



Data in Brief

Transcriptome of the freshwater amphipod *Gammarus pulex* hepatopancreas



E. Gismondi *, J.P. Thomé

Laboratory of Animal Ecology and Ecotoxicology (LEAE), University of Liege, B6c, Allée du 6 Août 11, 4000 Liege, Belgium

ARTICLE INFO

Article history:

Received 6 April 2016

Accepted 11 April 2016

Available online 16 April 2016

Keywords:

Gammarus pulex

Hepatopancreas

Transcriptome

Illumina HiSeq

Ecotoxicology

ABSTRACT

So far, ecotoxicological studies used biomarkers of exposure or of effects in order to investigate the impacts of contaminated areas on biota (Peakall, 1994 [6]). However, although these results are important in the ecotoxicological risk assessment, biomarkers are very specific and only provide information on the biological processes or physiological pathways targeted by the biomarkers experimenters choose to test (Monsinjon and Knigge, 2007 [5]). In recent years, proteomics have become a major tool in ecotoxicology, as they provide a global insight into the mechanism of action of pollutants without the need of hypothesis testing or any preconception on the biological processes likely impacted (Gismondi et al., 2015; Trapp et al., 2015 [7]; Truebano, 2016 [8]). However, the analysis of proteomic results is often limited due to the lack of database, especially for non-model organisms, such as *Gammarus* sp, commonly used as biological model in ecotoxicology (Sornom et al., 2012 [11]; Vellinger et al., 2013 [9]; Gismondi and Thomé, 2014 [1]; Lebrun et al., 2014 [3]). Here, we performed Illumina HiSeq sequencing to total RNA isolated from the hepatopancreas (i.e. detoxification tissue) of *Gammarus pulex* males and females coming from uncontaminated river and contaminated river (e.g. PCB, benzo(a)pyrene). Approximately 290 M paired-end reads were assembled, filtered and sorted into 39,801 contigs whose 10,878 were similar of proteins available in databases. The assembled contigs could represent a reference hepatopancreas transcriptome for *G. pulex*, and constitute an important resource for future investigations on the impacts of pollutants on invertebrate biota, since it would improve the understanding of the mechanisms of action involved in toxicity. In addition, the hepatopancreas transcriptome will also allow the identification of new potential biomarkers for the ecotoxicological risk assessments. Assembled contigs were deposited in the European Nucleotide Archive under the BioProject number PRJEB13055, with accession numbers FJVI01000001–FJVI01039801.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Direct link to deposited data

<http://www.ebi.ac.uk/ena/data/view/PRJEB13055>.

2. Experimental design, materials and methods

2.1. *Gammarus pulex* sampling and library preparation

Adult males and females *Gammarus pulex* (size average: 10 mm and 7 mm, respectively) were collected in the Blanc-Gravier (50°34'60"N and 5°34'60"E, Liège, Belgium), a stream of good physicochemical quality, as defined by the European Directive, and in the Vesdre River (50°36'00"N and 5°37'58"E, Vaux-sous-Chevremont, Belgium) which is a river contaminated by several organic micropollutants (e.g. PCB dioxine-like and nondioxin like, benzo(a)pyrene). After transferring the samples to the laboratory into the corresponding river water, each individual were dissected to sample the hepatopancreas tissue. In order to have a wide range of conditions in the analysed sample, and increase the probability of protein detection, 5 hepatopancreas of males and 5 hepatopancreas of females, coming from the uncontaminated

Specifications

Organism/cell line/tissue	<i>Gammarus pulex</i> hepatopancreas
Sex	Male and female
Sequencer or array type	Illumina HiSeq 2500
Data format	Raw data: FASTAQ file
Experimental factors	Sampling in uncontaminated river and contaminated river (e.g. PCB, benzo(a)pyrene)
Experimental features	TruSeq RNA libraries synthesised and sequenced to identify transcriptome of the hepatopancreas tissue of <i>Gammarus pulex</i> .
Consent	Data are publicly available
Sample source location	Liège, Belgium

* Corresponding author.

Table 1
MixS descriptors.

Item	Description
Investigation type	Eukaryote
Project name	Hepatopancreas transcriptome for <i>Gammarus pulex</i>
Latitude, longitude	50°34'60"N, 5°34'60"E 50°36'00"N, 5°37'58"E
Geolocalisation	Belgium, Liege
Collected by	Eric GISMONDI
Collected date	04-Aug-14
Environment	Freshwater river
Biome	ENVO:00000873
Feature	ENVO:01000297
Material	ENVO:00002011
Depth	<0.5 m
Sequencing method	Illumina HiSeq 2500
Assembly method	Velvet v 1.2.07
Assembly name	Hepatopancreas transcriptome for <i>Gammarus pulex</i>
Genome recovery	94×

Table 2
Assembly statistics.

Assembled bases	Number of contigs	Mean contig length	Median contig length	N50	GC content (%)
22,614,229	39,801	568	387	641	54.5

site, and 5 hepatopancreas of males and 5 hepatopancreas of females, coming from the contaminated site, were pooled (i.e. 10 hepatopancreas males and 10 hepatopancreas of females per pool). Indeed, some proteins could be expressed (or over-expressed) only in contaminated conditions, as well as differently according to gender [2]; thus, a mixture of uncontaminated/contaminated individuals and males/females allow to increase the probability of identifying more transcripts. Pooled samples were performed in duplicates, total RNA were immediately isolated using a Qiagen RNeasy Lipid Tissue MiniKit (Qiagen, Germany), following the manufacturer's instructions. TruSeq RNA libraries were synthesised and sequenced using 2 × 100 base paired-end (Illumina HiSeq 2500, GenoScreen, Lille, France), and MixS descriptors are presented in Table 1.

2.2. Assembly of reads and annotation

The sequencing allowed to produce an average of 290.8 M of paired-end reads of which 85.6% were of good quality (Q30 score). The reads were assembled using VELVET v1.2.07 Software [10]. All the sequences were then taken into further process of redundancy removing using CD-

HIT-EST v4.6 [4] with a sequence identity threshold of 90%, leaving 39,801 sequences corresponding to 10,878 potential genes (Table 2). The assembled transcriptome was annotated using BLATP v2.2.29 with an e-value threshold of 1×10^{-7} .

2.3. Deposition in the database

Assembled contigs were deposited in the European Nucleotide Archive under the BioProject number PRJEB13055, with accession numbers FJVI01000001-FJVI01039801 (<http://www.ebi.ac.uk/ena/data/view/PRJEB13055>).

Acknowledgments

This work was funded by the BelPD-COFUND grant obtained by Eric Gismondi in the University of Liège. RNA sequencing and transcriptome assembly and annotations were performed by GenoScreen (Lille, France). Authors are grateful to Sandy Contreras for her great assistance during deposition in the database.

References

- [1] E. Gismondi, J.P. Thomé, Effects of two PBDE congeners on the moulting enzymes of the freshwater amphipod *Gammarus pulex*. Environ. Pollut. 191 (2014) 119–125.
- [2] E. Gismondi, G. Mazzucchelli, E. De Pauw, C. Joaquim-Justo, J.P. Thomé, Gender differences in responses in *Gammarus pulex* exposed to BDE-47: a gel-free proteomic approach. Ecotoxicol. Environ. Saf. 122 (2015) 205–213.
- [3] J.D. Lebrun, D. Leroy, A. Giusti, C. Gourlay-Francé, J.P. Thomé, Bioaccumulation of polybrominated diphenyl ethers (PBDEs) in *Gammarus pulex*: relative importance of different exposure routes and multipathway modeling. Aquat. Toxicol. 154 (2014) 107–113.
- [4] W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22 (2006) 1658–1659.
- [5] T. Monsinjon, T. Knigge, Proteomic applications in ecotoxicology. Proteomics 7 (2007) 2997–3009.
- [6] D.B. Peakall, The role of biomarkers in environmental assessment (1). Introduction. Ecotoxicology 3 (1994) 157–160.
- [7] J. Trapp, C. Almunia, J.C. Gaillard, O. Pible, A. Chaumot, O. Geffard, J. Armengaud, Data for comparative proteomics of ovaries from five non-model, crustacean amphipods. Data in Brief 5 (2015) 1–6.
- [8] M. Truebano, O. Tills, J.J. Spicer, Embryonic transcriptome of the brackishwater amphipod *Gammarus chevreuxi*. Mar. Genomics (2016), <http://dx.doi.org/10.1016/j.margen.2016.02.002>.
- [9] C. Vellinger, E. Gismondi, V. Felten, P. Rousselle, K. Mehennaoui, M. Parant, P. Usseglio-Polatera, Single and combined effects of cadmium and arsenate in *Gammarus pulex* (Crustacea, Amphipoda): understanding the links between physiological and behavioural responses. Aquat. Toxicol. 140–141 (2013) 106–116.
- [10] D. Zerbino, E. Birney, Velvet: algorithms for *de novo* short read assembly using de bruijn graphs. Genome Res. 18 (2008) 821–829.
- [11] P. Sornom, E. Gismondi, C. Vellinger, S. Devin, J.F. Féraud, J.N. Beisel, Effects of sublethal cadmium exposure on antipredator behavioural and antitoxic responses in the invasive amphipod *Dikerogammarus villosus*. PLoS one 7 (8) (2012) e42435.