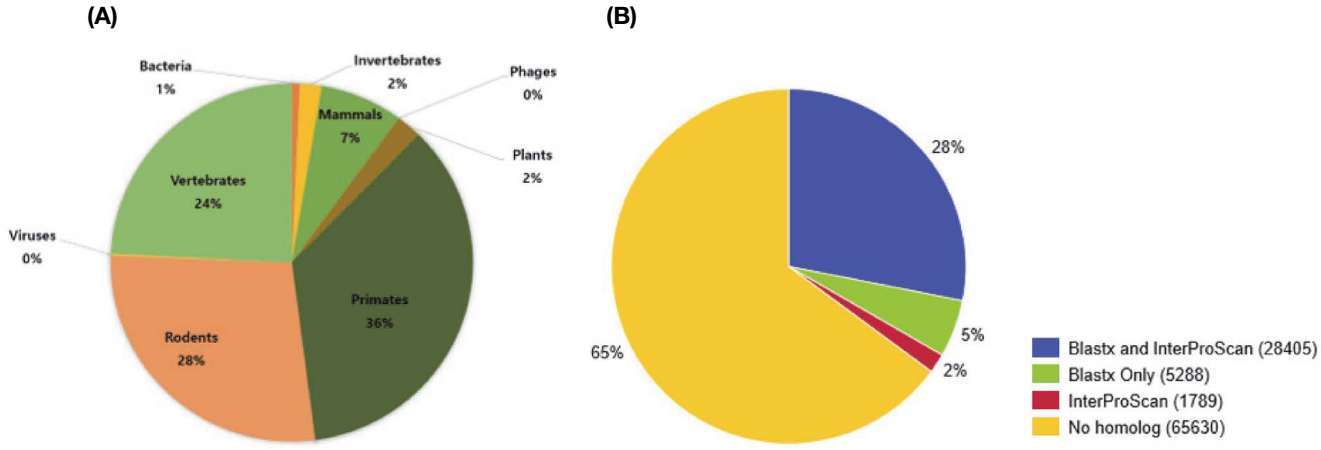

Fig. 2. Function annotation of the unigenes. (A) Distribution of the annotated unigenes based on BLAST search; (B) Annotation results of consensus between Blastx and InterProScan (v5).

Table 3. Statistics of functional annotation for the transcriptome of blass bloched rockfish

	Blastx & InterProScan	Blastn	Blastx	InterProScan	Total
	No.	No.	No.	No.	No.
Has homologous	28,405 (28.1%)	33,693 (33.3%)	5,288 (5.2)	31,063 (53.3%)	
No homologous		67,419 (66.7%)		27,191 (46.7%)	
All annotated unigenes					35,482
CDS of unigenes				58,254	
No homolog					65,630
All unigenes					101,112

는 665 bp였다. Unigene의 길이별 분포(Table 2)를 살펴보면, 200~500 bp 크기 구간이 63,236개로 가장 많았고, 4.0 kb 이상의 transcript는 621개였다. 가장 큰 전사체 크기는 18,619 bp였으며, 1 kb 크기 이상의 전사체는 17,752개로 전체 unigene의 17.6%에 해당하였다.

3. Unigene의 functional annotation

개볼락 피부 전사체를 어셈블리하여 확보한 unigene의 기능은 NCBI Nr (Non-redundant protein) 데이터베이스의 BLASTx 검색과 InterProScan tool을 이용한 functional annotation 결과로 파악하였다(Fig. 2, Table 3). BLAST 및 InterProScan 검색으로 예측한 코딩 서열로 unigene 35,482개(35%)의 기능을 annotation 할 수 있었다. NCBI Nr DB의 Blastn 검색(filter 기준: e-value≤1e-10, Best hits) 결과, 33,693개 unigene이 annotation 되었고, 이들 중 29,481개(87.5%)는 영장류(11,993개, 36%), 설치류(9,327개, 28%), 척추동물(8,161개, 24%)의 유전자와 상동성을 나타내었다(Fig. 2A). InterProScan 검색으로 확인한 코딩 서열(58,254개) 중 31,063개(53.3%)는 기능 정보가 알려져 있었고, 나머지 27,191개는 기능 정보가 명확하지 않았다. BLASTx와 InterProScan을 동시에 적용한 functional annotation 결과에서 기능이 확인되지 않은 65,630개 unigene은 향후 연구에서 Swiss-Prot, KEGG, COG 등의 데이터베이스를 추가 검색하여 annotation 하고자 한다.

4. Unigene의 발현량 계산 및 체색 타입별 차등발현유전자 발현 분석

개볼락 피부 전사체로부터 확보한 unigene의 발현량은 RSEM 프로그램을 이용하여 계산하였다(Table 4). 반점이나 마킹이 없는 Wild type (Bbrf-Wt) 피부샘플에서는 85,831개의 unigene이 발현되었고, 붉은색 반점을 가진 Color type (Bbrf-Ct) 피부샘플에서는 84,198개 unigene이 발현되는 것을 확인하였다. 체색 타입별 차등발현유전자를 FPKM (Fragment Per Kilobase of transcripts per Million mapped reads) 값을 기준(>1.0)으로 선정하여 기능을 확인한 결과, Wild type 피부샘플에서 발현되는 유전자 중 22,504개가 기능이 밝혀져 있고, Color type 피부샘플에서는 21,695개가 기능이 밝혀진 유전자였다.

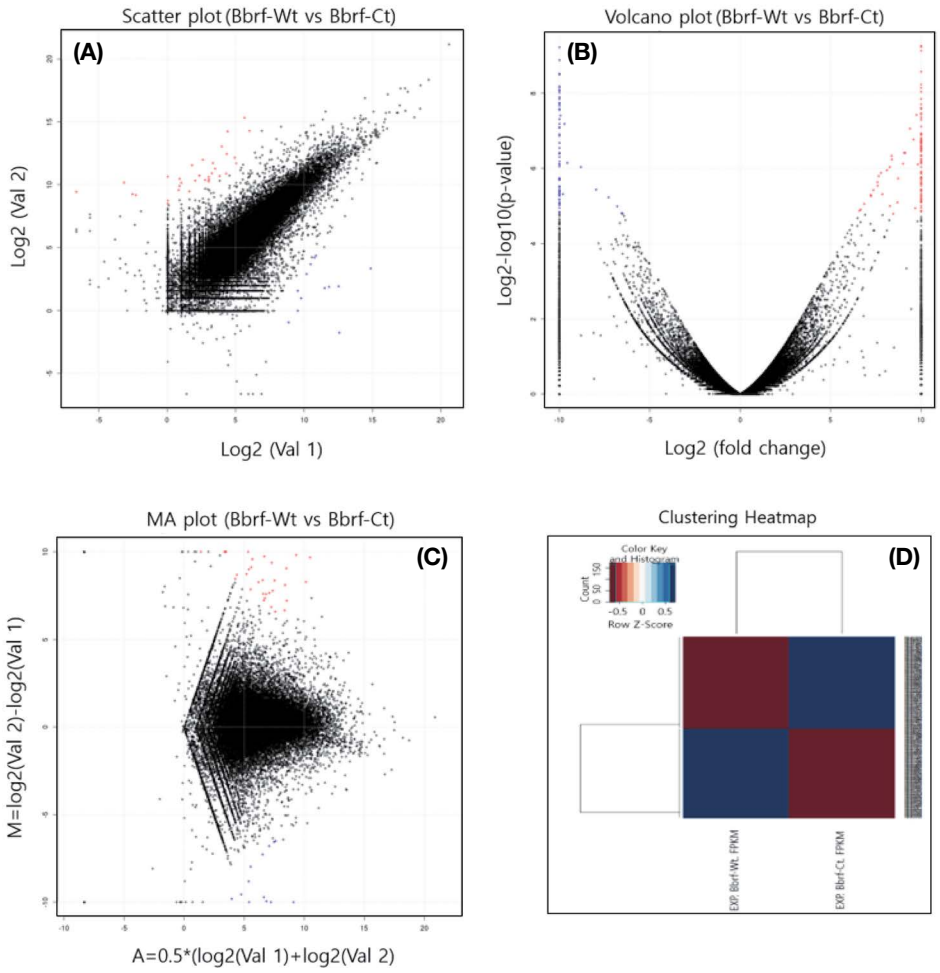

Fig. 3. Gene expression of differentially expressed genes in the two skin color types of blass bloched rockfish (*Sebastes pachycephalus*). (A) ScatterpPlot; (B) Volcano plot; (C) MA plot; (D) Clustering Heatmap.

Table 4. Expression information of unigenes derived from blass bloched rockfish by skin color types

Sample	Gene			Gene (>fpkm 1.0)		
	Expressed		Unexpressed	Expressed		Unexpressed
	Known	Novel		Known	Novel	
Bbrf-Wt	32,094	53,737	(15,281)	22,504	33,004	(4,611)
Bbrf-Ct	31,226	52,972	(16,914)	21,695	31,804	(6,620)

*Color types of skin sample: Bbrf-Wt, Wild type of Blass bloched rockfish; Bbrf-Ct, Color type of Blass bloched rockfish

개볼락의 체색 형성에 관여하는 유전자를 확보하기 위하여 FPKM 값이 1.0 이상인 unigene만 선택하여 체색 타입별 피부 전사체의 발현특성 분석을 진행하였다. 차등발현 후보유전자는 발현량이 2배 이상(|logFC|≥1)인 전사체를 선발하였고, FDR (False Discover Rate) 5% (FDR≤0.5)를 적용하여 분석하였다.

개볼락의 체색 타입에 따라 발현량 차이를 나타낸 916개의 전사체 중 164개를 TCC 프로그램을 활용(Cutoff: q-value< 0.01)하여 체색 타입별 차등발현유전자로 최종 선정하였다. 차등발현유전자를 분석하기 위한 체색 타입별 전사체의 발현양상, 발현 유사성, 발현값과의 상관관계 정보는 Fig. 3에 나타내었다. 체색 타입별 피부 전사체의 발현양상은 Scatter plot (Fig.

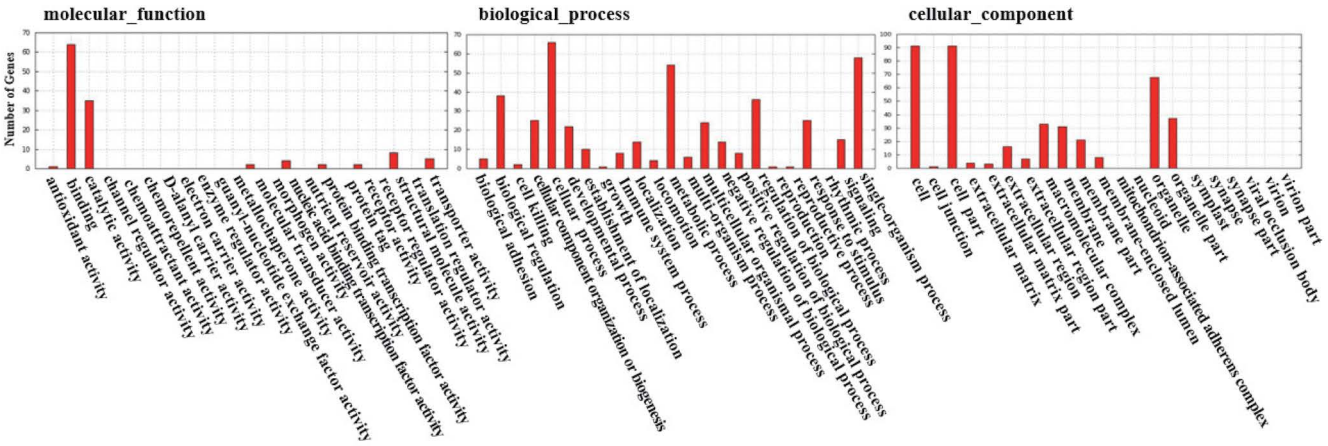

Fig. 4. Gene ontology classification of skin color types differentially expressed genes of blass bloched rockfish.

3A)으로 확인하였고, Wild type 피부샘플의 normalized data (log2)와 Color type 피부샘플의 normalized data (log2)를 X축과 Y축에 각각 표시하였다. 검정색 사선 아래 파란색 점은 발현량이 2배 이상 감소한 유전자들에 해당하고, 검정색 사선 위 빨간색 점은 2배 이상 증가한 유전자들에 해당한다. Volcano plot (Fig. 3B)은 Scatter plot의 기능과 거의 동일하고, 각 유전자의 log2 Fold Change와 p-value를 분석하기 위하여 사용되는 것으로 왼쪽의 파란색 점은 2배 이상 감소한 유전자들이고, 오른쪽 빨간색 점은 2배 이상 증가한 유전자들, 그리고 검은색 점 위는 p-value 0.05 이하인 유전자들이다. MA plot (Fig. 3C)은 유전자의 발현값과 p-value의 상관관계를 확인하기 위하여 사용했으며, 체색 타입별 피부 전사체의 유사성은 Clustering Heatmap (Fig. 3D)으로 확인하였다.

5. 체색 타입별 피부에서 차등발현하는 유전자의 Gene Ontology enrichment 분석

개볼락의 체색 타입별 피부에서 발현량 차이를 나타낸 차등발현유전자 164개 중 Color type에서 발현량이 증가하는 것은 83개였고, 발현량이 감소하는 것은 81개였다. 체색 타입 특이적으로 Color type 피부에서만 발현하는 유전자 51개와 Wild type 피부에서만 발현하는 유전자 68개를 확인하였으며, 이들 중 79개 (Color type 35개, Wild type 33개)가 기능이 밝혀진 유전자에 해당하였다. Table 5는 체색 타입별 차등발현유전자의 발현량 차이를 기준으로 상위 20개에 해당하는 유전자를 나타낸 것이다. Color type 피부에서 발현량이 15배 이상 증가한 유전자에는 *MYH7*, *CUL1*, *ZNF106*, *FAM73A*, *MYOT*, *TOMM7*, *CTL*, *VMA21*, *B2M*, *ISCA1*, *EMILIN1*, *CAC1S*, *LGALS9*, *CMAS*, *WDR33*, *ASCC3*, *ANXA6*, *SAMHD1*, *MRP63*, *ACOT11* 등이 있었다. 반면 Color type 피부에서 발현량이 15배 이상 감소한 유전자에는 *NMRK2*, *FLNC*, *ACADM*, *GM129*, *SLC38A2*, *PCBP2*, *ETNK1*, *RNF213*, *FOS*, *BAG4*, *MURC*, *MRPL18*, *HRSP12*, *RBM12B*, *SCEL*, *UBE2L3*, *SLC4A7*, *DCAF11*, *MST4*, *Ppan* 등이 있었다.

개볼락 체색 타입별 피부에서 차등적으로 발현하는 유전자에 대하여 Gene ontology enrichment 분석 결과 (Fig. 4), 생화학적 산물에 따라 구분하는 분자수준에서의 기능은 binding (52.03%), catalytic activity (28.46%), structural molecule activity (6.50%)의 순서로 분포하였다. 또한 생물학적인 대사과정에서의 역할에 따른 구분에서는 cellular process (15.10%), single-organism process (13.27%), metabolic process (12.36%) 순서로 나타났고, 세포의 구성요소 (Cellular component)에 따라 구분할 경우에는 세포 (22.14%), 세포 부위 (22.14%), 세포소기관 (16.55%)의 순서로 분포하는 것을 확인하였다.

고 찰

개볼락 (*Sebastes pachycephalus*)은 한국의 중부 및 남부 해역과 일본의 홋카이도 남쪽 바다에 분포하고, 서식 깊이와 서식 장소에 따라 체색의 변이가 심하여 다양하고 복잡한 체색 패턴을 보인다. Matsubara (1943; Kai *et al*., 2011에서 재인용)는 복잡한 체색 패턴을 가진 개볼락을 몸 전체에 분포한 반점과 마킹을 기준으로 4개의 아종으로 구분했으나, Nakabo (2002)는 개볼락을 등 (dorsum)에 있는 색을 가진 마킹의 유무와 등 가시 (dorsal spine) 아래에 있는 비늘의 유무 기준에 따라 2개의 종으로 구분할 수 있다고 보고하였다.

많은 생물에서 체색과 체색 패턴은 생존 및 번식기능과 연관된 중요한 표현형질로 (Protas and Patel, 2008; Rodgers *et al*., 2010), 체색 (coloration)은 주로 색소포 (chromatopore)나

Table 5. Top 20 up-regulated and down-regulated DEGs in the Color type skin

Gene ID	Gene name	Description	Fold changes
Up-regulated genes			
Bbrf14060	MYH7	Myosin-7	18.4
Bbrf24777	CUL1	Cullin-1	17.5
Bbrf28904	ZNF106	Zinc finger protein 106	17.4
Bbrf26698	FAM73A	Protein FAM73A	17.4
Bbrf30393	MYOT	Myotilin	17.3
Bbrf04388	TOMM7	Mitochondrial import receptor subunit TOM7 homolog	17.3
Bbrf27220	CTL	Galactose-specific lectin nattectin	17.0
Bbrf10316	VMA21	Vacuolar ATPase assembly integral membrane protein vma21	17.0
Bbrf29784	B2M	Beta-2-microglobulin	16.9
Bbrf17755	ISCA1	Iron-sulfur cluster assembly 1 homolog, mitochondrial	16.9
Bbrf28527	EMILIN1	EMILIN-1	16.5
Bbrf25183	CAC1S	Dihydropyridine-sensitive L-type skeletal muscle calcium channel subunit alpha-1	16.5
Bbrf19960	LGALS9	Galectin-9	16.5
Bbrf23658	CMAS	N-acylneuraminate cytidylyltransferase	16.3
Bbrf10202	WDR33	pre-mRNA 3' end processing protein WDR33	16.1
Bbrf26224	ASCC3	Activating signal cointegrator 1 complex subunit 3	15.9
Bbrf21247	ANXA6	Annexin A6	15.8
Bbrf28776	SAMHD1	Deoxynucleoside triphosphate triphosphohydrolase SAMHD1	15.7
Bbrf11844	MRP63	Ribosomal protein 63, mitochondrial	15.7
Bbrf28488	ACOT11	Acyl-coenzyme A thioesterase 11	15.4
Down-regulated genes			
Bbrf27579	NMRK2	Nicotinamide riboside kinase 2	−19.9
Bbrf26562	FLNC	Filamin-C	−17.5
Bbrf17534	ACADM	Medium-chain specific acyl-CoA dehydrogenase, mitochondrial	−17.1
Bbrf22079	GM129	Uncharacterized protein C1orf51 homolog	−16.9
Bbrf26028	SLC38A2	Sodium-coupled neutral amino acid transporter 2	−16.8
Bbrf20950	PCBP2	Poly(rC)-binding protein 2	−16.3
Bbrf73667	ETNK1	Ethanolamine kinase 1	−16.2
Bbrf23744	RNF213	E3 ubiquitin-protein ligase RNF213	−16.2
Bbrf13080	FOS	Proto-oncogene c-Fos	−16.2
Bbrf25540	BAG4	BAG family molecular chaperone regulator 4	−16.2
Bbrf28221	MURC	Muscle-related coiled-coil protein	−16.1
Bbrf25670	MRPL18	39S ribosomal protein L18, mitochondrial	−16.1
Bbrf08448	HRSP12	Ribonuclease UK114	−16.1
Bbrf21897	RBM12B	RNA-binding protein 12B	−16.1
Bbrf15845	SCEL	Sciellin	−16.0
Bbrf17325	UBE2L3	Ubiquitin-conjugating enzyme E2 L3	−16.0
Bbrf27072	SLC4A7	Sodium bicarbonate cotransporter 3	−16.0
Bbrf27793	DCAF11	DDB1- and CUL4-associated factor 11	−15.9
Bbrf26119	MST4	Serine/threonine-protein kinase MST4	−15.8
Bbrf21383	Ppan	Suppressor of SWI4 1 homolog	−15.5

색소세포(pigment cell)에 의해 합성되는 여러 가지 색상이 결정한다. 이들 색소세포의 다양화는 세포 인자, 영양소 인자, 생리적 요인, 유전적 요인 및 환경요인과 연관이 있고, 복잡한 생물학적인 과정을 거쳐 피부의 색상을 만든다(Jiang *et al*., 2014). 척추동물의 체색은 6 가지 색소세포에 의해 결정되는데(Braasch *et al*., 2007), 포유류 및 조류의 피부에서는 멜라닌 세포만 확인되고, 파충류는 황색소포, 적색소포, 홍색소포가 확인된 반면, 경골어류에서는 색소세포 6 가지 형태 모두가

확인된다(Mellgren and Johnson, 2002; Bagnara *et al*., 2007; Braasch *et al*., 2009).

어류의 체색 형성 관련 유전자 탐색 연구가 모델 어류인 제브라피쉬를 이용하여 진행되었고, *MATP*, *OCA4*, *SOX10*, *KIT*, *EDNRB*, *SLC24A5*를 비롯한 색소침착 과정에 관여하는 수십여 개의 유전자와 변이에 관한 내용이 밝혀졌다(Kelsh *et al*., 1996; Kelsh *et al*., 2004; Parichy, 2006). 또한 비모델 어류인 잉어(common carp)와 레드 틸라피아(Red tilapia)의 피부색 형성 관련 유전자 탐색 연구가 수행되었다. 잉어 전사체 분석 연구(Jiang *et al*., 2014)에서는 체색이 다른 두 품종에서 차등발현유전자를 확보하였고, 피부의 색소침착 과정에 관여하는 주요 유전자인 *TYPR1*, *SILV*, *ASIP* 및 *xCT* 유전자의 발현량을 비교하였다. 레드 틸라피아 전사체 분석 연구(Zhu *et al*., 2016)에서는 피부색 차이와 관련된 후보 유전자를 선별하고 유전육종에 적용할 수 있는 SSR 마커를 확보하였으며, *TYR, TYRP1, SILV, SOX10, SLC24A5, CBS* 및 *SLC7A11* 유전자의 발현량을 분석하였다. 어류 피부색 및 체색 패턴과 관련된 유전자를 발굴하고 활용하기 위한 연구가 진행되었으나, 어류의 다양한 체색 타입에 따른 상세한 분자특성 및 연관 기작에 관한 연구는 여전히 부족하다. 개볼락의 체색 패턴 형성 관련 유전적 특성 및 변이 등에 관한 연구로는 Kai *et al*.(2011)이 보고한 내용이 있다. 개볼락 몸 전체에 있는 반점 색상과 마킹 위치에 따른 체색 타입별 개체들을 대상으로 유전자 변이를 조사하였으나, 체색 타입과 관련된 유의미한 유전자 변이는 발견하지 못했다. 이외에 개볼락의 다양하고 복잡한 체색 패턴과 관련된 유전자 변이를 비롯한 유전 특성에 관한 연구 보고는 아직 없다.

RNA-seq은 전사물(transcript)의 전체에 해당하는 전사체(transcriptome)를 분석하여 발현의 차이를 확인하는 방법으로 NGS 플랫폼을 사용하며, 대용량 시퀀싱(high-throughput sequencing)을 통하여 발현되는 전사물의 동정과 정량을 동시에 수행할 수 있어 생명과학 연구의 표준 도구가 되었다. RNA-seq 분석법을 이용하여 연구대상 생물의 전사체 정보를 밝힌 연구 결과가 발표되었고(Wang *et al*., 2009; Ji *et al*., 2009; Liao *et al*., 2013), 연구대상 생물의 서로 다른 발달단계나 생리학적인 프로세스와 연관된 유의미한 발현유전자 및 중요한 경로에 대한 연구결과도 보고되었다(Samanta *et al*., 2006; Huan *et al*., 2012; Liao *et al*., 2013). 본 연구에서는 개볼락(*Sebastes pachycephalus*)의 다양한 체색 패턴과 관련된 유전자 탐색에 필요한 유전자 정보를 확보하기 위하여 반점이나 마킹이 전혀 없는 Wild type과 몸 전체에 붉은 반점이 있고 배 부분에 마킹이 있는 Color type으로 구분한 개볼락 시료의 피부 전사체를 RNA-seq 방법으로 분석하였다. RNA-seq 실험은 대용량 시퀀싱으로 얻은 read 개수가 체색 타입에 따라 특이적으로 발현하는 유전자 및 발현량 차이 확인이 가능한지를 고려하여 설계하였으며, 결과 분석 과정을 거쳐 개볼락 체색 타입별 피부에서 차등발현하는 유전자와 체색 타입 특이적으로 발현하는 유전자를 확보하였다.

RNA-seq 분석 연구를 성공하기 위한 중요한 전제조건은 대용량 시퀀싱 방법을 통하여 얻은 데이터로 연구목적에 부합하는 생물학적인 의문점을 설명할 수 있어야 하는데, 이것은 연구대상 생물에 적합한 라이브러리 형태, sequencing depth 및 반복실험 횟수를 고려한 실험 디자인을 통하여 가능하다. 또한 시퀀싱 실험 자체를 적절하게 수행하여 불필요한 편향(bias)이 없는 데이터를 얻어야 한다. 실험설계에서 중요한 요소인 sequencing depth (라이브러리를 얼마나 많이 읽을 것인가) 또는 라이브러리 크기는 각 샘플의 sequencing된 read의 수를 나타내는 것으로(Conesa *et al*., 2016), 샘플의 시퀀싱 수준을 크게 정하여 많이 읽으면 더 많은 transcript를 찾아낼 수 있고, 정량 결과의 정확도 또한 높지만(Mortazavi *et al*., 2008), 일반적으로 최적의 sequencing depth는 무조건 크게 정하지 않고, 실험의 목적에 따라 다르게 결정한다. 그 이유는 sequencing depth가 클수록 유전자 동정과 정량에는 유리하지만, transcriptional noise와 off-target transcript (표적 외 전사물)를 증가시키는 단점도 있기 때문이다(Tarazona *et al*., 2011). 일부 연구자는 진핵생물의 전사체에서 고도로 발현된 유전자의 정량은 500만 개의 mapped read로 충분하다고 주장하지만, 발현 수준이 낮은 유전자 및 전사물을 정확하게 정량하려면 최대 1억 개의 read가 필요하다는 다른 저자의 의견도 있다(Sims *et al*., 2014). 시료의 복잡성이 적은 단세포의 유전자 발현량 분석에는 보통 100만 read를 이용하는 것으로 실험을 설계하지만, 발현량이 많은 유전자는 5만 개 read로 정량이 가능했으며, 심지어는 2만 개 read 만으로도 비장 조직에서 세포 유형을 구별할 수 있었다(Jaitin *et al*., 2014).

체색 타입별 개볼락 피부의 전사체를 시퀀싱하여 얻은 염기서열 데이터는 FastQC로 quality control을 진행하여 clean read를 확보하였다. RNA-seq 데이터는 대용량 시퀀싱으로 raw read 획득, read 정렬 및 유전자 정량화 과정을 거쳐 확보하고, 각 단계마다 특정한 방법을 적용하여 데이터의 품질을 체크해야 한다. 먼저 raw read의 quality control은 sequencing error, PCR artifact 또는 contamination (라이브러리 섞임 현상)을 확인하기 위한 과정으로 염기서열의 품질, GC content, adaptor 유무, 지나치게 많은 k-mer 그리고 duplicated read 존재 여부를 검사한다. 전사체 데이터 분석에서 허용되는 duplication, k-mer, GC content 수준은 실험 및 연구대상 생물의 특성에 따라 다르지만, 이들 값은 동일 실험에서는 각 샘플별로 균일해야만 한다. 본 연구에서는 raw read의 quality control을 진행하여 PE clean read를 Wild type 피부에서 73,392,934개와 Color type 피부에서 73,454,304개를 각각 확보(Table 1)하였고, 체색 타입별 전사체 비교분석에 사용

이 가능한 것으로 판단하였다. 색상이 다른 잉어 전사체 연구(Jiang *et al*., 2014)에서는 2종의 피부 샘플을 시퀀싱하여 평균 33,376,990개의 PE clean read를 확보하여 분석에 사용하였으며, 레드 틸라피아 피부색 연구(Zhu *et al*., 2016)에서는 3가지 색상(WP; whole pink, PR; pink with scattered red spots, PB; pink with scattered black spots) 개체 각각의 피부 샘플을 시퀀싱하여 확보한 PE clean read 42,927,212개로 체색 타입별 전사체를 비교 분석하였다.

RNA-seq 분석은 표준염기서열(reference genome) 정보 이용이 가능한 경우에는 reference genome 또는 reference transcriptome에 맵핑하여 전사물을 식별한다. 연구목적이 주석이 없는(unannotated) 새로운 전사물 발굴이 아닌 정량 분석에 중점을 둔 경우에는 정보가 밝혀진 생물 종의 reference transcriptome에 맵핑한다. 이와 반대로 연구대상 생물의 게놈 서열 정보가 없는 경우에는 먼저 각 read를 더 긴 contig로 조립하여 transcriptome으로 만들고, 그 다음에는 정량을 위하여 각 read를 이 transcriptome에 다시 맵핑한다. 개볼락의 게놈 정보가 아직 밝혀지지 않아 본 연구에서는 *de novo* assembly 방식을 채택하였다. 체색 타입별 피부 샘플 간 비교 분석을 위하여 clean read를 합쳐서 하나의 input으로 만들었고, Trinity 프로그램을 이용하여 transcriptome을 조립하여 101,112개의 unigene을 확보하였다(Table 2). 잉어 피부색 연구(Jiang *et al*., 2014)에서는 transcriptome 조립 과정을 거쳐 198,781개의 unigene을 얻었다. 레드 틸라피아 피부색 연구(Zhu *et al*., 2016)에서는 실험어 1개체에서 9개 조직시료(뇌, 근육, 간, 창자, 심장, 신장, 난소, 비장, 피부)를 샘플링하였고, 이들 샘플을 pooling하여 만든 cDNA 라이브러리 1개와 3개 strain(WP, PB, PR)의 피부 cDNA 라이브러리를 각각 시퀀싱하여 얻은 read를 어셈블리하여 213,396개의 unigene을 확보하였다. 이와 같은 선행연구의 어류 피부 전사체 분석결과에서, RNA-seq 분석으로 얻은 read를 어셈블리하여 확보한 unigene의 개수는 시퀀싱에 사용한 cDNA 라이브러리의 개수에 따라 영향을 많이 받는다는 것을 알 수 있었다.

유전자의 기능분석을 위한 전사체 분석 연구는 차등발현 유전자의 분자생물학적 기능이나 관련 pathway를 찾기 위하여 수행하므로, RNA-seq 분석으로 확보한 전사체에 대하여 충분한 기능 주석 데이터(functional annotation data)가 반드시 있어야 한다. Gene Ontology (Ashburner *et al*., 2000), Bioconductor (Huber *et al*., 2015), DAVID (Huang *et al*., 2009), Babelomics (Medina *et al*., 2010) 등의 데이터베이스에는 대부분의 모델생물 종에 대한 주석 데이터가 포함되어 있다. 그러나, *de novo* transcriptome assembly 또는 재구성(reconstruction)을 통해 발견한 신규 전사물은 기능 정보가 부족하여 단백질 데이터베이스인 SwissProt (Bairoch *et al*., 2004)와 protein domain 데이터베이스인 InterPro (Hunter *et al*., 2011)를 이용하여 기능적인 주석(annotation)을 단다. 본 연구에서 확보한 unigene의 기능은 NCBI Nr (Non-redundant protein) 데이터베이스와 InterProScan tool을 이용하여 코딩 서열을 예측한 후, annotation을 진행하였다. 그 결과, BLASTn 이용으로 33,693개(33.3%)가 annotation 되었고, InterProScan을 이용하여 58,254개(57.6%)에 대한 CDS를 확인하였다(Fig. 2, Table 3). BLASTx와 InterProScan으로 annotation 되지 않은 65,630개 unigene은 Swiss-Prot, KEGG, COG 등의 데이터베이스를 추가 검색하면 annotation 될 것으로 판단하였으며, 향후 연구를 통하여 annotation을 완성하고자 한다. 레드 틸라피아 피부색 전사체 분석 연구(Zhu *et al*., 2016)의 경우, NCBI Nr을 비롯한 Swiss-Prot, KEGG, COG 등 4개의 데이터베이스를 검색하여 unigene의 91.75%를 function annotation 하였다.

개볼락 체색 타입별 특이발현유전자 및 차등발현유전자를 확보하기 위하여 전사체에 맵핑된 read의 수를 이용하여 전사체의 발현값을 측정하였다. RNA-seq 분석의 가장 일반적인 적용은 유전자 또는 전사체의 발현을 측정하는 것으로 transcript sequence에 맵핑된 read의 수가 기본값이다. Raw read의 수 그 자체는 transcript의 길이, 전체 read 수, sequencing bias가 고려되지 않아 발현량 비교에 부적합하므로 샘플 간 발현 차이를 비교하기 위해서는 정규화(normalization) 작업을 해야한다. RNA-seq 분석 후, 유전자 발현을 비교하기 위하여 다양한 방법으로 정규화하고, Cufflinks와 같은 프로그램을 이용하여 유전자 발현값을 측정하지만, 비교할 샘플의 특성 등에 따른 bias는 여전히 존재한다. 또한 유전자 길이 보정은 샘플 전체에서 동일한 유전자에 대한 발현량의 변화를 비교할 때는 필요 없지만, 샘플 내에서 유전자 발현을 비교할 때는 순위화(ranking)에 중요한 요소임을 확인해야 한다. Transcript-level을 정확하게 측정하기 위한 알고리즘과 bias를 보정하는 방법이 개발되고 있다(Conesa *et al*., 2016). 본 연구에서는 transcriptome mapping을 통해 발현을 측정하는 RSEM을 사용하였는데, RSEM은 expectation maximization 방법을 사용하여 TPM 값을 측정하는 알고리즘으로 multi-mapping read를 결과 중 transcript와 할당하고 샘플 내의 sequencing bias를 보정한다(Li and Dewey, 2011). 잉어 피부색의 유전적 변이 연구(Jiang *et al*., 2014)도 RSEM 알고리즘을 사용하였으며, 레드 틸라피아의 피부색 관련 전사체 분석연구(Zhu *et al*., 2016)에서는 Gene Transfer Format (GTF) 파일과 Cufflinks (V2.2.1) 프로그램을 사용하였다. Cufflinks (Trapnell *et al*., 2010)는 TopHat mapper (expectation-maximization 방식 사용)로 mapping한 결과를 바탕으로 transcript 발현값을 계산하는데, 이 방법을 통해 non-uniform read 분포로 인한 bias를 적절히 다룰 수 있다. Cufflinks는 PE read의 장점을 사용하기 위해 설계되었기에 GTF 정보를 이용하여 transcript를 확인하거나 데이터를 통해 새로이 transcript를 추론할 수도 있다. 본 연구에서는 개

볼락 체색 타입별 피부 샘플의 발현량을 FPKM 값을 활용하여 비교하였고, 체색 타입에 따라 차등발현하는 유전자 164개를 최종 확보하였다(Table 4). 레드 틸라피아의 피부색 관련 전사체 분석연구(Zhu *et al*., 2016)에서도 발현량을 FPKM 값을 활용하여 비교하였으며, 3가지 체색 타입 피부 전사체로부터 유의미한 148개의 차등발현유전자를 선발하였다. 발현양상 검증을 통하여 피부색소(pigment) 합성 관련 주요 유전자인 tyr (tyrosinases), tyrp1 (tyrosinase-related protein 1), silv (premelanosome protein, *pmel*/*silv*), sox10 (sex determining region Y-box 10), slc24a5 (solute carrier family 24 member 5), cbs (cystathionine beta synthase) 및 slc7a11 (solute carrier family 7 member 11)을 확보하였다. 잉어 피부색의 유전적 변이 연구(Jiang *et al*., 2014)에서는 체색 패턴이 다른 2개 strain (Yellow River carp, Xingguo red carp)에서 발현량 차이를 나타낸 tyrp1, silv, asip (agouti signaling protein) 및 xCT (cysteine/glutamate transporter)를 피부 색소형성에 관여하는 유전자로 확보하였다.

본 연구에서 확인한 164개의 체색 타입별 차등발현유전자(Differentially expressed gene, DEG)중 기능 정보가 알려진 79개에는 어류의 피부색소 합성에 관여하는 것으로 보고된 유전자나 멜라닌 생성의 주요 유전자로 알려진 tyr (tyrosinases)이나 tyrp1 (tyrosinase-related protein 1)에 해당하는 것은 확인되지 않았다. 그러나, Color type 피부 및 Wild type 피부 특이적으로 발현하는 유전자와 발현량이 체색 타입별로 15배 이상의 차이를 나타낸 다수의 유전자가 확인되어, 이들을 개볼락의 체색 패턴 형성 관련 발현유전자 분석을 위한 후보유전자로 선정하였다. 향후 연구에서는 이들 차등발현유전자의 발현양상 확인을 위하여 qRT-PCR 분석을 실시하는 등 추가 분석을 수행하고, 개볼락의 체색 형성과 관련된 핵심 유전자를 발굴하여 체색 타입별 유전자 변이 특성을 밝히고자 한다.

요 약

생물의 종 구분에 이용하는 지표 중 체색은 특징이 뚜렷한 형태 지표로서, 어류의 종 동정에 유용한 형태형질이다. 개볼락은 한국 중부와 남부, 일본 홋카이도 남쪽 등지에 분포하는 상업적으로 중요한 어종으로, 피부에 반점의 유무 및 마킹이 있는 위치에 따라 4개의 아종으로 구분하는 복잡한 체색 특성을 갖는다. 그러나 개볼락의 다양한 체색 패턴과 관련된 유전자 탐색 및 유전자 변이 발굴 등 체색 형성에 관여하는 유전자 규명에 관한 연구는 없다. 이에 따라 본 연구에서는 개볼락의 체색 패턴 관련 유전자 발굴 및 유전자 발현 특성을 규명하기 위한 기초 연구로 체색 타입별 피부 전사체를 프로파일링하였다. 개볼락을 Wild type (반점과 marking 없음)과 Color type (반점과 마킹 모두 있음)으로 구분하였고, 피부 전사체를 RNA-seq 방법을 이용하여 분석하였다. 개볼락 피부 전사체의 발현량을 비교하여 체색 타입별 차등발현유전자 164개를 확보하였다. 이들 차등발현유전자의 기능을 Gene ontology (GO) 분석으로 확인한 결과, 2개는 molecular function, 46개는 biological process, 6개는 cellular component 기능그룹에 속하였다. 차등발현유전자 중 CTL (Galactose-specific lectin nattectin), CUL1 (Cullin-1), CMAS (N-acylneuraminate cytidylyltransferase), NMRK2 (Nicotinamide riboside kinase 2), ALOXE3 (Hydroperoxide isomerase ALOXE3), SLC4A7 (Sodium bicarbonate cotransporter 3) 등은 특정 체색 타입 특이적인 발현양상을 나타냈다. 이번 연구는 개볼락의 체색 패턴 형성에 관여하는 전사체를 탐색한 첫 번째 연구로, 체색 형성 관련 기능유전자 발굴을 위한 후보유전자로 개볼락의 체색 타입별 차등발현유전자를 확보한 것에 의의가 있다. 향후에는 이들 후보유전자의 발현양상 및 기능을 분석하여 개볼락의 복잡한 체색 패턴과 관련된 기능유전자의 특성을 밝히고자 한다.

사 사

본 연구는 한국해양과학기술원 주요사업인 '한국 주변 해양생태계 변동 이해 및 대응 기반 연구(PE99813)'의 연구비 지원으로 수행되었습니다.

REFERENCES

Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin and G. Sherlock. 2000. Gene Ontology: tool for the unification of biology. Nat. Genet., 25: 25-29. https://doi.org/10.1038/75556.

Bagnara, J.T., P.J. Fernandez and R. Fujii. 2007. On the blue coloration of vertebrates. Pigment Cell Res., 20: 14-26. https://doi.org/10.1111/j.1600-0749.2006.00360.x.

Bairoch, A., B. Boeckmann, S. Ferro and E. Gasteiger. 2004. Swiss-Prot: juggling between evolution and stability. Brief Bioinformatics, 5: 39-55. https://doi.org/10.1093/bib/5.1.39.

Blanco, E., G. Parra and R. Guigó. 2007. Using geneid to identify genes. Curr. Protoc. Bioinformatics, 18: 4.3.1-4.3.28. https://doi.org/10.1002/0471250953.bi0403s18.

Braasch, I., M. Schart and J.N. Volff. 2007. Evolution of pigment synthesis pathways by gene and genome duplication in fish. BMC Evol. Biol., 7: 74. https://doi.org/10.1186/1471-2148-7-74.

Braasch, I., F. Brunet, J.N. Volff and M. Schart. 2009. Pigmentation pathway evolution after whole-genome duplication in fish. Genome Biol. Evol., 1: 479-493. https://doi.org/10.1093/gbe/evp050.

Conesa, A., P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M.W. Szcześniak, D.J. Gaffney, L.L. Elo, X. Zhang and A. Mortazavi. 2016. A survey of best practices for RNA-seq data analysis. Genome Biol., 17: 13. https://doi.org/10.1186/s13059-016-0881-8.

Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman and A. Regev. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol., 29: 644-652. https://doi.org/10.1038/nbt.1883.

Haas, B.J., A. Papanicolaou, M. Yassour, M. Grabher, P.D. Blood, J. Bowden, M.B. Couger, D. Eccles, B. Li, M. Lieber, M.D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C.N. Dewey, R. Henschel, R.D. LeDuc, N. Friedman and A. Regev. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc., 8: 1494-1512. https://doi.org/10.1038/nprot.2013.084.

Hoekstra, H.E. 2006. Genetics, development and evolution of adaptive pigmentation in vertebrates. Heredity, 97: 222-234. https://doi.org/10.1038/sj.hdy.6800861.

Huan, P., H. Wang and B. Liu. 2012. Transcriptomic analysis of the clam meretrix meretrix on different larval stages. Mar. Biotechnol., 14: 69-78. https://doi.org/10.1007/s10126-011-9389-0.

Huang, D.W., B.T. Sherman and R.A. Lempicki. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc., 4: 44-57. https://doi.org/10.1038/nprot.2008.211.

Hubbard, J.K., J.A.C. Uy, M.E. Hauber, H.E. Hoekstra and R.J. Safran. 2010. Vertebrate pigmentation: from underlying genes to adaptive function. Trends Genet., 26: 231-239. https://doi.org/10.1016/j.tig.2010.02.002.

Huber, W., V.J. Carey, R. Gentleman, S. Anders, M. Carlson, B.S. Carvalho, H.C. Bravo, S. Davis, L. Gatto, T. Girke, R. Gottardo, F. Hahne, K.D. Hansen, R.A. Irizarry, M. Lawrence, M.I. Love, J. MacDonald, V. Obenchain, A.K. Oleś, H. Pagès, A. Reyes, P. Shannon, G.K. Smyth, D. Tenenbaum, L. Waldron and M. Morgan. 2015. Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods, 12: 115-121. https://doi.org/10.1038/nmeth.3252.

Hunter, S., P. Jones, A. Mitchell, R. Apweiler, T.K. Attwood, A. Bateman, T. Bernard, D. Binns, P. Bork, S. Burge, E. de Castro, P. Coggill, M. Corbett, U. Das, L. Daugherty, L. Duquenne, R.D. Finn, M. Fraser, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, C. McMenamin, H. Mi, P. Mutowo-Muellenet, N. Mulder, D. Natale, C. Orengo, S. Pesseat, M. Punta, A.F. Quinn, C. Rivoire, A. Sangrador-Vegas, J.D. Selengut, C.J.A. Sigrist, M. Scheremetjew, J. Tate, M. Thimmajanarthanan, P.D. Thomas, C.H. Wu, C. Yeats and S.-Y. Yong. 2011. InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res., 40 (Database issue): D306-312. https://doi.org/10.1093/nar/gkr948.

Jaitin, D.A., E. Kenigsberg, H. Keren-Shaul, N. Elefant, F. Paul, I. Zaretsky, A. Mildner, N. Cohen, S. Jung, A. Tanay and I. Amit. 2015. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science, 343: 776-779. https://doi.org/10.1126/science.1247651.

Ji, P., G. Liu, J. Xu, X. Wang, J. Li, Z. Zhao, X. Zhang, Y. Zhang, P. Xu and X. Sun. 2012. Characterization of common carp transcriptome: sequencing, de novo assembly, annotation and comparative genomics. PLoS One, 7: e35152. https://doi.org/10.1371/journal.pone.0035152.

Jiang, Y., S. Zhang, J. Xu, J. Feng, S. Mahboob, K.A. Al-Ghanim, X. Sun and P. Xu. 2014. Comparative transcriptome analysis reveals the genetic basis of skin color variation in common carp. PLoS ONE, 9: e108200. https://doi.org/10.1371/journal.pone.0108200.

Kai, Y., K. Nakayama and T. Nakabo. 2011. Genetic and morphological divergence within the *Sebastes pachycephalus* complex (Scorpaeniformes: Scorpaenidae). Ichthyol. Res., 58: 333-343. https://doi.org/10.1007/s10228-011-0236-0.

Kai, Y. and T. Nakabo. 2013. Taxonomic review of the *Sebastes pachycephalus* complex (Scorpaeniformes: Scorpaenidae). Zootaxa, 3637: 541-560. https://doi.org/10.11646/zootaxa.3637.5.3.

Kelsh, R.N., M. Brand, Y.J. Jiang, C.P. Heisenberg, S. Lin, P. Haffter, J. Odenthal, M.C. Mullins, F.J. M. van Eeden, M. Furutani-Seiki, M. Granato, M. Hammerschmidt, D.A. Kane, R.M. Warga, D. Beuchle, L. Vogelsang and C. Nüsslein-Volhard. 1996. Zebrafish pigmentation mutations and the processes of neural crest development. Development, 123: 369-389. https://dev.biologists.org/content/develop/123/1/369.

Kelsh, R.N., C. Inoue, A. Momoi, H. Kondoh, M. Furutani-Seiki, K. Ozato and Y. Wakamatsu. 2004. The Tomita collection of medaka pigmentation mutants as a resource for understanding neural crest cell development. Mech. Dev., 121: 841-859. https://doi.org/10.1016/j.mod.2004.01.004.

Lamason, R.L., M-A.P.K. Mohideen, J.R. Mest, A.C. Wong, H.L. Norton, M.C. Aros, M.J. Jurynec, X. Mao, V.R. Humphreville, J.E. Humbert, S. Sinha, J.L. Moore, P. Jagadeeswaran, W. Zhao, G. Ning, I. Makalowska, P.M. McKeigue, D. O'Donnell, R. Kittles, E.J. Parra, N.J. Mangini, D.J. Grunwald, M.D. Shriver, V.A. Canfield, K.C. Cheng. 2005. SLC24A5, a putative cation exchanger, affects pigmenta-

tion in zebrafish and humans. Science, 310: 1782-1786. https://doi.org/10.1126/science.1116238.

Leclercq, E., J.F. Taylor and H. Migaud. 2010. Morphological skin colour changes in teleosts. Fish Fish., 11: 159-193. https://doi.org/10.1111/j.1467-2979.2009.00346.x.

Li, B. and C.N. Dewey. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics, 12: 323. https://doi.org/10.1186/1471-2105-12-323.

Liao, X., L. Cheng, P. Xu, G. Lu, M. Wachholtz, X. Sun and S. Chen. 2013. Transcriptome analysis of crucian carp (*Carassius auratus*), an important aquaculture and hypoxia-tolerant species. PLoS One, 8: e62308. https://doi.org/10.1371/journal.pone.0062308.

Matsubara, K. 1943. Studies on the scorpaenoid fishes of Japan. Anatomy, phylogeny and taxonomy II. Trans. Sigenkagaku Kenkyusyo, 2: 171-486.

Medina, I., J. Carbonell, L. Pulido, S.C. Madeira, S. Goetz, A. Conesa, J. Tárraga, A. Pascual-Montano, R. Nogales-Cadenas, J. Santoyo, F. García, M. Marbà, D. Montaner and J. Dopazo. 2010. Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. Nucleic Acids Res., 38: W210-213. https://doi.org/10.1093/nar/gkq388.

Mellgren, E.M. and S.L. Johnson. 2002. The evolution of morphological complexity in zebrafish stripes. Trends Genet., 18: 128-134. https://doi.org/10.1016/s0168-9525(01)02614-2.

Miller, C.T., S. Beleza, A.A. Pollen, D. Schluter, R.A. Kittles, M.D. Shriver and D.M. Kingsley. 2007. Cis-regulatory changes in *Kit ligand* expression and parallel evolution of pigmentation in sticklebacks and humans. Cell, 131: 1179-1189. https://doi.org/10.1016/j.cell.2007.10.055.

Mortazavi, A., B.A. Williams, K. McCue, L. Schaeffer and B. Wold. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods, 5: 1-8. https://doi.org/10.1038/nmeth.1226.

Nakabo, T. 2002. Scorpaenidae. In: Nakabo, T. (ed.), Fishes of Japan with pictorial keys to the species, English ed. Tokai Univ. Press, Tokyo, Japan, pp. 565-595, 1524-1528.

Parichy, D.M. 2006. Evolution of danio pigment pattern development. Heredity, 97: 200-210. https://doi.org/10.1038/sj.hdy.6800867.

Pertea, G., X. Huang, F. Liang, V. Antonescu, R. Sultana, S. Karamycheva, Y. Lee, J. White, F. Cheung, B. Parvizi, J. Tsai and J. Quackenbush. 2003. TIGR Gene indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics, 19: 651-652. https://doi.org/10.1093/bioinformatics/btg034.

Protas, M.E. and N.H. Patel. 2008. Evolution of coloration patterns. Annu. Rev. Cell Dev. Biol., 24: 425-446. https://doi.org/10.1146/annurev.cellbio.24.110707.175302.

Roberts, R.B., J.R. Ser and T.D. Kocher. 2009. Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. Science, 326: 998-1001. https://doi.org/10.1126/science.1174705.

Rodgers, G.M., J.L. Kelley and L.J. Morrell. 2010. Colour change and assortment in the western rainbowfish. Anim. Behav., 79: 1025-1030. https://doi.org/10.1016/j.anbehav.2010.01.017.

Rodionov, V.I., J. Yi, A. Kashina, A. Oladipo and S.P. Gross. 2003. Switching between microtubule- and actin-based transport systems in melanophores is controlled by cAMP levels. Curr. Biol., 13: 1837-1847. https://doi.org/10.1016/j.cub.2003.10.027.

Samanta, M.P., W. Tongprasit, S. Istrail, R.A. Cameron, Q. Tu, E.H. Davidson and V. Stolc. 2006. The transcriptome of the sea urchin embryo. Science, 314: 960-962. https://doi.org/10.1126/science.1131898.

Sims, D., I. Sudbery, N.E. Ilott, A. Heger, and C.P. Ponting. 2014. Sequencing depth and coverage: key considerations in genomic analyses. Nat. Rev. Genet., 15: 121-132. https://doi.org/10.1038/nrg3642.

Sun, J., T. Nishiyama, K. Shimizu and K. Kadota. 2013. TCC: an R package for comparing tag count data with robust normalization strategies. BMC Bioinformatics, 14: 219. https://doi.org/10.1186/1471-2105-14-219.

Tarazona, S., F. Garcia-Alcalde, J. Dopazo, A. Ferrer and A. Conesa. 2011. Differential expression in RNA-seq: a matter of depth. Genome Res., 21: 2213-2223. https://doi.org/10.1101/gr.124321.111.

Trapnell, C., B.A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M.J. van Baren, S.L. Salzberg, B.J. Wold and L. Pachter. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol., 28: 511-515. https://doi.org/doi.org/10.1038/nbt.1621.

Wang, C., M. Wachholtz, J. Wang, X. Liao and G. Lu. 2014. Analysis of the skin transcriptome in two oujiang color varieties of common carp. PLoS One, 9: e90074. https://doi.org/10.1371/journal.pone.0090074.

Yu, H.J. and J.K. Kim. 2014. New record of *Sebastes nudus* and redescription of *Sebastes pachycephalus* (Pisces: Scorpaenidae) from Korea. Fish Aquat. Sci., 17: 129-136. https://doi.org/10.5657/FAS.2014.0129.

Zhu, W., L. Wang, Z. Dong, X. Chen, F. Song, N. Liu, H. Yang and J. Fu. 2016. Comparative transcriptome analysis identifies candidate genes related to skin color differentiation in red tilapia. Sci. Rep., 6: 31347 https://doi.org/10.1038/srep31347.