Comparative Biochemistry and Physiology, Part D 8 (2013) 344–351

Contents lists available at ScienceDirect

Comparative Biochemistry and Physiology, Part D

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

# Characterization and analysis of a transcriptome from the boreal spider crab *Hyas araneus*

Lars Harms [a,*], Stephan Frickenhaus [b], Melanie Schiffer [a], Felix C. Mark [a], Daniela Storch [a], Hans-Otto Pörtner [a], Christoph Held [c], Magnus Lucassen [a]

[a] Integrative Ecophysiology, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany
[b] Scientific Computing, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany
[c] Functional Ecology, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany

ARTICLE INFO

*Article history:*
Received 22 May 2013
Received in revised form 5 September 2013
Accepted 30 September 2013
Available online 10 October 2013

*Keywords:*
Spider crab (*Hyas araneus*)
Crustacea
Transcriptome
Markov Clustering Algorithm
Anti-viral immunity

ABSTRACT

Research investigating the genetic basis of physiological responses has significantly broadened our understanding of the mechanisms underlying organismic response to environmental change. However, genomic data are currently available for few taxa only, thus excluding physiological model species from this approach. In this study we report the transcriptome of the model organism *Hyas araneus* from Spitsbergen (Arctic). We generated 20,479 transcripts, using the 454 GS FLX sequencing technology in combination with an Illumina HiSeq sequencing approach. Annotation by Blastx revealed 7159 blast hits in the NCBI non-redundant protein database. The comparison between the spider crab *H. araneus* transcriptome and EST libraries of the European lobster *Homarus americanus* and the porcelain crab *Petrolisthes cinctipes* yielded 3229/2581 sequences with a significant hit, respectively. The clustering by the Markov Clustering Algorithm (MCL) revealed a common core of 1710 clusters present in all three species and 5903 unique clusters for *H. araneus*. The combined sequencing approaches generated transcripts that will greatly expand the limited genomic data available for crustaceans. We introduce the MCL clustering for transcriptome comparisons as a simple approach to estimate similarities between transcriptomic libraries of different size and quality and to analyze homologies within the selected group of species. In particular, we identified a large variety of reverse transcriptase (RT) sequences not only in the *H. araneus* transcriptome and other decapod crustaceans, but also sea urchin, supporting the hypothesis of a heritable, anti-viral immunity and the proposed viral fragment integration by host-derived RTs in marine invertebrates.
© 2013 Elsevier Inc. All rights reserved.

## 1. Introduction

The great spider crab, *Hyas araneus*, is a benthic decapod crab that lives on sublitoral rocky or sandy substrates to a depth of 50 m (Hayward and Ryland, 1990). Within the North-East Atlantic region it is distributed along a latitudinal gradient from the English Channel up to the Arctic regions of Spitsbergen, where it represents one of the most prominent brachyuran crabs (Zittier et al., 2012). The size of its distribution range and the corresponding cline in environmental conditions make *H. araneus* an ideal species to study the effects of environmental changes as well as functional differentiation between populations. For example, decreased larval developmental rates in Arctic compared to temperate populations suggest adaptation to the polar cold (Walther et al., 2010). Elevated seawater $PCO_2$ (as projected by ocean acidification scenarios) caused an increase in metabolic rate during larval development pointing to higher metabolic costs in larvae (Schiffer et al., 2012). Adult *H. araneus* displayed increased heat sensitivity under elevated $CO_2$ levels with potential consequences for biogeographical boundaries (Walther et al., 2009). However, the genetic basis of these responses to environmental changes has so far only been investigated for a limited number of candidate genes. For example, hyastatin, a peptide involved in haemolymph antimicrobial defense, has been isolated, and the importance of the cys-containing region for the antimicrobial activity and a possible multifunctional character has been demonstrated (Sperstad et al., 2009). The reason for the small number of studies is likely the lack of genomic information in databases like the National Center for Biotechnology Information (NCBI). Currently, only 26 nucleotide sequences of *H. araneus* are published in NCBI.

In recent years, Next Generation Sequencing (NGS) has made it possible to approach this problem by sequencing and assembly of entire genomes of ecologically relevant species (for review see Wheat, 2010). However, for non-model organisms, sequencing a transcriptome rather than the genome to obtain the genetic data is advantageous for many reasons. The generation of sequence data is quick, it is relatively cost-effective and can thus provide the genetic basis for studies with fewer resources. Further, transcriptome sequencing can provide both

*Abbreviations:* MCL, Markov Clustering Algorithm; CEGMA, Core Eukaryotic Genes Mapping Approach.

* Corresponding author.
*E-mail address:* LHarms@gmx.net (L. Harms).

1744-117X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.cbd.2013.09.004

expression and coding data, using RNA-seq (Martin and Wang, 2011). Using different tissues and differentially treated animals it is possible to capture variations in coding sequences, stress induced sequences as well as differences in the expression level. Respective approaches have already been applied to a number of marine invertebrates to achieve insights into expression information (Giant Ezo scallop (Hou et al., 2011); common octopus (Zhang et al., 2012); 2 Mollusca, 2 Arthropoda, 2 Annelida, 2 Memertea, 2 Porifera (Riesgo et al., 2012); pearl oyster (Shi et al., 2013)), thereby expanding the existing genetic resources massively.

Thus, the objective of the present study was to fundamentally characterize the transcriptome of *H. araneus*. For analyzing specific homologies within decapod transcriptomes and for identifying common and specific gene clusters of the selected group of species we introduced the Markov Clustering Algorithm (MCL) clustering approach.

To develop an extensive transcriptome of *H. araneus* we combined the 454 and Illumina sequencing technologies on normalized and common cDNA libraries constructed from pooled samples of multiple tissues from animals treated with different environmental conditions (see Materials and methods). We assembled the sequences to reconstruct transcripts potentially representing the *H. araneus* transcriptome. Because no reference genome is available for *H. araneus* we assembled the transcriptome *de novo*. There are several *de novo* tools available, but none represent the perfect solution (Kumar and Blaxter, 2010). To obtain a comprehensive and high-quality de novo assembly of the *H. araneus* transcriptome, we tested different assembling tools and compared the resulting assemblies. In the second part we analyzed the functionally annotated transcriptome for particular features and compared the identified sequences with available sequence information of other decapod crustaceans using the MCL-clustering to reveal homologies within the selected group of species.

The approach illustrates a potential methodological framework and may promote further transcriptome studies in non-model organisms. The transcriptome obtained for *H. araneus* will become essential for future analyses and annotations and also provide useful information for future functional genomic studies in crustaceans.

## 2. Materials and methods

### 2.1. Sample preparation and RNA extraction

Adult specimens of the boreal spider crab *H. araneus* were collected in the Kongsfjord (N 78° 58.635′; E 11° 29.454′) at the west coast of Spitsbergen (Norway). Animals were acclimated for 10 weeks in flow through aquaria systems to 6 different treatments of 3 seawater $PCO_2$ values (390, 1120 and 1960 μatm) combined with two temperatures (5 and 10 °C), respectively. Tissue samples comprising of all 6 gill-arches, tegument, heart, hepatopancreas, testis and pincer muscle were collected from four to six animals per treatment and directly frozen in liquid nitrogen. Samples were stored at −80 °C until used for RNA extraction. Total tissue RNA was extracted by using the RNeasy Mini Kit according to the "Purification of Total RNA from Animal Tissue" protocol (QIAGEN, Hilden, Germany). RNA quantities were determined by a NanoDrop 2000c spectrometer (PeqLab, Erlangen, Germany), and RNA was analyzed for quality by microfluidic electrophoresis in an Agilent 2100 Bioanalyzer (Agilent Technologies).

### 2.2. Sequencing and assembly

To generate the transcriptome of the non-model organism *H. araneus*, two different sequencing approaches were used. First, a 454 pyrosequencing approach based on normalized cDNA libraries was applied, serving as a basis for the assembly. Using samples from multiple tissues and differentially treated animals (rearing temperature and $PCO_2$ level) as well as using a normalized cDNA libraries allow for a comprehensive transcriptome, capturing variations in coding sequences, stress induced sequences as well as low expressed genes. Two separate cDNA libraries were sequenced by 454: a library exclusively based on gill samples and a library based on samples of a mixture of tissues. For the *H. araneus* gill library, the same amount of RNA was collected from each gill of 4 animals per treatment and pooled in one sample. The same was done for all other tissues to prepare the material for a mixed tissue library. Both mixtures were used for the library constructions by the Max Planck Institute for Molecular Genetics (Berlin, Germany). Total RNA of the two pools (gill and mixed tissue) was used for cDNA synthesis using the SMART protocol (Mint-Universal cDNA synthesis kit, Evrogen, Moscow, Russia). The cDNA was subsequently normalized using duplex-specific nuclease and re-amplified thereafter following the instructions of the "Trimmer Kit" (Evrogen, Moscow, Russia). Sequencing libraries were prepared from cDNA using the "GS FLX Titanium General Library Preparation Kit" (Roche, Basel, Switzerland). Before sequencing, the libraries were amplified by polymerase chain reaction (PCR) using the 'GS FLX Titanium LV emPCR Kit' (Roche, Basel, Switzerland) (De Gregoris et al., 2011). Sequencing was performed by the Max Planck Institute for Molecular Genetics (Berlin, Germany) on a 454 Genome Sequencer FLX using the Titanium chemistry (Roche). Initial quality control and filtering of adapters and barcodes was performed at the Max Planck Institute for Molecular Genetics (Berlin, Germany). Both cleaned libraries were combined for the subsequent *de novo* assembly. To optimize the quality of the *de novo* transcriptome assembly, we compared two different assembler programs: GS De Novo Assembler version 2.6 (Newbler, Roche) and MIRA 3.0 (Chevreux et al., 1999). We tested each program with the following main assembly parameters: minimum percentage identities of 95%, and minimum overlap length of 40 bp for MIRA, and 40 bp for the GS De Novo Assembler. The "-cdna" mode was used for the GS De Novo Assembler. The final de novo assembly by GS De Novo Assembler was chosen based on basic assembly metrics and performance in terms of completeness and contiguity.

Secondly, an Illumina sequencing approach was used to enhance the 454 based transcriptome. Six different cDNA libraries based on samples of the six different treatments were sequenced. For each treatment, total RNA from all gills of 4 animals was pooled and used for the library construction by GATC Biotech (Konstanz, Germany). Libraries for each treatment were constructed according to the 'SMART protocol for Illumina sequencing' (Clontech, Mountain View, CA, USA). Illumina single-end sequencing was performed on a HiSeq 2000 Sequencer by GATC Biotech. Initial quality control and filtering of adapters was performed by GATC Biotech. In addition, obtained raw reads were quality controlled by FastQC (Babraham Institute, Cambridge, UK) and cleaned using the FastX-Toolkit (Hannon Lab — Cold Spring Harbor Laboratory, NY, USA). Quality control was performed using the following parameters: minimum quality score of 20, minimum percentage of bases within the quality score of 90 and a minimum length of 25 bases. To enhance the set of GS De Novo Assembler-assembled contigs, obtained Illumina-data from the six libraries were combined and reads were assembled *de novo* with ABySS version 1.3.2 (Simpson et al., 2009) with $k = 26$, minimum overlap length of 30 bp and minimum sequence identity of 0.9. Considerable overlaps with the GS De Novo Assembler-assembled 454-data were detected with blastn (word size 8), removing Abyss-contigs above E-value $10^{-10}$ and length below 500 bp. The transcriptome of *H. araneus* was deposited in the 'European Nucleotide Archive' (ENA) at the 'European Molecular Biological Laboratory-European Bioinformatics Institute' (EMBL-EBI) (Accession range: HAAI01000001–HAAI01019199).

### 2.3. Functional annotation

Functional annotation of the *H. araneus* transcriptome was accomplished using the Blast2GO software v.2.6.0 (Conesa et al., 2005; Gotz et al., 2008). Homology searches were performed using Blastx against the NCBI non-redundant protein database. Blast searches were performed with an E-value cut-off of $1E^{-3}$. For the Gene Ontology (GO)

Download English Version:

https://daneshyari.com/en/article/1978596

Download Persian Version:

https://daneshyari.com/article/1978596

Daneshyari.com