Marine Genomics xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

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

De novo transcriptome sequencing of an estuarine amphipod *Grandidierella japonica* exposed to zinc

Kyoshiro Hiki[a,*], Nobuyoshi Nakajima[b], Haruna Watanabe[c], Fumiyuki Nakajima[a], Tomohiro Tobino[d]

[a] Department of Urban Engineering, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-8656, Japan
[b] Center for Environmental Biology and Ecosystem Studies, National Institute for Environmental Studies, Onogawa 16-2, Tsukuba, Ibaraki 305-8506, Japan
[c] Center for Environmental Risk Research, National Institute for Environmental Studies, Onogawa 16-2, Tsukuba, Ibaraki 305-8506, Japan
[d] Environmental Science Center, The University of Tokyo, Hongo 7-3-1, Bunkyo-Ku, Tokyo 113-0033, Japan

ARTICLE INFO

Keywords:
De novo transcriptome sequencing
Grandidierella japonica
Zinc
Amphipod
Biomarker

ABSTRACT

Despite the ecological and ecotoxicological importance of the amphipod *Grandidierella japonica*, there is little molecular information on this species. We performed de novo transcriptome sequencing of *G. japonica* under zinc exposure to develop toxicant responsive transcriptome resources. Using the Ion Torrent platform we obtained 11 million raw reads, which were assembled into 105,166 transcripts (N50: 511 bases). The obtained sequences include many stress- and defence-related genes such as antioxidant genes and members of the cytochrome P450 family as well as potential reference genes. Our transcriptome data of *G. japonica* should be a useful reference for future ecological and ecotoxicological studies.

1. Introduction

Amphipods are organisms living in direct contact with sediment and are thus among the most susceptible organisms to estuarine and marine sediment contamination (Long et al., 1990; Gesteira and Dauvin, 2000). They have been widely used as test species for sediment toxicity assessments (USEPA, 1994). The estuarine amphipod *Grandidierella japonica* is a species native to Japan, and is widely distributed globally, such as along the Pacific coast of North America, Australia and Europe probably due to anthropogenic dispersal (Chapman and Dorman, 1975; Myers, 1981; Pilgrim et al., 2013; Marchini et al., 2016). The sensitivity of *G. japonica* to metals and various organic pollutants was reported to be comparable to those of other brackish and marine amphipods frequently used in sediment toxicity assessments such as *Ampelisca abdita*, *Eohaustorius estuarius* and *Melita plumulosus* (Lee et al., 2005; King et al., 2006). *G. japonica* has a relatively high tolerance to ammonia and differences in grain size of the sediment (Nipper et al., 1989; Lee et al., 2005). This high tolerance makes this species attractive for sediment toxicity tests; ammonia and fine particles are often confounding factors in these tests. In addition, this species reproduces rapidly (Greenstein and Tiefenthaler, 1997) and is easily cultured in the laboratory. As a consequence, *G. japonica* has been used for estuarine and marine sediment toxicity assessments, especially in areas where no other sentinel amphipod species is found (Nipper et al., 1989).

Despite its ecotoxicological value, to date, the sequence information of *G. japonica* is available only for cytochrome *c* oxidase subunit I of mitochondrial DNA (Pilgrim et al., 2013), 18S ribosomal RNA (Ito et al., 2008) and few cDNA fragments (Hiki et al., 2017). This sequence information is insufficient to understand the mechanistic basis for sediment toxicity. More sequence information on stress- and defence-related genes, such as members of heat shock protein family and cytochrome P450 family, is required to efficiently use the molecular responses of this species for environmental risk assessments.

In this study, we aimed to obtain transcriptome sequence information of *G. japonica* and to determine the sequences of genes related to chemical stress. We performed de novo transcriptome analysis of this species when exposed to zinc solution, using the Ion Torrent platform. Zinc was selected as a contaminant of concern in Japan and as a model toxicant that is known to induce the expression of stress- and defence-related genes (e.g. cytochrome P450) in another amphipod species (Hook et al., 2014). The transcriptome data obtained in this study will serve as reference for other crustacean species, as well as a resource for future ecotoxicological studies using this species.

* Corresponding author.
E-mail addresses: hikik@env.t.u-tokyo.ac.jp, hiki.1225@gmail.com (K. Hiki).

https://doi.org/10.1016/j.margen.2017.11.011
Received 23 October 2017; Received in revised form 24 November 2017; Accepted 27 November 2017

Table 1
Characteristics of *G. japonica* transcriptome sequencing project in compliance with the MIGS standard.

Item	Description
Investigation type	Eukaryote transcriptome
Project name	*Grandidierella japonica* transcriptome
Collected date	3 Oct 2016
Environment (biome)	ENVO_00002030 (aquatic biome)
Environment (feature)	ENVO_00000241 (tidal mudflat)
Environment (material)	ENVO_00002006 (water)
Isolation and growth condition	Hiki and Nakajima, 2015 (https://doi.org/10.2166/wst.2015.304)
Tissue type	Whole body
Developmental stage	Juvenile
Sequencing technology	Ion PGM
Assembly	Trinity v.2.3.1
Finishing strategy	Contigs
Data accessibility	Bioproject PRJDB5548

2. Methods and analysis

2.1. Sources of organisms

The amphipods were collected from a tidal flat on 23 May 2009 at the mouth of the Nekozane River in Chiba (35°39′20″ N, 139°55′08″ E), Japan, and were maintained in aquariums containing river sediment and artificial seawater (Table 1). The aquariums were maintained at 25 °C under a photoperiod of 16 light (L):8 dark (D). The culture sediment was collected from Komatsugawa tidal flats (35°41′16″ N, 139°51′06″ E). Juvenile amphipods, which were defined as individuals retained on a 500-μm nylon mesh after being separated from the adults with a 710-μm mesh, were collected from culture aquariums on 29 September 2016, and used for transcriptome analysis after exposure to zinc for 4 days. Zinc exposure was performed in a glass beaker containing a thin layer of quartz sand and 120 mL of seawater (30‰ salinity) spiked with zinc sulphate (purity: 99.5%, Wako Pure Chemical Industries, Ltd., Osaka, Japan) at a nominal concentration of 156 μg Zn/L. The beakers were maintained at 25 °C under a photoperiod of 16L:8D 24 h prior to beginning exposure. Half of the overlying water was replaced with new spiked seawater every 24 h. Aeration and additional food were not provided during exposure. After 96 h of exposure, surviving individuals were soaked in RNAlater™ (Ambion, Austin, TX, USA) after cutting of the thorax with tweezers to allow RNAlater™ to infiltrate into the body and were then stored at 4 °C for < 12 h until RNA extraction. The measured pH, dissolved organic carbon concentrations and zinc concentrations in the overlying water after 96 h of exposure were 7.9, 17.8 mg/L and below the detection limit in the control, and were 7.9, 11.5 mg/L and 128.4 μg Zn/L in the zinc treatment. The mortality was > 85% in both treatments.

2.2. RNA isolation and cDNA library construction

All surviving amphipods (17 for zinc and 18 for control) were pooled by treatment and total RNA of them was extracted using TRI Reagent (MRC, Cincinnati, OH, USA). Only one replicate with 17 and 18 individuals was used for each treatment. Since the RNA integrity number is not reliable for some invertebrate species (Ishikawa, 1977), RNA integrity was considered acceptable only if total RNA on a 1.0% denaturing agarose gel showed 18S rRNA bands and did not show broad smears below these bands. RNA purity was verified to have A260/A280 ratios of between 1.8 and 2.1 using Nanodrop™ 1000 (Thermo Scientific, Waltham, MA, USA). The removal of DNA contamination and rRNA was performed using DNA-free (Ambion) and Oligotex-dT30 (Takara, Shiga, Japan).

The cDNA library was constructed using Ion Total RNA-Seq Kit v2 (Life Technologies, CA, USA) following the manufacturer's protocol. The size distribution and concentration of the constructed libraries were checked using TapeStation 2200 (Agilent Technologies, CA, USA). Emulsion PCR was performed using the One Touch 2 system following the Ion PGM™ OT2 Template 400 kit protocol (Life Technologies). Enriched spheres were loaded into the Ion 318™ chip and sequenced on Ion PGM using the Ion PGM™ Hi-Q™ sequencing kit (Life Technologies).

2.3. Sequencing, assembly and annotation

A total of 11,713,816 raw reads were obtained from the two cDNA libraries. After the removal of low-quality reads by filtering those with an average quality score < 10 as initial screening and by trimming reads of 3′ ends with a quality score < 20 using the Fastx toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), 8,972,234 reads were retained. After the removal of rRNA using SortMeRNA (ver. 2.1; Kopylova et al., 2012) with the eight databases supplied with the software, 8,141,808 reads were retained. These reads were assembled into 105,116 transcripts with an N50 of 511 bases by using Trinity (ver. 2.3.1; Grabherr et al., 2011) with the default parameters. The assembled transcripts were clustered into 78,572 unigenes using the CD-HIT-EST program (ver. 4.6.6; Huang et al., 2010) at a sequence identity threshold of 98% to reduce redundancy. After mapping the clean reads back onto the clustered unigenes by using RSEM (ver. 1.3.0; Li and Dewey, 2011), a total of 77,708 unigenes (98.9%) had mapped reads. This high proportion indicates the successful de novo assembly of *G. japonica*. For the control and zinc exposure libraries, 69,731 unigenes (88.7%) and 74,758 unigenes (95.1%) had mapped reads, respectively.

The unclustered transcripts and the longest sequences in each cluster were regarded as unigenes for functional analysis. A total of 21,461 unigenes were annotated against the non-redundant nucleotide (nt; downloaded on 27 December 2016), Clusters of Orthologous Groups (COG; 2014 update) and Swissprot (downloaded on 15 December 2016) databases using BLAST + (ver. 2.5.0) with an E-value cut-off of 10^{-5}. Among them, 4935 unigenes were hit against the nt database, 6186 against the COG database and 20,241 against the Swissprot database. The low proportion of annotated unigenes (25.8%) was due to the limited length of assembled sequences (Zhang et al., 2016). The proportion of unigenes with homologies to known sequences was 13% for unigenes of 200–1000 bases in length, but was increased with increasing length of unigenes, and reached 73% for unigenes of > 1300 bases in length (Fig. 1). Further deep sequencing of *G. japonica* is required to provide more comprehensive functional

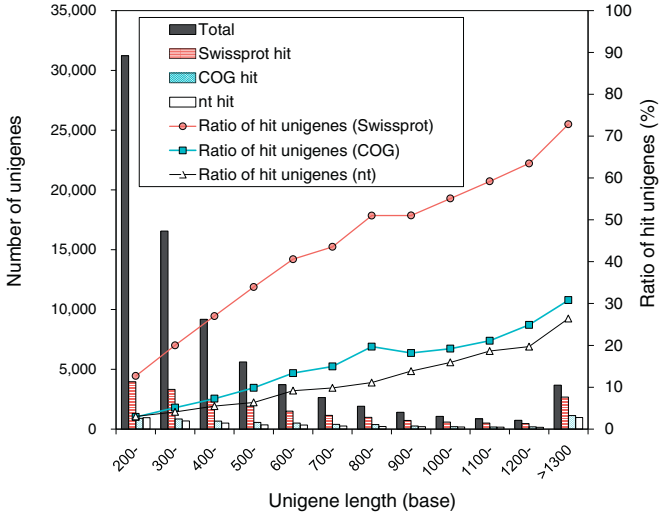

Fig. 1. Length distribution of all unigenes and unigenes hit against databases. The ratios of hit unigenes against the Swissprot, COG and nt databases are shown as lines (right axis).