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], and Nam Gyu Park[a,†]

[a]Department of Biotechnology, College of Fisheries Sciences, Pukyong National University, 45 Yongso-ro, Nam-gu, Busan 48513, Korea

[b]Division of Plant Biotechnology, Institute of Environmentally-Friendly Agriculture (IEFA), College of Agriculture and Life Sciences, Chonnam National University, Gwangju 500-757, Korea

[c]School of Biological and Chemical Sciences, Queen Mary University of London, London, E1 4NS UK

[†]Corresponding author: Nam Gyu Park, Department of Biotechnology, College of Fisheries Sciences, Pukyong National University, 45 Yongso-ro, Nam-gu, Busan 48513, Korea, Telephone: +82 51 6295867, Fax: +82 51 6295863, e-mail: ngpark@pknu.ac.kr

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 549,625 contigs with a mean length of 835 nucleotides (nt), an N50 of 1,473 nt, and GC ratio of 42.52%. 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. 2,408, 1,727, 2,667, 3,321, 2,687, and 8,560 specifically expressed genes were identified for BW, CE, TF, SM, PC and GN, 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.

Keywords: transcriptome, multiple tissues, differential expressed genes, echinoderm, starfish, *Patiria pectinifera*

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) and tissue regeneration (Thorndyke et al., 2001).

Recently, transcriptome sequencing of emerging model marine 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 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 and is also of interest from economic and environmental perspectives (Kim et al., 2016; Mita et al., 2009). 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.gov/genomes/all/GCA/000/204/965/GCA_000204965.1_ASM204 96v1). A total of 408,167,516 reads (97.66%) were obtained and were subjected to *de novo* assembly. The assembly produced a total of 549,625 contigs of a total length of 459,072,158 nt, with a mean length of 835 nt, an N50 of 1,473 nt, and GC ratio of 42.52% in a size range of 224 to 35,675 nt, using

Trinity Assembler (v2.0.6) (Grabherr et al., 2011), which is optimized to present the best *de novo* results compared to a variety of assemblies with paired-end reads. This Transcriptome Shotgun Assembly (TSA) project has been deposited at DDBJ/EMBL/GenBank under the accession GFOQ00000000. The version described in this paper is the first version, GFOQ01000000. Subsequently, the output from the assembler was processed to predict open reading frames (ORFs) using TransDecoder v3.0.0, generating a total of 126,136 ORFs with a total length of 145,293,354 nt, a mean length of 1,151 nt, an N50 of 1,569 nt, and GC ratio of 51.272% in a size range of 297 to 33,894 nt. The predicted ORFs were used in subsequent stages of annotation and analysis, and are available in the supplementary material (Appendix A).

To identify the function of predicted genes, all the predicted ORFs were aligned against the NCBI non-redundant (NR) databases using a cut off e-value of $< 10^{-3}$, returning 102,187 hits (supplementary material 1). The top 10 species with hits are shown in Fig. 1A. Most of the ORFs aligned to the genome sequence of the sea urchin *Strongylocentrotus purpuratus* (48.08%), followed by the acorn worm *Saccoglossus kowalevskii* (15.91%), and the lancelet *Branchiostoma floridae* (7.302%). This result is as expected because as an echinoderm *P. pectinifera* is a deuterostomian invertebrate, a superphylum that includes two chordate subphyla that are closely related to vertebrates (Urochordata and Cephalochordata) and the Ambulacraria (Hemichordata and Echinodermata) (Adoutte et al., 2000). To further investigate the biological functions of the predicted ORFs they were annotated to the Gene Ontology (GO) database for biological process (BP), cellular component (CC), and molecular function (MF) and to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database for enzyme commission (EC) number assignment via mapping to the KEGG pathway map, using Blast2Go program (Conesa et al., 2005). GO classification revealed that 51,075 (40.49%) annotated ORFs were categorized into 30 functional groups, and 10,963 (8.69%) transcripts had significant matches in the KEGG pathways (Fig. 1B and C and supplementary material 1).

2.3. Differentially expressed genes (DEGs) analysis

The expression values of all samples were estimated to a fragments per kilobase per million mapped reads (FPKM) value by RSEM (Li and Dewey, 2011) and were further normalized by the quantile normalization method with the R-package preprocessCore. A total of 91,124 transcripts, approximately three-fourths of the predictied ORFs in the transcriptome of *P. pectinifera*, were expressed in at least one tissue/organ based on the following criteria (FPKM $\geq$ 0.3 and Expected read count $\geq$ 5), and 23,412 of the expressed transcripts were present in all tissues/organs (supplementary material 1). The tissue-specific transcripts (i.e. significantly higher expression in a single tissue relative to all tissues) were defined using the ROKU method based on the Shannon entropy and

outlier detection to scan expression profiles for ranking tissue-specific genes using Akaike's information criterion (AIC) procedure (Kadota et al., 2006). The upper-limit option specifying the maximum 25 percentage of tissue transcripts as outliers to each gene was set as 0.25. A total of 21,370 tissue/organ-specific transcripts were detected as follows: BW (2,408), GN (8,560), PC (2,687), CE (1,727), SM (3,321), and TF (2,667) (Fig. 2A and supplementary material 2). Differentially expressed genes (DEGs) were identified by pairwise comparisons between two tissues/organs using the edgeR program (Robinson et al., 2010). For the non-replicated samples, we set the biological coefficient of variation (BCV) parameter as 0.4 and the parameters for DEGs were set as a log fold change (logFC ≥ 4) and a false discovery rate (FDR ≤ 0.01). The highest number of DEGs (27,840) was detected in the GN vs. PC comparison, representing 17,014 of up-regulated genes in GN and 10,826 of up-regulated genes in PC, and the lowest number of DEGs (12,385) was detected in the BW vs. CE comparison, representing 6,010 of upregulated genes in BW and 6,375 of up-regulated genes in CE. The DEGs between different tissue/organ samples are summarized in Fig. 2B and details are shown in supplementary material 3. To assess transcriptome similarity between tissues/organs, principal component analysis (PCA) and hierarchical clustering were performed (Fig. 2C, D and E). Both analyses revealed three discrete groupings reflecting their biological functions: the body wall (BW) and its associated tissues/organs (CE and TF) constituting epidermis, organs involved in feeding and digestion (PC and SM), and an organ (GN) involved in reproductive functions formed a separate, stand-alone cluster.

2.4. Conclusions

Although several studies have reported transcriptomic analysis of starfish species, there have been no studies that investigated differentially expressed genes in different starfish tissues/organs. In this study, RNA-seq employing paired-end Illumina HiSeq™ 2500 sequencing technology was used to generate transcriptome data and profile global gene expression in six tissues/organs from the starfish *P. pectinifera*. Our results revealed that numerous genes are differentially expressed in the six tissues/organs, which group together into three clusters in accordance with biological functions. These data provide a basis for identification of gene networks associated with specific biological functions in adult starfish tissues/organs. Therefore, the *P. pectinifera* transcriptome data obtained in this study provide a valuable resource for future research on many aspects of starfish biology, including gene/protein evolution, neuroendocrinology, innate immunity, tissue regeneration, and ecology.

Acknowledgments

This research was supported by Korea Ministry of Environment (MOE) as "Eco-innovation Program (201300030002) and co-supported by Basic Science Research Program through the National

173 Research Foundation of Korea (NRF) funded by the Ministry of Education
174 (2016R1D1A3B03934086).

175 The authors declare that they have no conflicts of interest with the contents of this article.

References

Adoutte, A., Balavoine, G., Lartillot, N., Lespinet, O., Prud'homme, B., de Rosa, R. 2000. The new animal phylogeny: reliability and implications. Proc. Nat. Acad. Sci. USA 97, 4453-4456.

Arnone, M.I., Andrikou, C., Annunziata, R. 2016. Echinoderm systems for gene regulatory studies in evolution and development. Curr. Opin. Genet. Dev. 39, 129-137.

Audic, S., Claverie, J.-M. 1997. The significance of digital gene expression profiles. Genome Res. 7, 986-995.

Benjamini, Y., Hochberg, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Series B Stat. Methodol. 289-300.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics (Oxford, England) 21, 3674-3676.

Ferguson, J.C. 1964. Nutrient transport in starfish. II. Uptake of nutrients by isolated organs. Biol. Bull. 126, 391-406.

Franco, C.F., Santos, R., Coelho, A.V. 2011. Proteome characterization of sea star coelomocytes–The innate immune effector cells of echinoderms. Proteomics 11, 3587-3592.

Gabre, J.L., Martinez, P., Nilsson Skold, H., Ortega-Martinez, O., Abril, J.F. 2015. The coelomic epithelium transcriptome from a clonal sea star, *Coscinasterias muricata*. Mar. Genomics 24 (3), 245-248.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644-652.

Hennebert, E., Wattiez, R., Flammang, P. 2011. Characterisation of the carbohydrate fraction of the temporary adhesive secreted by the tube feet of the sea star *Asterias rubens*. Mar. Biotechnol. 13, 484-495.

Kadota, K., Ye, J., Nakai, Y., Terada, T., Shimizu, K. 2006. ROKU: a novel method for identification of tissue-specific genes. BMC Bioinformatics 7, 294.

Kim, C.H., Kim, E.J., Go, H.J., Oh, H.Y., Lin, M., Elphick, M.R., Park, N.G. 2016. Identification of a novel starfish neuropeptide that acts as a muscle relaxant. J. Neurochem. 137, 33-45.

Li, B., Dewey, C.N. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12, 323.

Martin, J.A., Wang, Z. 2011. Next-generation transcriptome assembly. Nat. Rev. Genet. 12, 671-682.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17, pp. 10-12.

Mita, M., Yoshikuni, M., Ohno, K., Shibata, Y., Paul-Prasanth, B., Pitchayawasin, S., Isobe, M., Nagahama, Y. 2009. A relaxin-like peptide purified from radial nerves induces oocyte maturation and ovulation in the starfish, *Asterina pectinifera*. Proc. Nat. Acad. Sci. USA 106, 9507-9512.

Motokawa, T. 2011. Mechanical mutability in connective tissue of starfish body wall. Biol. Bull. 221, 280-289.

Paine, R.T. 1966. Food web complexity and species diversity. Am. Nat. 100, 65-75.

Richardson, M.F., Sherman, C.D. 2015. De Novo Assembly and Characterization of the invasive northern pacific seastar transcriptome. PLoS one 10, e0142003.

Robinson, M.D., McCarthy, D.J., Smyth, G.K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139-140.

Semmens, D.C., Dane, R.E., Pancholi, M.R., Slade, S.E., Scrivens, J.H., Elphick, M.R. 2013. Discovery

221 of a novel neurophysin-associated neuropeptide that triggers cardiac stomach contraction and
222 retraction in starfish. J. Exp. Biol. 216, 4047-4053.
223 Semmens, D.C., Mirabeau, O., Moghul, I., Pancholi, M.R., Wurm, Y., Elphick, M.R. 2016.
224 Transcriptomic identification of starfish neuropeptide precursors yields new insights into
225 neuropeptide evolution. Open Biol. 6, 150224.
226 Stewart, M.J., Stewart, P., Rivera-Posada, J. 2015. De novo assembly of the transcriptome of
227 *Acanthaster planci* testes. Mol. Ecol. Resour. 15, 953-966.
228 Thorndyke, M.C., Chen, W.C., Beesley, P.W., Patruno, M. 2001. Molecular approach to echinoderm
229 regeneration. Microsc. Res. Tech. 55, 474-485.
230 Yu, X., Lin, J., Zack, D.J., Qian, J. 2006. Computational analysis of tissue-specific combinatorial gene
231 regulation: predicting interaction between transcription factors in human tissues. Nucleic acids Res.
232 34, 4925-4936.

Figure legends

Fig. 1 Annotation of the *P. pectinifera* transcriptome to NCBI NR, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Ontology (GO) database. A, Distribution of top 10 species with hits with a cut off e-value of $< 10^{-3}$. B, Top 10 GO terms for each of the GO categories: molecular function (MF), biological process (BP), and cellular component (CC). C, Top 10 processes in KEGG pathways with hits using enzyme commission (EC) annotations.

Fig. 2 Statistics of differentially expressed genes (DEGs) in tissues/organs of *P. pectinifera*. A, The number of tissue specific genes in six tissues/organs in *P. pectinifera*. B, The number of differentially expressed genes between different tissue/organ samples, with specific criteria ($\log2FC \geq 4$ and $FDR < 0.01$). C, Principal component analysis based on all expressed genes showing three distinct groups of tissues. D, Cluster dendrogram between tissues/organs with AU/BP values (%) using Ward's method. E, Correlation heatmap of the transcriptome between tissues/organ of *P. pectinifera*. BW, body-wall; GN, gonad; PC, pyloric caeca; CE, coelomic epithelium; SM, stomach; and TF, tube feet.

248 Table legend

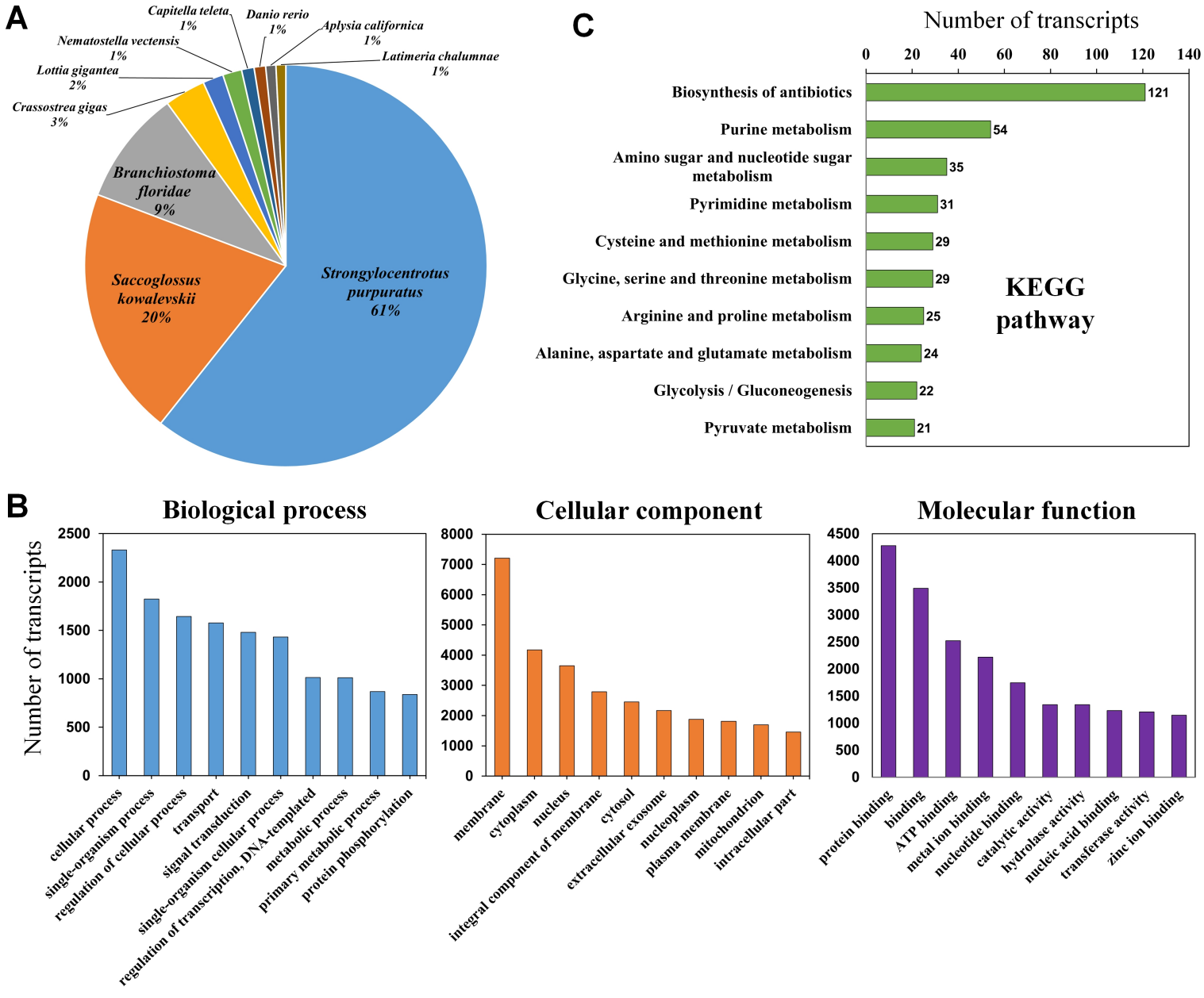
249 **Table 1** MixS descriptors and statistics for the sequencing and *de novo* assembly of the starfish
250 *Patiria pectinifera* transcriptome

Appendix A. Supplementary data legends

Supplementary material 1 List of the predicted open reading frames (ORFs) from the *de novo* assembled transcriptome of *P. pectinifera,* including the results of annotations to NCBI NR, Gene Ontology (GO), and Kyoto Encyclopedia of Genes and Genomes (KEGG) database.

Supplementary material 2 List of tissue-specific genes including the results of BLAST, GO, and KEGG annotations.

Supplementary material 3 Differentially expressed genes (DEGs) in *P. pectinifera* transcriptome, based on comparison of six tissues/organs


A
Capitella teleta 1%
Danio rerio 1%
Aplysia californica 1%
Nematostella vectensis 1%
Latimeria chalumnae 1%
Lottia gigantea 2%
Crassostrea gigas 3%
Branchiostoma floridae 9%
Saccoglossus kowalevskii 20%
Strongylocentrotus purpuratus 61%
C
Number of transcripts
0 20 40 60 80 100 120 140
Biosynthesis of antibiotics 121
Purine metabolism 54
Amino sugar and nucleotide sugar metabolism 35
Pyrimidine metabolism 31
Cysteine and methionine metabolism 29
Glycine, serine and threonine metabolism 29
Arginine and proline metabolism 25
Alanine, aspartate and glutamate metabolism 24
Glycolysis / Gluconeogenesis 22
Pyruvate metabolism 21
KEGG pathway
B
Biological process
Number of transcripts
cellular process
single-organism process
regulation of cellular process
transport
signal transduction
single-organism cellular process
regulation of transcription, DNA-templated
metabolic process
primary metabolic process
protein phosphorylation
Cellular component
membrane
cytoplasm
nucleus
integral component of membrane
cytosol
extracellular exosome
nucleoplasm
plasma membrane
mitochondrion
intracellular part
Molecular function
protein binding
binding
ATP binding
metal ion binding
nucleotide binding
catalytic activity
hydrolase activity
nucleic acid binding
transferase activity
zinc ion binding

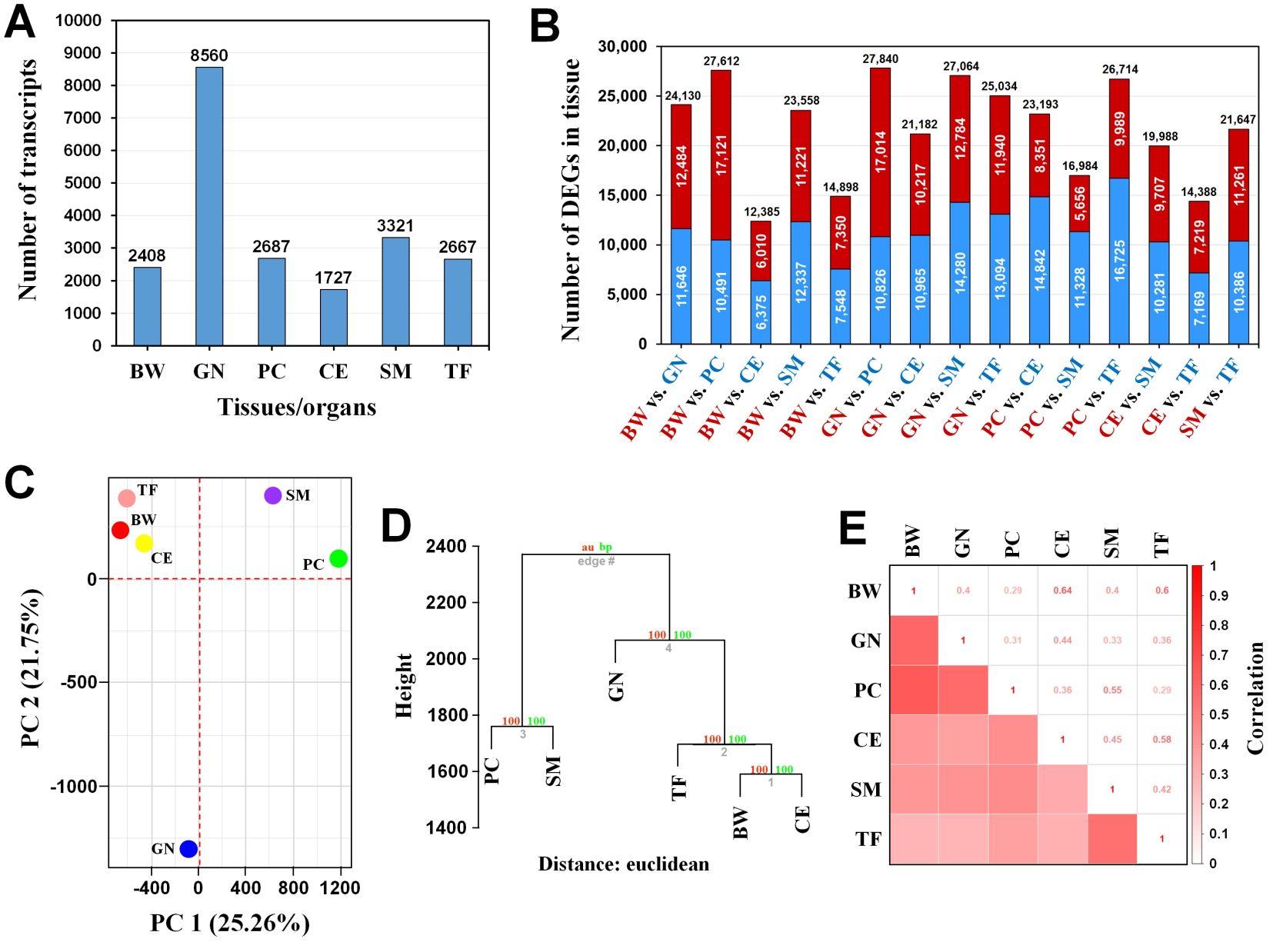

A
Number of transcripts
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
2408
8560
2687
1727
3321
2667
BW
GN
PC
CE
SM
TF
Tissues/organs
B
Number of DEGs in tissue
30,000
25,000
20,000
15,000
10,000
5,000
0
24,130
12,484
11,646
27,612
17,121
10,491
12,385
6,010
6,375
23,558
11,221
12,337
14,898
7,350
7,548
27,840
17,014
10,826
21,182
10,217
10,965
27,064
12,784
14,280
25,034
11,940
13,094
23,193
8,351
14,842
16,984
5,656
11,328
26,714
9,989
16,725
19,988
9,707
10,261
14,388
7,219
7,169
21,647
11,261
10,386
BW vs. GN
BW vs. PC
BW vs. CE
BW vs. SM
BW vs. TF
GN vs. PC
GN vs. CE
GN vs. SM
GN vs. TF
PC vs. CE
PC vs. SM
PC vs. TF
CE vs. SM
CE vs. TF
SM vs. TF
C
PC 2 (21.75%)
0
-500
-1000
TF
BW
CE
SM
PC
GN
-400
0
400
800
1200
PC 1 (25.26%)
D
Height
2400
2200
2000
1800
1600
1400
au bp
edge #
100 100
4
100 100
3
100 100
2
100 100
1
PC
SM
GN
TF
BW
CE
Distance: euclidean
E
BW
GN
PC
CE
SM
TF
1 0.4 0.29 0.64 0.4 0.6
1 0.31 0.44 0.33 0.36
1 0.36 0.55 0.29
1 0.45 0.58
1 0.42
1
Correlation
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

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 (97.6)	68,288,434 (97.8)	67,210,592 (97.7)	74,127,038 (97.7)	69,813,734 (97.8682092 02)	408,468,004 (97.7)
Contamination removal	Number, [b](%)	67,068,478 (97.7)	61,888,058 (97.7)	68,209,202 (97.7)	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	549,625	459,072,158	835	224	35,675	1,473	42.5	