Marine Genomics 28 (2016) 21–24

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen

Genomics/technical resources

Comparative transcriptome analysis of three color variants of the sea cucumber *Apostichopus japonicus*

Jihoon Jo [a,1], Jongsun Park [b,1], Hyun-Gwan Lee [c], Elizabeth M.A. Kern [d], Seongmin Cheon [a], Soyeong Jin [d], Joong-Ki Park [d], Sung-Jin Cho [e], Chungoo Park [a,*]

[a] School of Biological Sciences and Technology, Chonnam National University, Gwangju 500-757, Republic of Korea
[b] InfoBoss, Co. Ltd., 201 room, 134, Gangseo-ro 17-gil, Hwageok-dong, Gangseo-go, Seoul 07766, Republic of Korea
[c] School of Dentistry, Chonnam National University, Gwangju 500-757, Republic of Korea
[d] Division of EcoScience, Ewha Womans University, Seoul 120-750, Republic of Korea
[e] Department of Biology, College of Natural Sciences, Chungbuk National University, Cheongju, Chungbuk 361-763, Republic of Korea

ARTICLE INFO

Article history:
Received 22 February 2016
Received in revised form 21 March 2016
Accepted 22 March 2016
Available online 19 April 2016

Keywords:
Sea cucumber
Apostichopus japonicas
Body-color variations
Transcriptome
RNA-seq

ABSTRACT

The sea cucumber *Apostichopus japonicus* Selenka 1867 represents an important resource in biomedical research, traditional medicine, and the seafood industry. Much of the commercial value of *A. japonicus* is determined by dorsal/ventral color variation (red, green, and black), yet the taxonomic relationships between these color variants are not clearly understood. We performed the first comparative analysis of *de novo* assembled transcriptome data from three color variants of *A. japonicus*. Using the Illumina platform, we sequenced nearly 177,596,774 clean reads representing a total of 18.2 Gbp of sea cucumber transcriptome. A comparison of over 0.3 million transcript scaffolds against the Uniprot/Swiss-Prot database yielded 8513, 8602, and 8588 positive matches for green, red, and black body color transcriptomes, respectively. Using the Panther gene classification system, we assessed an extensive and diverse set of expressed genes in three color variants and found that (1) among the three color variants of *A. japonicus*, genes associated with RNA binding protein, oxidoreductase, nucleic acid binding, transferase, and KRAB box transcription factor were most commonly expressed; and (2) the main protein functional classes are differently regulated in all three color variants (extracellular matrix protein and phosphatase for green color, transporter and potassium channel for red color, and G-protein modulator and enzyme modulator for black color). This work will assist in the discovery and annotation of novel genes that play significant morphological and physiological roles in color variants of *A. japonicus*, and these sequence data will provide a useful set of resources for the rapidly growing sea cucumber aquaculture industry.

1. Introduction

Sea cucumbers represent commercially and medically valuable members of the group Echinodermata, which itself is one of the most abundant and ecologically successful marine invertebrate clades. Sea cucumbers can regenerate lost tissues and organs within a few months and are of particular interest in biomedical research. More than 1250 species of sea cucumber have been identified from the sea floor worldwide and approximately 20 of them are edible. One especially important species as a source of seafood and traditional medicine is the Pacific sea cucumber *Apostichopus japonicus* Selenka 1867, which is mainly found off the coasts of northeast Asia (including northern China, Korea, Japan, and far east Russia) (Choe and Oshima, 1961; Kanno and Kijima, 2002). This species exhibits three dorsal/ventral color variants (red, green, and black), which differ in their biological and morphological attributes (*e.g.*, shape of ossicle, habitat preference, spawning period, and polian vesicles) (Choe and Oshima, 1961; Hongsheng Yang, 2015).

The body color of *A. japonicus* is an important trait affecting the price and taste of its products, and the rare red color type is the most favored and expensive. Recently, global exploitation of sea cucumbers to meet consumer demand has generated rising conservation concern (Bordbar et al., 2011; Purcell, 2014). In order to manage this natural resource and to establish genetically based breeding systems for sea cucumber aquaculture, it is necessary to investigate genomic relationships among color variants. Unfortunately, there are so far no published genome data for any species of the Stichopodidae, and transcriptome data for all three color variants of *A. japonicus* are likewise unavailable. In this study we present a comprehensive analysis of *A. japonicus* color variant transcriptomes and provide a global view of potential molecular mechanisms that may be used to assist the rapidly growing sea cucumber aquaculture industry and to facilitate molecular selection in breeding programs.

* Corresponding author.
E-mail address: chungoo@jnu.ac.kr (C. Park).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.margen.2016.03.009
1874-7787/

Table 1
General feature and transcriptome sequencing project information of *Apostichopus japonicus* in accordance with MixS recommendation.

MIGS-id	Property	Red	Evidence code
	Classification	Domain Animalia	TAS Choe and Oshima (1961)
		Phylum Echinodermata	TAS Choe and Oshima (1961)
		Class Holothuroideaia	TAS Choe and Oshima (1961)
		Order Aspidochirotida	TAS Choe and Oshima (1961)
		Family Stichopodidae	TAS Choe and Oshima (1961)
		Genus Apostichopus	TAS Choe and Oshima (1961)
		Species *A. japonicus*	TAS Choe and Oshima (1961)
MIGS-ID 4	Geographic isolation	Republic of Korea/Geomun-do	NAS
MIGS-ID 4.1	Latitude	34°1′35″N	NAS
MIGS-ID 4.2	Longitude	127°18′45″E	NAS
MIGS-ID 4.3	Depth	Not recorded	IDA
MIGS-ID 5	Sample collection time	2015–03-27	NAS
MIGS-ID 6	Habitat	Marine habitat	NAS
MIGS-ID 6.1	Temperature	15.5 °C	TAS An et al. (2007)
MIGS-ID 8	Ploidy	Diploid	TAS Okumura et al. (2009)
MIGS-ID 11	Estimated size	1.2 Gb	IDA
MIGS-ID 29	Sequencing platform	Illumina Hiseq 2000	IDA
MIGS-ID 30	Assembler	Trinity	IDA
MIGS-ID 31	Finishing quality	Complete	IDA
MIGS-ID 32	Gene calling method	Stand-alone BLAST (version 2.2.13)	IDA
	BIOPROJECT ID	PRJEB12167	IDA

TAS: Traceable Author Statement.
NAS: Non-traceable Author Statement.
IDA: Inferred from Direct Assay.

2. Data description

2.1. Sample collection and RNA extraction

We obtained red, green, and black wild specimens of *A. japonicus* from the same geographic region, near the coast of Geomun-do in South Korea (34°1′35″N, 127°18′45″E) (Table 1). Total RNAs of sea cucumber specimens were extracted from the body wall of live tissues. About 5 g of each tissue were treated for 10 min at RT with RNAlater RNA Stabilization Reagent (Qiagen, cat. no. 76104). Tissues were frozen with liquid nitrogen and then crushed to a fine powder with a pestle and mortar. Total RNA was extracted from the powder with 10 ml of Tri-Reagent (Molecular Research Center, Inc., Cincinnati, OH, USA) following the manufacturer's protocol. The RNA pellets were dissolved in DEPC-treated water. The water with total RNAs was treated with DNase I (Sigma, cat. no. AMPD1) to eliminate gDNA contamination.

2.2. Transcriptome sequencing

After qualifying and quantifying the total RNA using a NanoDrop1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and Bioanalyzer 2100 (Agilent Technologies, Palo Alto CA, USA), 1 μg of total RNA was used with the TruSeq RNA library preparation kit (Illumina, San Diego, USA) according to the manufacturer's instruction manuals. The mRNA in total RNA was converted into a library of template molecules suitable for subsequent cluster generation using the reagents provided in the Illumina TruSeq RNA Sample Preparation Kit. Briefly, the mRNA was first purified from the total RNA using polyA selection, and then was chemically fragmented and converted into single-stranded cDNA using random hexamer priming. Next, the second strand was generated to create double-stranded cDNA ready for TruSeq library construction. The short ds-cDNA fragments were connected with sequencing adapters, and suitable fragments were separated by agarose gel electrophoresis. Finally, the TruSeq RNA libraries were built by PCR amplification, quantified using qPCR according to the qPCR Quantification Protocol Guide, and qualified using Agilent Technologies 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). We performed paired-end sequencing (2 × 100 bp) using the HiSeq 2000 platform (Illumina, San Diego, CA, USA).

2.3. De novo transcriptome assembly

After the sequencing run, the resulting raw reads in FASTQ format were processed using Trimmomatic (version 0.33) (Bolger et al., 2014) to obtain clean reads by removing those containing adapter sequences, poly-N sequences, or low quality bases (below a mean Phred score of 20). About 55 million clean reads were obtained for each library (Table 2). These high-quality reads were *de novo* assembled separately for each sample using Trinity (Grabherr et al., 2011) and SOAPdenovo-Trans (Li et al., 2010) which are two of the most popular, state-of-the-art de Bruijn graph assemblers. A de Bruijn graph is composed of nodes and edges, where nodes represent nucleotide sequences of fixed length *k* (*k*-mers) and edges connect nodes that overlap by *k*-1. Trinity was carried out with default *k*-mer length of 25. Because SOAPdenovo-Trans can be executed with different *k*-mer lengths, which is very important for obtaining high quality assembly transcripts, we ran 13 replicates for each *k* value ranging from 31 to 79 with 4 step intervals and determined a *k*-mer size of 47 as the optimal value in all three samples. Several parameters were taken into consideration when assessing the quality of two assemblies: number of transcripts produced, average length of transcripts, coverage, and N50 value (data not shown). Trinity assembler showed better performance in all three samples and resulting assemblies were chosen for this study (Table 2).

Table 2
Summary statistics for assembled scaffolds from the sea cucumber *Apostichopus japonicus* using the Trinity *de novo* assembly program.

Body color	Green	Red	Black
Number of reads sequenced	65,681,314	59,066,456	56,794,634
Number of reads processed	64,229,838	57,777,354	55,589,582
%[a]	97.79	97.82	97.88
Total assembled bases (bp)	93,981,462	98,863,699	90,886,316
Total trinity 'genes'	79,794	79,817	77,530
Total trinity 'transcripts'	102,995	101,841	98,871
Average length (bp)	912.49	970.77	919.24
GC content (%)	38.30	38.79	38.37
N10 (bp)	4076	4913	4130
N30 (bp)	2526	2918	2586
N50 (bp)	1595	1800	1608

Note. All statistics are based on all transcript contigs.
[a] % (Number of reads processed / Number of reads sequenced) * 100.

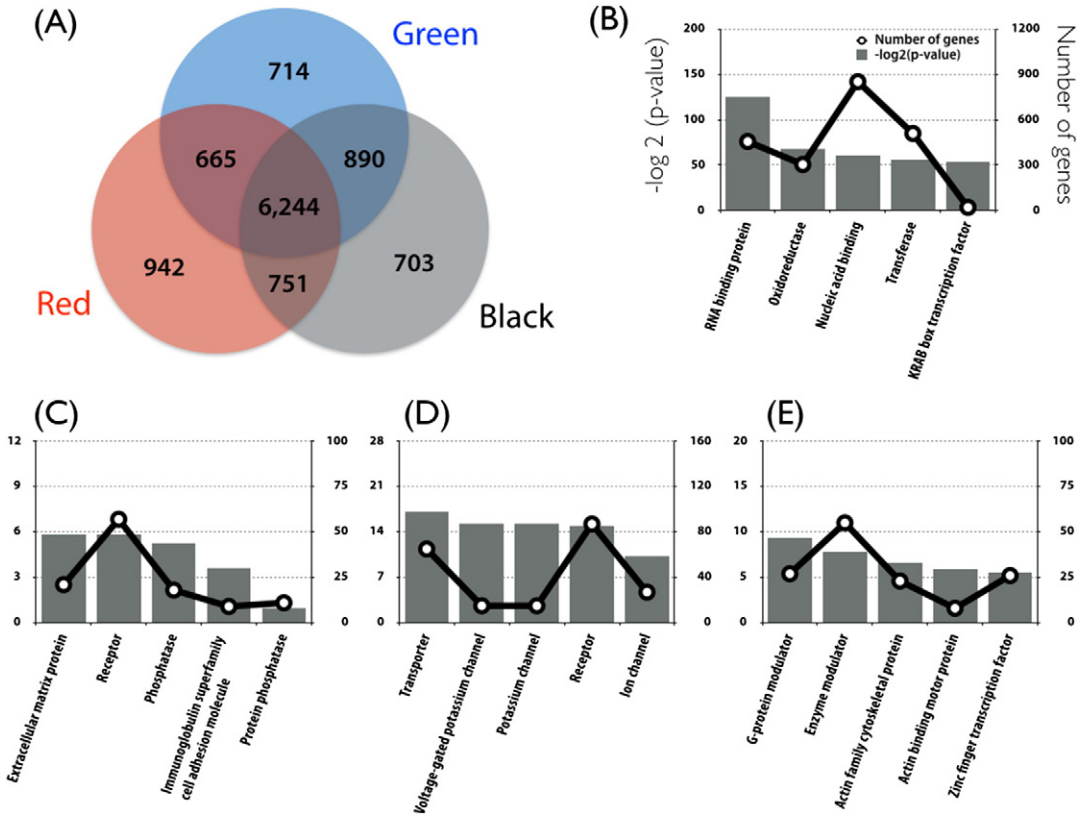

Fig. 1. Comparison of the transcriptome of three color variants in the sea cucumber *Apostichopus japonicus*. (A) Venn diagram showing the overlap in terms of numbers of genes expressed among different colored individuals. (B–E) Classification of expressed transcripts assigned to protein classes by PANTHER protein class ontology for common (B) and color-specific (green (C), red (D), and black (E)) genes in *A. japonicus*. The five most enriched proteins classes are shown. Bars represent transformed *p*-values. Broken lines represent the number of genes.

2.4. Functional annotation

Because so far there are no published genomes of any Stichopodidae species, we compared three sets of sea cucumber transcripts against the Uniprot/Swiss-Prot database (UniProt, 2015), which contains high quality, manually curated and reviewed proteins, using BLASTx implemented in stand-alone BLAST (version 2.2.13) (Altschul et al., 1990) with an *E*-value cutoff of 10^{-5}. All BLAST outputs were inspected and the best result was selected based on the lowest *E*-value. Following the above procedure, 8513, 8602, and 8588 unique putative transcripts were identified from green, red, and black color variants of *A. japonicus*. Among them, 6244 transcripts were commonly expressed in all three color variants while about one thousand genes were color specific (Fig. 1A). The target transcripts from the three color variants were classified into functional categories based on the Panther gene classification system (http://www.pantherdb.org) (Thomas et al., 2003). Commonly expressed transcripts among the three color variants were associated with the functional classes of RNA binding protein, oxidoreductase, nucleic acid binding, transferase, and KRAB box transcription factor (Fig. 1B). Intriguingly, the main protein functional classes are differently regulated in all three color variants (extracellular matrix protein and phosphatase for green color, transporter and potassium channel for red color, and G-protein modulator and enzyme modulator for black color) (Fig. 1C–E).

2.5. Data deposition

All raw sequence data were deposited in the European Nucleotide Archive (ENA) database under accession numbers PRJEB12167. The accession numbers of *de novo* assembled transcripts of *A. japonicus* range from HADD01000001 to HADD01101841, from HADE01000001 to HADE01102995, and from HADF01000001 to HADF01098871 for red, green, and black body colors, respectively.

3. Conclusion

In this paper, we present assembled transcriptomes from three color variants of *A. japonicus* and reveal their protein expression difference. Although whether this difference is strongly coupled with the body color still needs to be investigated, our transcriptome data will provide a comprehensive view of understanding genetic, physiological, and evolutionary relationships among color variants *A. japonicus*, and will be invaluable resources for the rapidly growing sea cucumber aquaculture industry.

Acknowledgments

This work was supported by research grants from the Next-Generation BioGreen 21 Program (CABX, no. PJ010001), Rural Development Administration, Republic of Korea to S.C., and from the Marine Biotechnology Program (PJT200620, Genome Analysis of Marine Organisms and Development of Functional Applications and Marine Mollusk Resource Bank of Korea) funded by the Ministry of Oceans and Fisheries of the Republic of Korea to C.P., and from the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2015R1C1A1A02036896) to C.P.

References

Choe, S., Oshima, Y., 1961. On the morphological and ecological differences between two commercial forms, "green" and "red", of the Japanese common sea cucumber, *Stichopus japonicus* Selenka. Nippon Suisan Gakkaishi 27, 97–105.

Kanno, M., Kijima, A., 2002. Quantitative and qualitative evaluation on the color variation of the Japanese sea cucumber *Stichopus japonicus*. Suisanzoshoku 50, 63–69.
Hongsheng Yang, J.-F.H., Mercier, Annie, 2015. The Sea Cucumber *Apostichopus japonicus*: History, Biology and Aquaculture. Academic Press.
Bordbar, S., Anwar, F., Saari, N., 2011. High-value components and bioactives from sea cucumbers for functional foods—a review. Mar. Drugs 9, 1761–1805.
Purcell, S.W., 2014. Value, market preferences and trade of Beche-de-mer from Pacific Island sea cucumbers. PLoS One 9, e95075.
Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.
Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652.
Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., Li, S., Yang, H., Wang, J., Wang, J., 2010. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. 20, 265–272.
UniProt, C., 2015. UniProt: a hub for protein information. Nucleic Acids Res. 43, D204–D212.
Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H.Y., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., Narechania, A., 2003. PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. 13, 2129–2141.
An, Z., Dong, Y., Dong, S., 2007. Temperature effects on growth-ration relationships of juvenile sea cucumber *Apostichopus japonicus* (Selenka). Aquaculture 272, 644–648.
Okumura, S.I., Kimura, K., Sakai, M., Waragaya, T., Furukawa, S., Takahashi, A., Yamamori, K., 2009. Chromosome number and telomere sequence mapping of the Japanese sea cucumber *Apostichopus japonicus*. Fish. Sci. 75, 249–251.