Marine Genomics 28 (2016) 37–39

Contents lists available at ScienceDirect

Marine Genomics

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

Genomics/technical resources

De novo assembly and annotation of the Antarctic copepod (*Tigriopus kingsejongensis*) transcriptome

Hui-Su Kim [a,1], Bo-Young Lee [a,1], Jeonghoon Han [a], Young Hwan Lee [a], Gi-Sik Min [b], Sanghee Kim [c,*], Jae-Seong Lee [a,*]

[a] Department of Biological Science, College of Science, Sungkyunkwan University, Suwon 16419, South Korea
[b] Department of Biological Sciences, College of Natural Sciences, Inha University, Incheon 22212, South Korea
[c] Division of Life Sciences, Korea Polar Research Institute, Incheon 21990, South Korea

ARTICLE INFO

Article history:
Received 29 March 2016
Received in revised form 19 April 2016
Accepted 19 April 2016
Available online 4 May 2016

Keywords:
Transcriptome
RNA-seq
Antarctic
Copepod
Tigriopus kingsejongensis
Lipid metabolism

ABSTRACT

The whole transcriptome of the Antarctic copepod (*Tigriopus kingsejongensis*) was sequenced using Illumina RNA-seq. De novo assembly was performed with 64,785,098 raw reads using Trinity, which assembled into 81,653 contigs. TransDecoder found 38,250 candidate coding contigs which showed homology to other species by BLAST analysis. Functional gene annotation was performed by Gene Ontology (GO), InterProScan, and KEGG pathway analyses. Finally, we identified a number of expressed gene catalog for *T. kingsejongensis* that is a useful model animal for gene information-based polar research to uncover molecular mechanisms of environmental adaptation on harsh environments. In particular, we observed highly developing lipid metabolism in *T. kingsejongensis* directly compared to those of the Far East Pacific coast copepod *Tigriopus japonicus* at the transcriptome level.

1. Introduction

Of crustaceans, copepods are widely distributed and play an important role in the meiobenthic food web as an essential linker between the phytoplankton and higher trophic levels. To date the potential of using copepods as model organisms has been highlighted in diverse research areas (e.g. aquaculture, ecology, environmental biology, toxicology, and genomics). The copepod genus *Tigriopus* (family Harpacticidae) comprising of four well studied model species such as *Tigriopus brevicornis* (northern Europe), *Tigriopus californicus* (western America), *Tigriopus fulvus* (southern Europe), and *Tigriopus japonicus* (western Pacific) has been recognized as an intertidal model animal over the past three decades (Raisuddin et al., 2007). Particularly, *Tigriopus* are widely distributed in tidal pools all over the world due to their flexible physiology in response to dynamic environmental changes (e.g. salinity, temperature). In the genus *Tigriopus*, a species from the Kerguelen Island (latitude 49° S, longitude 69° E) of the southern Indian Ocean was previously identified (Soyer et al., 1987) and was annotated as *Tigriopus kerguelensis*. Recently, a new species of *Tigriopus* was identified from the King George Island of Antarctica and was registered with phenotypic characters as *Tigriopus kingsejongensis* (Park et al., 2014). In this manuscript, we analyzed transcriptome information of the Antarctic copepod *T. kingsejongensis* to provide public and practical use for genomic resource-based ecological and environmental studies.

2. Data description

2.1. Copepod culture

The Antarctic copepod *T. kingsejongensis* were kindly provided by Dr. Sanghee Kim (Korea Polar Research Institute, KIOST, Incheon, South Korea) and were maintained at the aquarium facility of the Department of Biological Science, Sungkyunkwan University (Suwon, South Korea). In brief, the copepods were maintained at 14 °C with a light/dark ratio of 12L/12D and artificial seawater (TetraMarine Salt Pro, Tetra™, Cincinnati, OH, USA; 5.71 ± 0.19 mg O_2/l) adjusted to 30 practical salinity units (psu). The copepods were fed with a green alga *Chlorella vulgaris* (approximately 6×10^4 cells/ml). MixS descriptors are presented in Table 1.

2.2. Illumina sequencing

Approximately two hundred adult *T. kingsejongensis* were homogenized in three volumes of TRIZOL® reagent (Invitrogen, Paisley,

* Corresponding authors.
E-mail addresses: sangheekim@kopri.re.kr (S. Kim), jslee2@skku.edu (J.-S. Lee).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.margen.2016.04.009
1874-7787/

Table 1
MIxS descriptors.

Item	Description
Investigation_type	Eukaryote
Project_name	*Tigriopus kingsejongensis* assembly
Lat_lon	−60° 14′ S, −58° 47′ W
Geo_loc_name	Antarctica: Maxwell bay
Collected_by	Korea Polar Research Institute
Collection_date	Nov.-13
Environment	Tidal pool
Biome	ENVO:01000108
Feature	ENVO:00000317
Material	ENVO:00002149
Temp	14 °C
Salinity	30 psu
Sequencing method	Illumina HiSeq 2000
Assembly method	Trinityrnaseq v 2.0.6
Assembly name	*Tigriopus kingsejongensis* assembly
Genome coverage	40×

Scotland) with a tissue grinder. Total RNA was extracted according to the manufacturers' instructions. DNA digestion was performed using DNase I (Sigma, St. Louis, MO, USA). Total RNA was quantified by absorption of light at A260 and quality checked by analyzing the ratios A230/260 and A260/280 using a spectrophotometer (QIAxpert®, Qiagen, Hilden, Germany). Synthesis of the paired-end library and the sequencing with Illumina HiSeq™ 2000 (Illumina, San Diego, CA, USA) were performed at the National Instrumentation Center for Environmental Management (NICEM, Seoul National University, Seoul, South Korea). Briefly, adaptor-ligated fragments were separated by size on an agarose gel, and the desired range of cDNA fragments (200 ± 25 bp) was excised from the gel. Proper fragments were selected and purified and subsequently PCR amplified to create the final cDNA library template. Image deconvolution, base calling, and quality value calculations were performed using the Illumina GA pipeline (ver. 1.6) according to the manufacturer's instructions.

2.3. Transcriptome assembly

The assembled transcriptome was annotated using Trinity ver. 2.0.6 (https://trinityrnaseq.github.io/) with an e-value threshold of 1×10^{-5} (Table 2). Briefly, Illumina HiSeq™ 2000 produced 139,923,976 reads representing a total of 21,073,987,179 nucleotides. The raw reads were cleaned by filtering out adaptor-only reads, trimming of adaptor sequences, empty nucleotides, ('N' in the end of reads), and low quality sequences (reads containing more than 50% bases with Q-value ≤ 20). Then de novo assembly of the clean reads was performed to generate non-redundant unigenes. The large contigs of the sequence reads, filtered by the quality control, were constructed using the de novo assembler Trinity (ver. 2.0.6; Grabherr et al., 2011). Using the sequence reads that passed quality control, we assembled 81,653 contigs with Trinity with a size range of 224 to 8427 bp. TransDecoder (http://transdecoder.sourceforge.net/) was used to identify candidate coding regions from the assembled transcripts and/or contigs; the candidate coding regions were used for BLAST analysis against the NCBI non-redundant (nr) protein database. Among the assembled contigs, TransDecoder found 38,250 contigs that contain candidate coding regions with a size range of 297 to 7809 bp. (Table 2, Fig. S1). Average contig size, median contig size, and N50 value were 954 bp, 699 bp, and 1283 bp, respectively.

Table 2
Assembly and annotation statistics.

Assembled bases	Number of contigs	Mean contig length	Median contig length	N50	GC content
36,497,199	38,250	954	699	1283	54.16

Annotation statistics

NCBI NR Blast	SignalP	InterProScan	GO (annotated)
23,918	3388	24,083	16,815

2.4. Gene Ontology and KEGG pathway analysis

BLAST analysis found that 22,977 unigenes (60%) had positive matches (e-value < 1e−06) to homologous genes of other species (Supplementary file S1). Most of contigs are related to those of the intertidal copepod *T. japonicus* (25,030 contigs; 65%) (Fig. S2A), and over 79% showed their highest BLAST hit with those of Arthropoda species (Fig. S2B). A total of 6723 conserved domains in the assembled transcripts was identified and annotated using InterProScan5 (Supplementary file S2) (Zdobnov and Apweiler, 2001). Gene Ontology (GO) and KEGG pathway analysis of the contigs were performed using Blast2GO (Conesa et al., 2005). All the results of GO analysis were analyzed over the second level, and GO terms related to the top domains were described in Fig. S3. In addition, detailed GO distributions in three GO categories (biological process, cellular component, and molecular function) were incorporated in Fig. S4A–C. Analysis of KEGG pathway showed that most annotated sequences were related to metabolism pathways with 116 predicted pathways (Supplementary file S3). Analysis of the KEGG pathway showed that most annotated sequences were related to metabolic pathways, including amino acid metabolism, nucleotide metabolism, carbohydrate metabolism, and glycan biosynthesis and metabolism (Fig. S5).

2.5. Lipid metabolism

Temperature is one of the abiotic key factors in the marine environment. One of the most consistent biochemical responses of organisms to low temperature is a modulation in fatty acid unsaturation of both membrane and depot lipids in a cell (Cossins and Prosser, 1978; Hazel and Williams, 1990). Thus, we analyzed the composition of lipid-relevant pathways and the numbers of genes involved in their pathways from the GO and KEGG results of *T. kingsejongensis* transcriptome with those of *T. japonicus* that was recently analyzed (Kim et al., 2015). Interestingly, unique enzymes (#32) of *T. kingsejongensis* involved in the lipid metabolism of GO terms were more highly detected than

Table 3
Numbers of sequences, pathways, enzymes and unique enzymes involved in the high level of pathways.

High-level of pathways	#seqs	#pathways	#enzyme	#unique enzymes
Amino acid metabolism	362	14	117	75
Biosynthesis of antibiotics	168	1	73	73
Biosynthesis of other secondary metabolites	42	12	22	16
Carbohydrate metabolism	354	15	152	91
Energy metabolism	121	6	44	39
Glycan biosynthesis and metabolism	250	13	35	22
Immune system	23	1	2	2
Lipid metabolism	244	14	37	32
Metabolism of cofactors and vitamins	116	11	46	45
Metabolism of other amino acids	149	9	41	37
Metabolism of terpenoids and polyketides	30	5	18	17
Nucleotide metabolism	357	2	53	45
Signal transduction	65	2	12	12
Translation	41	1	20	20
Xenobiotics biodegradation and metabolism	74	10	26	21

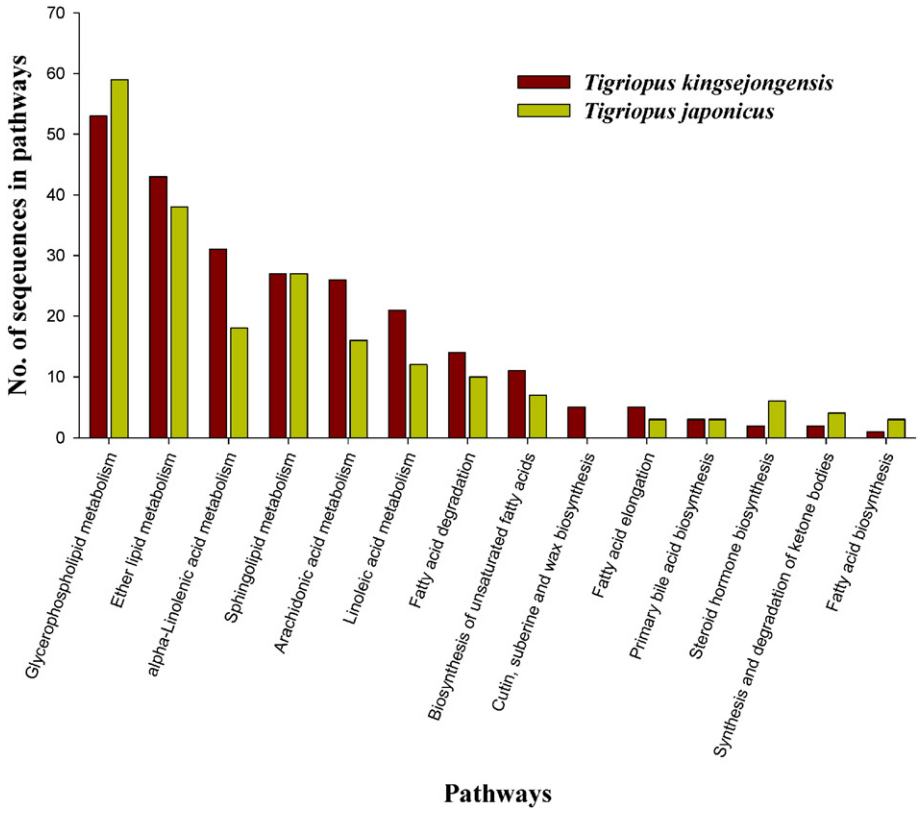

Fig. 1. Comparison of the numbers of unigenes involved in the lipid metabolism category from the KEGG results of *T. kingsejongensis* and *T. japonicus*.

those of *T. japonicus* (#24) (Table 3) (Kim et al., 2015). KEGG analysis also revealed that 244 genes were involved in the lipid metabolism from 22,977 unigenes of *T. kingsejongensis*, while *T. japonicus* bore only 206 lipid-relevant pathways genes even though a high number of *T. japonicus* unigenes (39,507 genes) were mapped to the KEGG database than those of *T. kingsejongensis* (Fig. 1, Supplementary file S3). Thus, we conclude that *T. kingsejongensis* has developed storage and utilization (e.g. lipid bilayer's viscosity) of lipid metabolisms for ecologic and energetic adaptations to life in polar oceans. As for pathways, the "cutin, suberine and wax biosynthesis" category is notable as five unigenes of *T. kingsejongensis* were mapped to the category, while no unigene was identified in the *T. japonicus* KEGG pathway. This result strongly supports previous findings that calanoid copepods from high latitudes are extensively accumulating storage lipids, almost mainly as wax esters (Lee et al., 1971; Kattner and Hagen, 1995).

2.6. Data deposition

The sequences of the transcript contigs were deposited to the Transcriptome Shotgun Assembly (TSA) database in GenBank (Accession no. GDFW00000000).

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.margen.2016.04.009.

Acknowledgments

We thank Prof. Hans-U. Dahms for his comments on the manuscript. This work was supported by a grant from the Marine Biotechnology Program (PJT200620, Genome analysis of marine organisms and development of functional application) funded by the Ministry of Oceans and Fisheries of Korea to Jae-Seong Lee and also supported by the project titled "Korea-Polar Ocean Development: K-POD" (project no. PM15050) funded by the Ministry of Oceans and Fisheries, Korea.

References

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 21, 3674–3676.

Cossins, A.R., Prosser, C.L., 1978. Evolutionary adaptation of membranes to temperature. Proc. Natl. Acad. Sci. U. S. A. 75, 2040–2043.

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.

Hazel, J.R., Williams, E.E., 1990. The role of alterations in membrane lipid composition in enabling physiological adaptation of organisms to their physical environment. Prog. Lipid Res. 29, 167–227.

Kattner, G., Hagen, W., 1995. Polar herbivorous copepods — different pathways in lipid biosynthesis. ICES J. Mar. Sci. 52, 329–335.

Kim, H.-S., Lee, B.-Y., Won, E.-J., Han, J., Hwang, D.-S., Park, H.G., Lee, J.-S., 2015. Identification of xenobiotic biodegradation and metabolism-related genes in the copepod *Tigriopus japonicus* whole transcriptome analysis. Mar. Genomics 24, 207–208.

Lee, R.F., Nevenzel, J.C., Paffenhofer, G.A., 1971. Importance of wax esters and other lipids in the marine food chain: phytoplankton and copepods. Mar. Biol. 9, 99–108.

Park, E.-O., Lee, S., Cho, M., Yoon, S.H., Lee, Y., Lee, W., 2014. A new species of the genus *Tigriopus* (Copepoda: Harpacticoida: Harpacticidae) from Antarctica. Proc. Biol. Soc. Wash. 127, 138–154.

Raisuddin, S., Kwok, K.W.H., Leung, K.M.Y., Schlenk, D., Lee, J.-S., 2007. The copepod *Tigriopus*: a promising marine model organism for ecotoxicology and environmental genomics. Aquat. Toxicol. 83, 161–173.

Soyer, J., Thiriot-Quievreux, C., Colomines, J.C., 1987. Description de deux espèces jumelles du groupe *Tigriopus angulatus* (Copepoda, Harpacticoida) dans les archipels Crozet et Kerguelen (Terres Australes et Antarctiques Françaises). Zool. Scr. 16, 143–154.

Zdobnov, E.M., Apweiler, R., 2001. InterProScan—an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17, 847–848.