Gene 557 (2015) 28–34

Contents lists available at ScienceDirect

Gene

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

The eyestalk transcriptome of red swamp crayfish *Procambarus clarkii*

Chiara Manfrin [a], Moshe Tom [b], Gianluca De Moro [a], Marco Gerdol [a], Piero Giulio Giulianini [a], Alberto Pallavicini [a,*]

[a] Department of Life Sciences, University of Trieste, Via Giorgieri 5, Trieste 34127, Italy
[b] Israel Oceanographic and Limnological Research, P.O.B 8030, Haifa 31080, Israel

ARTICLE INFO

Article history:
Received 15 July 2014
Received in revised form 5 November 2014
Accepted 1 December 2014
Available online 3 December 2014

Keywords:
Procambarus clarkii
Red swamp crayfish
Transcriptome
Eyestalk
Illumina sequencing
Neuroendocrinology
Vision

ABSTRACT

The red swamp crayfish (*Procambarus clarkii*, Girard 1852) is among the most economically important freshwater crustacean species, and it is also considered one of the most aggressive invasive species worldwide. Despite its commercial importance and being one of the most studied crayfish species, its genomic and transcriptomic layout has only been partially studied. Illumina RNA-sequencing was applied to characterize the eyestalk transcriptome and identify its most characterizing genes. A collection of 83,170,732 reads from eyestalks was obtained using Illumina paired-end sequencing technology. A de novo assembly was performed with the Trinity assembly software generating 119,255 contigs (average length of 1007 bp) and identifying the first sequenced transcriptome in this species.

The eyestalk is a major site for the production of neurohormones and controls a variety of physiological functions such as osmotic regulation, molting, epidermal color patterns and reproduction. Hence, its transcriptomic characterization is interesting and potentially instrumental to the elucidation of genes which have not been comprehensively described yet. Moreover, the availability of such a large amount of information supported the characterization of molecular families which have never been described before. The *P. clarkii* eyestalk transcriptome reported here provides a resource for improving the knowledge of the still incompletely defined neuroendocrinology of this species and represents an important source of data for all the interested carcinologists.

1. Introduction

The red swamp crayfish (*Procambarus clarkii*, Girard 1852) is considered one of the most invasive species worldwide. Because of its high fitness, adaptability to changing environment, and high fecundity, this species has colonized a wide range of aquatic habitats. Originally native to Mexico and Louisiana (USA), this crayfish is now widely distributed worldwide as a result of the introduction of commercial harvests for the food industry (ISSG, 2012). Its expansion has undoubtedly contributed to a loss of biodiversity in the invaded habitats, especially in the areas where autochthon crayfishes were not able to compete with the new invader. The eyestalk — a group of organs of sight and visual perception — contains the neurosecretory X-organ/sinus gland complex (Ollivaux et al., 2006; Christie et al., 2010; Webster et al., 2012), which produces a superfamily of hormones named Crustacean Hyperglycemic Hormone (CHH) that controls many fundamental physiological functions such as molting, osmoregulation, modulation of glycemia, reproduction (Brown, 1934; Scharrer, 1952; Huberman and Aguilar, 1989; De Kleijn, 1994; Chung and Webster, 2003; Katayama et al., 2013; Turner et al., 2013) and behavioral responses, such as aggression (Aquiloni et al., 2012) and anxiety (Lok et al., 1977).

A thorough analysis of the main hormone products produced in the eyestalk was helpful in identifying variants, hormone-like transcripts and patterns of expression revealing aspects never considered before. The absence of the complete sequenced genome for this species makes the transcriptomic information derived from eyestalk among the only ready-to-use dataset currently available for researchers. To date, in fact, most of the molecular studies regarding the genus *Procambarus* are focused on population genetics (Li et al., 2012; Liu et al., 2013; Shen et al., 2013; Zhu et al., 2013) and immunity (Du et al., 2013; Zhang et al., 2013). In the past years, major efforts were devoted to increase knowledge of crustacean neurohormones and their structures, but unfortunately the genomic information of these animals is still limited (Hopkins, 2012; Webster et al., 2012). For the first time, this study reports the characterization of the eyestalk transcriptome of *P. clarkii*, significantly increasing the molecular knowledge base of this species. Prior to this study, 556 *P. clarkii* nucleotide sequences were filed within public databases, whereas about 1532 belonged to the *Procambarus* genus. Consequently, thanks to the application of molecular techniques to investigate different aspects of environmental biology and the development of new sequencing methods, easily applicable to

Abbreviations: CHH, Crustacean Hyperglycemic Hormone; MIH, Molt Inhibiting Hormone; ITP, Ion Transport Peptide; FPKM, Fragments Per Kilobase per Million mapped reads; ORF, Open Reading Frame; GPCR, G Protein-Coupled Receptor; SARA, Smad Anchor for Receptor Activation.

* Corresponding author at: Department of Life Sciences, University of Trieste, Via L. Giorgieri 5, Trieste, Italy.
E-mail address: pallavic@units.it (A. Pallavicini).

http://dx.doi.org/10.1016/j.gene.2014.12.001
0378-1119/

non-model organisms, it is possible to study topics which up to now have been excluded due to technical limitation. As an example of applied research, transcriptome data mining represents a new way to identify hormones that could be used as autocidal methods to hamper the diffusion of the red swamp crayfish.

2. Materials and methods

2.1. Animal maintenance

Two *P. clarkii* females and two males, collected in Sesto al Reghena (12°50′32″E, 45°51′34″N), were used as biological sources to produce the eyestalk transcriptome. All individuals were adult, in intermolt and non-reproductive stages. Before the eyestalk ablation, they were kept for a week in 120-liter tanks provided with closed circuit for filtered and aerated tap water at ~18 °C. Animals were fed on pellets (Sera granular, Heisenberg, Germany) three times a day.

2.1.1. Ethical note

The following experimental procedures comply with the current applicable laws of Italy, the country where they were performed. No specific permits are required for studies that do not involve endangered or protected invertebrate species. Individuals were maintained in appropriate laboratory conditions to guarantee their welfare and responsiveness. Upon completion of experiments, crayfish were euthanized by hypothermia.

2.2. RNA extraction, sequencing and de novo transcriptome assembly

Total RNA from 8 frozen eyestalks was extracted by homogenization in TriReagent RNA isolation solution (Sigma-Aldrich) and purified with RNeasy MinElute Cleanup Kit (Qiagen). First dissections were performed removing the retina up to the 4th ganglionic swelling because this region is fully pigmented and negatively impacts the quality of extracted RNA. Therefore, as showed in Fig. 1, the eyestalk tissues and components analyzed in this study included: hypodermic and muscle tissues, 1st–3rd ganglionic swellings (*medulla terminalis*, *medulla interna* and *medulla externa*), sinus gland, and X-organ. In order to obtain high-quality RNA, tissues were dissected and immediately stored in liquid nitrogen. RNA concentration and quality were assessed using Qubit RNA assay (Invitrogen) and Agilent 2100 Bioanalyzer (Agilent Technologies). An average of about 7 μg specimen was extracted, and 5 μg of pooled RNA was sent for sequencing to the Institute of Applied Genomics (IGA) in Udine (Italy). cDNA libraries were created with the TruSeq RNA Sample Prep Kit v2 (Illumina) and then were 100 bp paired-end sequenced using Illumina HiSeq2000. The raw sequencing reads were trimmed according to the quality of bases and the presence of residual Illumina adapters.

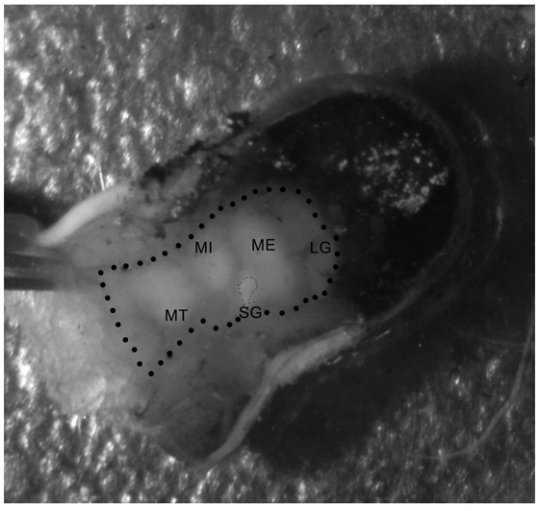

Fig. 1. Internal anatomy of the eyestalk. Marked by the dotted line, the dissected area from which RNA was extracted: LG: lamina ganglionaris, ME: medulla externa, MI: medulla interna, SG: sinus gland, and MT: medulla terminalis.

The de novo assembly of the trimmed reads was performed with Trinity software package (Grabherr et al., 2011), using the default options, allowing the formation of contigs longer than 250 base pairs. The Trinity software is aimed at creating a set of contigs, which is composed of all the possible expressed transcript variants per gene. From this highly redundant contig set we selected as representative the longest transcript with a minimum average coverage of 5 from each gene-related group ('comp' in the Trinity jargon).

2.3. Bioinformatic analysis

The eyestalk non-redundant transcriptome was annotated using the Trinotate suite (http://trinotate.sourceforge.net/). In detail, sequence similarities were detected by BLASTX (Altschul et al., 1990) against the UniProtKB/Swiss-Prot database. Codified peptides were predicted with Transdecoder. Functional domains were identified within the PFAM domain database (Punta et al., 2012) using HMMER (Finn et al., 2011). The prediction of signal peptide and transmembrane domains was performed through SignalP (Petersen et al., 2011) and tmHMM (Krogh et al., 2001), respectively. Finally, transcripts were assigned to eggNOG (Powell et al., 2012) and Gene Ontology (Ashburner et al., 2000) functional categories. We determined the percentage of full-length transcripts using an internal tool of Trinity. Large intergenic non-coding RNAs (lincRNAs) were detected within a selected set of transcripts with length >2 kb and without an Open Reading Frame (ORF) longer than 64 codons using Coding Potential Calculator (CPC) (Kong et al., 2007). RepeatMasker version open-4.0.3 (Smit et al., 2010) was used for detecting interspersed repeats, low complexity DNA sequences, and transposable elements using default conditions. Insertions, deletions and multi-nucleotide variations were identified via CLC Genomics Workbench v.6.03 (Aarhus, Denmark).

Gene expression data was obtained mapping sequence reads on the non-redundant eyestalk assembly using the RNA-seq tool of the CLC Genomics Workbench v.6.03 with the following parameters: minimum allowed length = 0.75, similarity fractions = 0.95 and maximum number of matching contigs = 10. Intact mapped paired-end reads were exclusively used for the calculation of expression values, which are given as Fragments Per Kilobase per Million mapped reads (FPKM).

The phylogenetic analysis was carried out by MEGA 6 (Tamura et al., 2013), initially testing the most suitable substitution model, and then using the Maximum Likelihood algorithm evaluated with 1000 bootstrap replica. All branches supported by less than 50% of replica were collapsed together.

3. Results and discussion

3.1. Transcriptome and assembly characterization

Deep transcriptome sequencing generated 83,170,732 trimmed nucleotide reads of which 80,649,690 are paired-end reads. Reads were filed within the Sequence Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) under the accession number SRR870673. Reads were assembled into 119,255 contigs by Trinity, then reduced to 46,770 as previously described in Materials and methods. Although Trinity assembled contigs were reduced by 60.8%, the 92.32% of reads was back mapped, meaning that only a negligible portion of the sequence data generated was excluded from gene expression analysis. The final non-redundant transcriptome was filed within the NCBI Transcriptome Shotgun Assembly (TSA) database under the accession number GARH01000000.