Marine Genomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

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

Method paper

Transcriptomics reveals tissue/organ-specific differences in gene expression in the starfish *Patiria pectinifera*

Chan-Hee Kim[a], Hye-Jin Go[a], Hye Young Oh[a], Yong Hun Jo[b], Maurice R. Elphick[c], Nam Gyu Park[a,*]

[a] Department of Biotechnology, College of Fisheries Sciences, Pukyong National University, 45 Yongso-ro, Nam-gu, Busan 48513, Republic of Korea
[b] Division of Plant Biotechnology, Institute of Environmentally-Friendly Agriculture (IEFA), College of Agriculture and Life Sciences, Chonnam National University, Gwangju 500-757, Republic of Korea
[c] School of Biological and Chemical Sciences, Queen Mary University of London, London E1 4NS, UK

ARTICLE INFO

Keywords:
Transcriptome
Multiple tissues
Differential expressed genes
Echinoderm
Starfish
Patiria pectinifera

ABSTRACT

Starfish (Phylum Echinodermata) are of interest from an evolutionary perspective because as deuterostomian invertebrates they occupy an "intermediate" phylogenetic position with respect to chordates (*e.g.* vertebrates) and protostomian invertebrates (*e.g. Drosophila*). Furthermore, starfish are model organisms for research on fertilization, embryonic development, innate immunity and tissue regeneration. However, large-scale molecular data for starfish tissues/organs are limited. To provide a comprehensive genetic resource for the starfish *Patiria pectinifera*, we report *de novo* transcriptome assemblies and global gene expression analysis for six *P. pectinifera* tissues/organs – body wall (BW), coelomic epithelium (CE), tube feet (TF), stomach (SM), pyloric caeca (PC) and gonad (GN). A total of 408 million high-quality reads obtained from six cDNA libraries were assembled *de novo* using Trinity, resulting in a total of 549,598 contigs with a mean length of 835 nucleotides (nt), an N50 of 1473 nt, and GC ratio of 42.5%. A total of 126,136 contigs (22.9%) were obtained as predicted open reading frames (ORFs) by TransDecoder, of which 102,187 were annotated with NCBI non-redundant (NR) hits, and 51,075 and 10,963 were annotated with Gene Ontology (GO) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) using the Blast2GO program, respectively. Gene expression analysis revealed that tissues/organs are grouped into three clusters: BW/CE/TF, SM/PC, and GN, which likely reflect functional relationships. 2408, 8560, 2687, 1727, 3321, and 2667 specifically expressed genes were identified for BW, GN, PC, CE, SM and TF, respectively, using the ROKU method. This study provides a valuable transcriptome resource and novel molecular insights into the functional biology of different tissues/organs in starfish as a model organism.

1. Introduction

Starfish are deuterostomian invertebrates belonging to the phylum Echinodermata that are recognized as fascinating animals with many features of special interest (Arnone et al., 2016). These include their status as a canonical example of a keystone species in ecology (Paine, 1966) and as model organisms for research on neuroendocrinology (Semmens et al., 2016), innate immunity (Franco et al., 2011), temporary adhesion (Hennebert et al., 2014), and tissue regeneration (Thorndyke et al., 2001).

Recently, transcriptome sequencing of emerging marine model organisms has proven to be an efficient method for relatively low cost gene discovery and analysis of differential gene expression (Martin and Wang, 2011). For these reasons, several research groups have utilized high throughput sequencing technology for molecular level characterization of various biological processes in several starfish species, including *Asterias amurensis* (Richardson and Sherman, 2015), *Asterias rubens* (Semmens et al., 2013; Semmens et al., 2016), *Acanthaster planci* (Stewart et al., 2015) and *Coscinasterias muricata* (Gabre et al., 2015). However, these transcriptomic studies were limited to analysis of whole larvae or just one adult starfish tissue/organ. Starfish have many different tissues and organs that are responsible for a variety of biological

Abbreviations: BW, body wall; CE, coelomic epithelium; TF, tube feet; SM, stomach; PC, pyloric caeca; GN, gonad; BUSCO, Benchmarking Universal Single-Copy Orthologs; KOG, EuKaryotic Orthologous Groups; ORFs, open reading frames; NR, non-redundant; GO, gene ontology; KEGG, Kyoto Encyclopaedia of Genes and Genomes; DEGs, differentially expressed genes; BP, biological process; CC, cellular component; MF, molecular function; FPKM, fragments per kilobase per million mapped reads; RSEM, RNA-Seq by expectation maximization; AIC, Akaike's information criterion; FDR, false discovery rate; PCA, principal component analysis

* Corresponding author.
E-mail address: ngpark@pknu.ac.kr (N.G. Park).

http://dx.doi.org/10.1016/j.margen.2017.08.011
Received 9 June 2017; Received in revised form 29 August 2017; Accepted 29 August 2017

processes. For example: 1. the mechanical state of the body-wall determines body stiffness and posture in starfish (Motokawa, 2011), 2. the coelomic epithelium is involved in wound healing, regeneration, and hematopoiesis (Gabre et al., 2015), 3. the stomach (pyloric and cardiac stomach) and pyloric caeca enable intake, digestion, absorption and storage of nutrients (Ferguson, 1964), 4. a multitude of tube feet enable locomotion, utilising secretion of adhesive materials for adhesion (Hennebert et al., 2011), and 5. gonads are, of course, essential for reproduction (Stewart et al., 2015). It is not known, however, if the physiological roles of different tissues/organs in starfish are reflected in their gene expression profiles. Therefore, it is of interest to comprehensively analyze differentially expressed genes (DEGs) in starfish tissues/organs.

The starfish species *Patiria pectinifera* is widely distributed in the northern Pacific Ocean and has been used as a model organism for studying different aspects of starfish physiology (Davydov et al., 1990; Kim et al., 2016; Mita et al., 2009) and is also of interest from economic and environmental perspectives (Jo et al., 2013; Popov et al., 2016). In this study, we characterise the *P. pectinifera* transcriptome by RNA-seq employing paired-end Illumina HiSeq™ 2500 sequencing technology and subsequent *de novo* assembly to generate a comprehensive set of reference contigs for gene discovery and for analysis of DEGs among six different tissues/organs. Our study provides a genetic resource for future comparisons with other echinoderm transcriptomes and for functional analysis of gene expression.

2. Data description

2.1. RNA isolation and illumina sequencing

Live specimens of the starfish species *Patiria pectinifera* (approximate diameter 8 cm) were collected at low tide from the coast of Cheongsapo of Busan, Korea (Table 1). Approval by the local institution/ethics committee was not required for this work because experimental work on starfish is not subject to regulation and *P. pectinifera* is not an endangered or protected species. Six tissues/organs, gonad (GN), pyloric caeca (PC), coelomic epithelium (CE), stomach (SM), tube feet (TF), and body-wall (BW, excluding CE) were dissected from six individual specimens of *P. pectinifera* and then total RNA was extracted using RNeasy Total RNA Isolation kit (Qiagen, USA) according to the manufacturer's instructions. The concentration, quality, and integrity of RNA preparations were determined using a NanoDrop-2000 spectrophotometer (Thermo, USA) and a Bioanalyzer 2100 (Agilent Technologies, USA). Then the RNA preparations were disrupted into short fragments. Double-stranded cDNA was synthesized with sequencing adapters using Illumina TruSeq™ RNA Library Prep Kit v2 (San Diego, CA, USA) following the manufacturer's instructions. Finally, six RNA-seq libraries were subjected to paired-end sequencing with a read length of 2 × 101 nucleotides on an Illumina HiSeq 2500 platform. Illumina HiSeq 2500 produced a total of 417,972,264 reads representing a total of 42,215,198,664 nucleotides from six tissues/organs, with the maximum number of reads (75,840,222) generated from the SM library and the minimum number of reads (63,360,640) generated from CE library (Table 1). The raw reads were deposited in the Sequencing Read Archive (SRA) of NCBI with accession numbers SRR5229423, SRR5229424, SRR5229425, SRR5229426, SRR5229427, SRR5229428 for TF, SM, CE, PC, GN and BW, respectively.

2.2. De novo assembly and functional annotation

After sequencing was completed, Illumina TruSeq adapter sequences, low-quality bases, and reads under minimum length were trimmed from the reads using CutAdapt v1.10 with -q 20, -m 30 parameters (Martin, 2011). Then, reads were filtered into clean reads. From these reads, contamination removal was performed by Bowtie2 v2.2.9 against the bacterial and ocean metagenome databases downloaded from NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria, ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria_DRAFT, ftp://ftp.ncbi.nlm.nih.

Table 1
MixS descriptors and statistics for the sequencing and *de novo* assembly of the transcriptome of the starfish *Patiria pectinifera*.

MixS descriptors

Item	Description	Item	Description
Investigation type	Eukaryote	Biome	ENVO:00002149
Project name	PRJNA371229	Feature	ENVO:01000687
Lat_lon	35.16°N, 129.19°E	Material	ENVO:00002019
Geo_loc_name	South Korea: Cheongsapo	Temp.	25 °C
Collection_date	2015-05-12T14:00 + 02:00	Seq_meth	Illumina HiSeq 2500

Sequencing (Illumina HiSeq2500; paired-end, 2 × 101) stats

Process		Tissues: Body-wall	Coelomic epithelium	Gonad	Pyloric caeca	Stomach	Tube feet	Total
Raw read	Number	68,676,738	63,360,640	69,973,024	68,743,218	75,840,222	71,378,422	417,972,264
	Total size (bp)	6,936,350,538	6,399,424,640	7,067,275,424	6,943,065,018	7,659,862,422	7,209,220,622	42,215,198,664
Adapter trimming	Number, [a](%)	67,107,554 (97.7)	61,920,652 (97.7)	68,288,434 (97.6)	67,210,592 (97.8)	74,127,038 (97.7)	69,813,734 (97.8)	408,468,004 (97.7)
Contamination removal	Number, [b](%)	67,068,478 (97.7)	61,888,058 (97.7)	68,209,202 (97.5)	67,180,392 (97.7)	74,104,076 (97.7)	69,717,310 (97.7)	408,167,516 (97.7)

[a](%), (number of adapter trimmed read/number of raw read) × 100
[b](%), (number of contamination removed read /number of adapter trimmed read) × 100

Assembly (*De novo* assembly; Trinity 2.0.6) stats

Process	Number of contigs	Total size (bp)	Mean size (bp)	Minimum size (bp)	Maximum size (bp)	N50 (bp)	GC, %	Data accessibility DDBJ/EMBL/GenBank
De novo	549,598	459,064,184	835	224	35,675	1473	42.5	GFOQ00000000