Marine Genomics 18 (2014) 147–150

Contents lists available at ScienceDirect

Marine Genomics

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

Genomics/Technical resources

Liver transcriptome characterization of the endangered freshwater silverside *Basilichthys microlepidotus* (Teleostei: Atherinopsidae) using next generation sequencing

Caren Vega-Retter *, David Véliz *

Departamento de Ciencias Ecológicas, Instituto de Ecología y Biodiversidad, Núcleo Milenio de Ecología y Manejo sustentable de Islas Oceánicas, Universidad de Chile, Las Palmeras 3425, Ñuñoa, Santiago, Chile

ARTICLE INFO

Article history:
Received 11 September 2014
Received in revised form 14 October 2014
Accepted 14 October 2014
Available online 25 October 2014

Keywords:
Freshwater fish
Conservation
RNA-Seq
Transcriptome

ABSTRACT

Twenty eight out of 44 Chilean freshwater fishes have been considered to fall within the vulnerable or threatened category. Thus, information about the biology, ecology and the effects of the anthropic activity on these species is fundamental to take appropriate conservation measures. In Chile the endemic silverside *Basilichthys microlepidotus* inhabits mainly rivers surrounded by cities; during the last 10 years it was categorized as an endangered species, thus more basic information is needed in order to elaborate a plan of protection for this species. To this end, the transcriptome of *B. microlepidotus* was sequenced, assembled and characterized. A total of 7.8 million reads (1.05 Gb) were obtained from the sequencing and 5.93 million reads (0.83 Gb) were used for the de novo assembly, obtaining a total of 31,523 contigs. Of these, 13,724 contigs with expression in all the individuals used were retained for the functional annotation. 7938 sequences were successfully annotated; the biological processes class was the most highly represented, followed by molecular function and cellular component. These sequence data provide a useful new molecular resource for future studies on gene expression and the effects of the human activity on *B. microlepidotus*, which will facilitate obtaining more information about that, as well as the developing of appropriate conservation strategies for this species.

1. Introduction

Chilean freshwater systems have a reduced number (44) of native fish species; 64% of them have been considered to fall within the vulnerable or threatened category (Vila et al., 2006). The main factors responsible for this situation are habitat fragmentation, invasive species and pollution, all of them produced by human activities. Knowledge of the biology and ecology of these fishes is limited (Habit et al., 2006; Vila et al., 2006), thus studies analyzing the effects of anthropic activity on native species are fundamental to take appropriate conservation measures for each species.

Basilichthys microlepidotus is an atherinopsid endemic to Chile that inhabits lakes and rivers from 28°S to 39°S (Quezada-Romegialli et al., 2010; Veliz et al., 2012). It is a microphagous species, feeding on insect larvae, small invertebrates, filamentous algae and detritus (Duarte et al., 1971). It has been pointed out that it can survive in highly polluted rivers (Vega-Retter et al., 2014). Considering that *B. microlepidotus* is indicated as an endangered species (Vila et al., 2006), future conservation measures will need information about its health, stress responses and adaptive responses to human activity.

Transcriptomics studies using Next-Generation Sequencing generate a large amount of data that contribute to the understanding of how species interact with their environment and their response to the current environmental change (Vera et al., 2008). The aim of this study was to characterize the liver transcriptome of *B. microlepidotus* in order to facilitate future studies on gene expression and the effects of the human activity, and the development of appropriate conservation strategies for this species.

2. Data description

2.1. Sample collection and sequencing

Three individuals of *B. microlepidotus* were collected in the Maipo River basin; the liver tissues were transported in RNA-later (Life Technologies) to the laboratory. RNA extraction and purification were performed with the PureLink™ RNA Mini Kit (Ambion) and the MicroPoly(A) Purist™ kit (Ambion), respectively. Total RNA was checked using an Agilent Model 2100 Bioanalyzer at OMICS Solutions (Santiago, Chile). Three separate barcoded libraries were constructed with the Ion Total RNA-Seq Kit v2 (Life Technologies) and sequenced in an Ion Torrent platform using the Ion 318 chip in OMICS Solutions (Santiago, Chile). Short read and quality filtration were performed with PRINSEQ (Schmieder and Edwards, 2011) and TRIMMOMATIC (Bolger et al., 2014) software. More details are given in the Supplementary methods.

* Corresponding authors.
E-mail addresses: carenvega@ug.uchile.cl (C. Vega-Retter), dveliz@uchile.cl (D. Véliz).

http://dx.doi.org/10.1016/j.margen.2014.10.004
1874-7787/

Table 1
Summary of the data obtained from the Ion Torrent sequencing of *B. microlepidotus*, trimming process, of de novo assembly and from the contigs retained for annotation.

	Total
Obtained from sequencing	
Reads (million reads)	7.8
Nucleotides (Gb)	1.05
Mean read length (bp)	141.8
After trimming for de novo assembly	
Reads (million reads)	5.93
Nucleotides (Gb)	0.83
Mean read length (bp)	146.8
Shortest sequence length (bp)	60
Longest sequence length (bp)	250
De novo assembly	
Contigs	31,523
Non-redundant contigs	31,285
Shortest contig length (bp)	200
Longest contig length (bp)	6022
Mean contig length (bp)	593.43
N50	653
Contigs retained for annotation	
Contigs	13,724
Shortest contig length (bp)	200
Longest contig length (bp)	6022
Mean contig length (bp)	836.8

2.2. Transcriptome assembly

A total of 7.8 million reads were obtained from the sequencing performed. After the trimming process 5.93 million reads were retained for the de novo assembly performed with the MIRA assembler (Cheveruex et al., 1999). 31,523 contigs were obtained; 31,285 of these were non-redundant contigs. To have a representative set of the liver contigs of *B. microlepidotus*, reads of each individual were mapped back to the assembled transcriptome using the alignment program TMAP (http://github.com/iontorrent/TMAP/tarball/tmap.0.3.7) (for more details see Supplementary methods) and contigs showing expression in the three individuals were chosen. In total 13,724 contigs (Supplementary information 1) with an average length of 836.8 bp were retained for the functional annotation (Table 1; Fig. 1A). The raw sequence data is accessioned in the NCBI Sequence Read Archive (SRA accession SRP046041).

2.3. Functional annotation

The Blastx function was performed with a minimum E-value score of 1.0E−06 and the gene ontology (GO) terms of molecular function, cellular component, and biological process were assigned to the 13,724 retained contigs using the Blast2GO software (Conesa et al., 2005). A total of 2803 sequences presented Blast results and 7938 (57.8%) sequences were successfully annotated (Fig. S1, Supplementary information

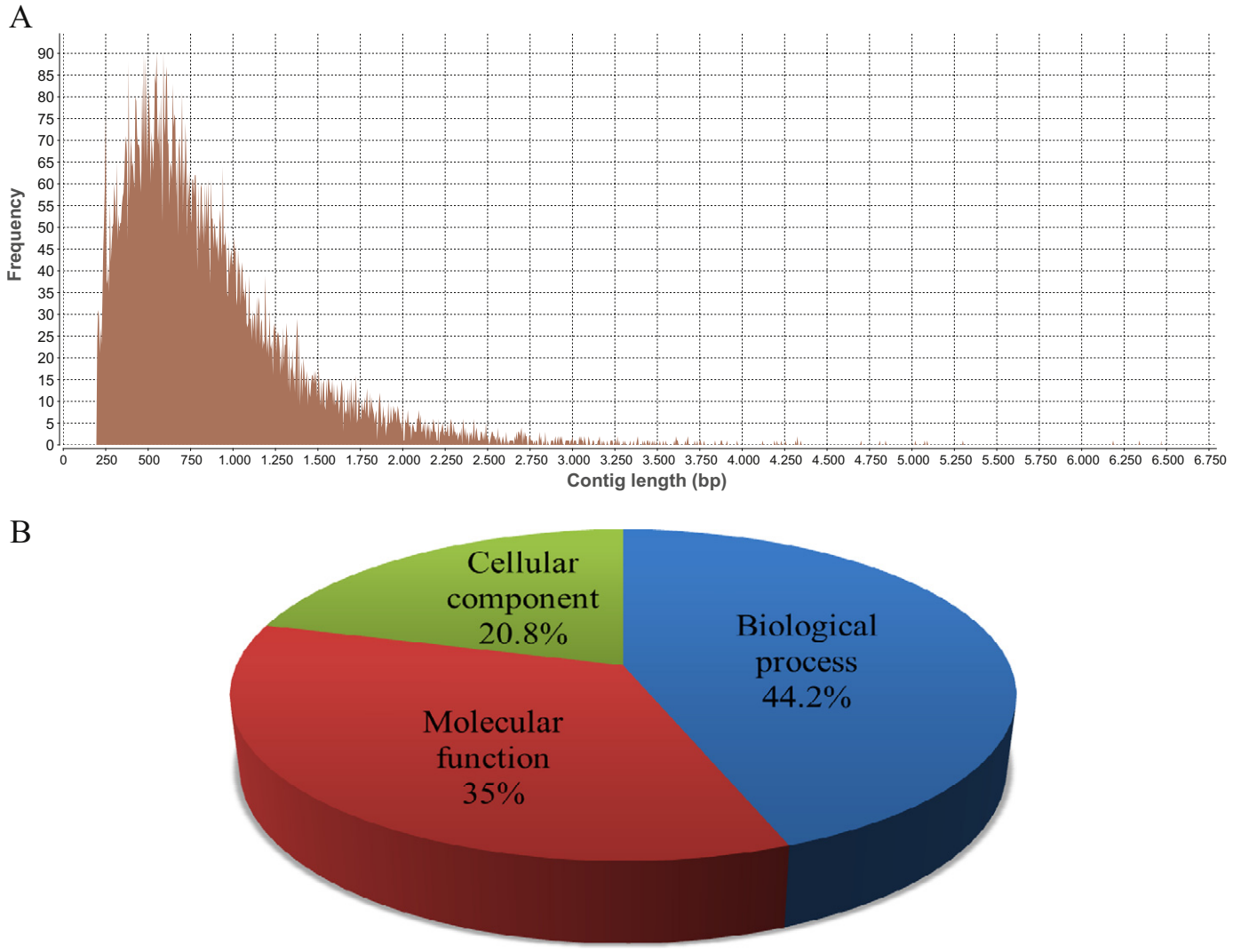

Fig. 1. Length statistics of the contigs with expression in the three individuals of *B. microlepidotus* (A). Of the 7938 genes with GO annotations, 44.2% belonged to biological processes while 35% and 20.8% were annotated as molecular function and cellular components, respectively (B).

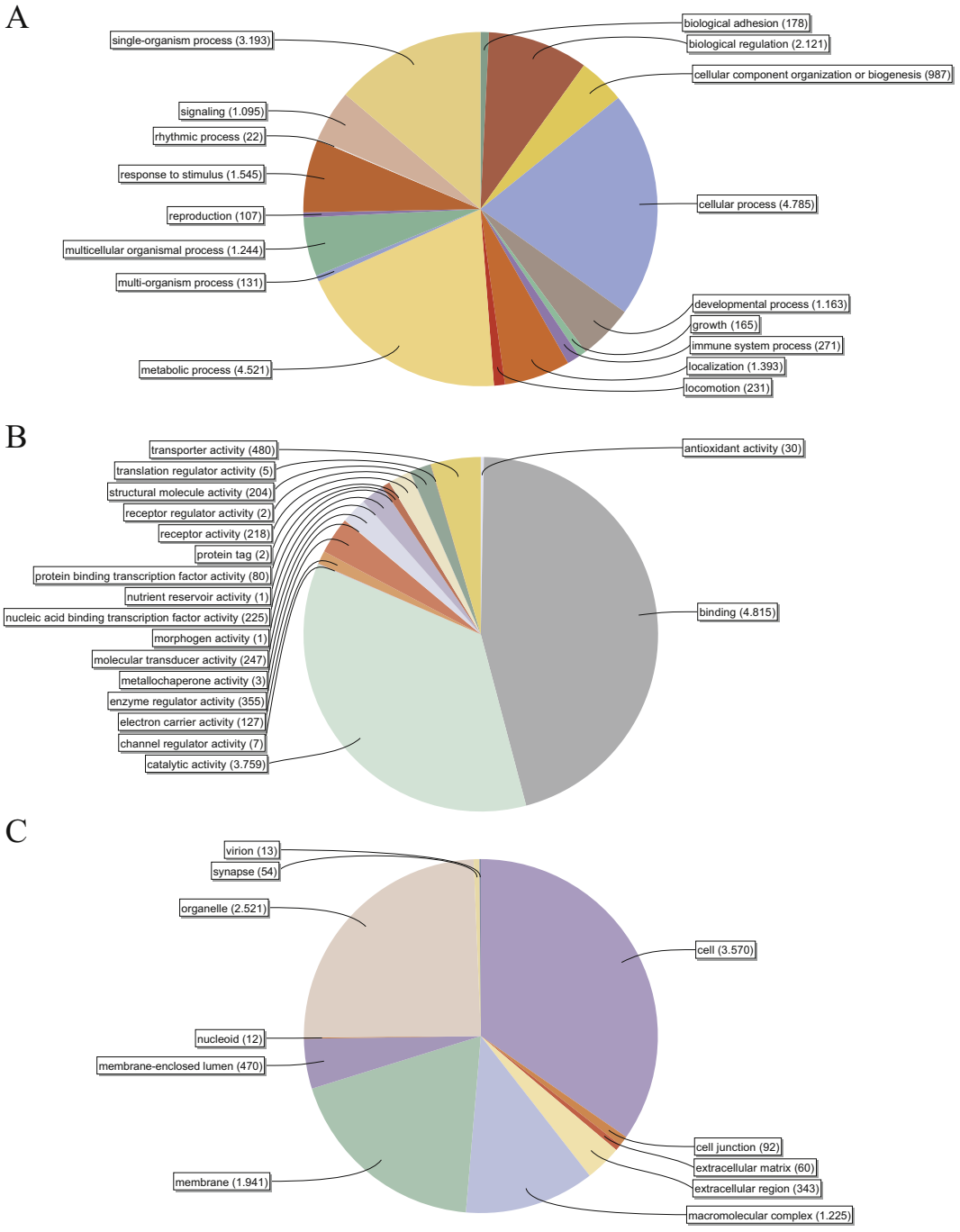

Fig. 2. Gene ontology annotations based on Blast2GO analysis (level 2). Biological process (A), molecular function (B), cellular component (C).

2). As expected, the species distribution of the Blast hits showed that most hits correspond to fish species (Fig. S2, Supplementary information 2). A total of 40,814 annotations (Supplementary information 3) for the 13,724 contigs were obtained; the biological processes class was the most highly represented (44.2%), followed by molecular function (35%) and cellular component (20.8%) (Fig. 1B). These proportions were similar to these described for *Oncorchynchus mykiss* (Fox et al., 2014). For *B. microlepidotus*, the biological processes involved mainly the diversity of gene expression, with predominance of cellular, metabolic and single-organism processes (Fig. 2A), while the GO annotations for molecular functions were mostly represented by binding and catalytic activity (Fig. 2B). The cellular component class was mainly composed of cell, organelle, membrane and macromolecular complex components (Fig. 2C). See Supplementary methods for details regarding functional annotation.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.margen.2014.10.004.

Acknowledgments

We thank C Quezada-Romegialli, JP Oyanedel and P Muñoz-Rojas for support during field work and Dr. Arne Nolte for support during the analyses. The authors thank R Espejo and Omics-Solutions Chile for sequencing. DV thanks Basal Grant PFB 023, ICM P05-002 and Nucleo Milenio NC120030; CVR thanks Conicyt Doctoral Fellowship 21090188 and doctoral thesis fellowship 24121005. All analyses were conducted in Chile and complied with its existing laws (Resolución Exenta No. 3329 Subsecretaria de Pesca).

References

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina Sequence Data. Bioinformatics 30, 2114–2120.

Cheveruex, B., Wetter, T., Suhai, S., 1999. Genome sequence assembly using trace signals and additional sequence information. Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB). 99, pp. 45–56.

Conesa, A., Götz, 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.

Duarte, W., Feito, R., Jara, C., Moreno, C., Orellana, A.E., 1971. Ictiofauna del sistema hidrográfico del río Maipo. Bol. Mus. Nac. Hist. Nat. 32, 227–268 (in Spanish).

Fox, S.E., Christie, M.R., Marine, M., Priest, H.D., Mockler, T.C., Blouin, M.S., 2014. Sequencing and characterization of the anadromous steelhead (*Oncorchynchus mykiss*) transcriptome. Mar. Genomics 15, 13–15.

Habit, E., Dyer, B., Vila, I., 2006. Estado de conocimiento de los peces dulceacuícolas de Chile. Gayana 70, 100–113.

Quezada-Romegialli, C., Fuentes, M., Veliz, D., 2010. Comparative population genetics of *Basilichthys microlepidotus* (Atheriniformes: Atherinopsidae) and *Trichomycterus areolatus* (Siluriformes: Trichomycteridae) in north central Chile. Environ. Biol. Fish 89, 173–186.

Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27, 863–864.

Vega-Retter, C., Muñoz-Rojas, P., Vila, I., Copaja, S., Veliz, D., 2014. Living in a highly polluted environment: the case of the silverside *Basilichthys microlepidotus* (Teleostei: Atherinopsidae) in the Maipo River basin, central Chile. Popul. Ecol. 56, 569–579.

Veliz, D., Catalan, L., Pardo, R., Acuña, P., Diaz, A., Poulin, E., Vila, I., 2012. The genus *Basilichthys* (Teleostei: Atherinopsidae) revisited along its Chilean distribution range (21° to 40° S) using variation in morphology and mtDNA. Rev. Chil. Hist. Nat. 85, 49–59.

Vera, J.C., Wheat, C.W., Fescemyer, H.W., Frilander, M.J., Crawford, D.L., Hanski, I., Marden, J.H., 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Mol. Ecol. 17, 1636–1647.

Vila, I., Pardo, R., Dyer, B., Habit, E., 2006. Peces límnicos: diversidad, origen y estado de conservación. In: Vila, I., Veloso, A., Schlatter, R., Ramírez, C. (Eds.), Macrófitas y vertebrados de los sistemas límnicos de chile. Editorial Universitaria, Santiago, pp. 73–102 (in Spanish).