The coelomic epithelium transcriptome from a clonal sea star, *Coscinasterias muricata*

Jonatan L Gabre[a], Pedro Martinez[b,d], Helen Nilsson Sköld[c], Olga Ortega-Martinez[e], Josep F Abril[b,f,*]

[a] Department of Chemistry and Molecular Biology, University of Gothenburg, Gothenburg, Sweden

[b] Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Barcelona, Catalonia, Spain

[c] Sven Lovén Centre of Marine Sciences Kristineberg, University of Gothenburg, Kristineberg, Fiskebäckskil, Sweden

[d] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Catalonia, Spain

[e] Biological and Environmental Science Department, University of Gothenburg, Sweden

[f] Institut de Biomedicina de la Universitat de Barcelona (IBUB), Barcelona, Catalonia, Spain

* Corresponding author: Josep F. Abril (jabril@ub.edu)

Abstract

Coscinasterias is a cosmopolitan genus of large asteroid sea stars with the ability of somatic fission as a clonal reproductive strategy. During fission, the animals tear themselves apart across their central disc, where after the lost body parts are regenerated. Here, we have sequenced and subsequently analysed the transcriptome of the coelomic epithelium of a clonal *C. muricata* specimen from New Zealand. Out of the total 389,768 raw reads, 11,344 contigs were assembled and grouped into functions. Raw read and assembled contig sequences are available at NCBI (BioSample: SAMN03371637), while the annotated assembly can be accessed through the project transcriptome browser (compgen.bio.ub.edu/gbrowse/starfish_transcriptome/). Our data is valuable for future detailed exploration of the coelomic epithelium functions as well as for a better understanding of sea star physiology.

Keywords

Pyrosequencing, RNA-Seq, Transcriptome, Echinoderm, Sea Star, Stem Cells, Regeneration, Clonal

---

1. Introduction

Coscinasterias is a cosmopolitan genus of large and fissiparous asteroid sea stars that are common in shallow waters at many continents (Waters and Roy, 2003). Sea stars are exclusively marine deuterostomes with features of special interest such as their key stone ecological functions, innate immunity, tissue regeneration and clonal potential. One species in this genus, *Coscinasterias muricata* from New Zealand is particularly interesting because it reproduces only sexually in certain geographical areas and mostly by cloning in others. Upon

fission, any of these sea stars tears itself apart across the central disc to form two or more separate pieces. Thereafter, the wounds heal and missing body parts regenerate. The coelomic epithelium is a tissue layer that covers the dorsal inside of sea stars (see Fig. 1a). It responds to several stimuli by extensive cell proliferation and it is involved in a range of important processes such as wound healing, regeneration and haematopoiesis (Holm et al. 2008, Hernroth et al. 2010). Despite the obvious wide relevance of this tissue, a molecular survey of its transcriptome has not been conducted, limiting deeper analysis of its functions. In order to provide a larger toolbox for further molecular analysis of sea stars and the coelomic epithelium in particular, we here provide de novo sequencing and subsequent annotation of the transcriptome of the coelomic epithelium from this clonal sea star, *C. muricata*.

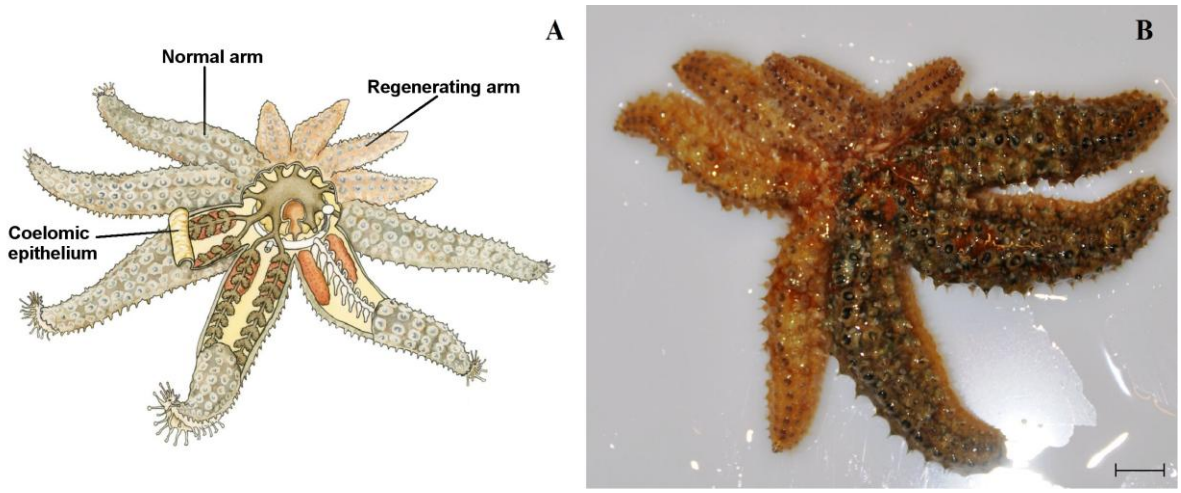

Figure 1. A) Diagram of relevant anatomical structures of *Coscinasterias muricata.* B) Photograph of an individual sea star collected from the clonal population in Otago Harbour, New Zealand. Coelomic epithelium from normal arms (darker arms to the right) was used to isolate the RNA used in this study. Scale bar 1 cm.

2. Data description

2.1 Study object and sample preparation

The analysed sea star of the species *C. muricata*, was originally collected from a highly clonal population off Portobello Aquarium in Otago Harbour, New Zealand (Sköld et al.

2003) and brought to Sven Lovén Centre of Marine Sciences at Kristineberg, University of Gothenburg, Sweden. The animals were kept at 15°C in running sea water (PSU 33) and fed *ad libitum* with small blue mussels. Spontaneous fission was observed during their maintenance and followed by regeneration of missing arms. The individual sampled for this study (Fig. 1b) had a diameter of 10 cm and was undergoing regeneration after fission; the presence of eggs indicated that it was a female. Coelomic epithelium was collected using sterile forceps from arms of normal length of the sea star.

Tissue was frozen in liquid nitrogen and homogenised with glass pestles. Total RNA extraction was performed with the RiboPure kit (Applied Biosystems, Foster City, CA, USA), following manufacturer´s protocol. Concentration was determined photometrically at 260 nm (NanoDrop ND-1000, Seqlab, Erlangen, Germany) and purity was estimated using the Agilent 2100 Bioanalyzer system (Agilent, Santa Clara, CA, USA).

2.2 Transcriptome assembly

A number of 389,768 single end sequences were obtained from a 454 Titanium run, sequencing conducted by Amplicon Express (Pullman, WA, USA). Trimmomatic version 0.32 (www.usadellab.org/cms/?page=trimmomatic) was used to remove adapters and 454 primer sequences, as well as low quality sequences, leaving a total of 346,201 sequences. These were evaluated using FastQC version 0.11.2 (Babraham Institute, www.bioinformatics.babraham.ac.uk/projects/fastqc/) and later assembled by SOAPdenovo version 1.03 (Beijing Genomics Institute, soap.genomics.org.cn/SOAPdenovo-Trans.html) with a kmer length of 63, resulting in 11,369 contigs (Fig. 2, upper panel). To check for contamination, contigs were compared with NCBI-BLASTN against a database made of the genomic sequences from the NCBI viral, bacterial, and fungal databases (GenBank and genomes). In total, 25 sequences were removed, leaving a final number of 11,344 contigs for

the functional annotation step. Out of those, 2,924 were longer or equal than 500 bp (Fig. 2) with an average length of 933 bp.

Sequence Stats	#Seqs	Total.Length	Avg.Length
Raw Reads	389,768	147,624,619bp	378.75bp
Trimmed and Clean Reads	346,201	88,571,400bp	255.84bp
Assembled Transcripts	11,369	5,212,378bp	458.50bp
VBF-Clean Transcripts	11,344	5,204,316bp	458.77bp

Assembly Stats	#Seqs	N50	GC.Pct	Longest Sequence
All Contigs	11,344	529bp	43.61%	9,225bp
#Contigs >500bp	2,924	948bp	44.68%	

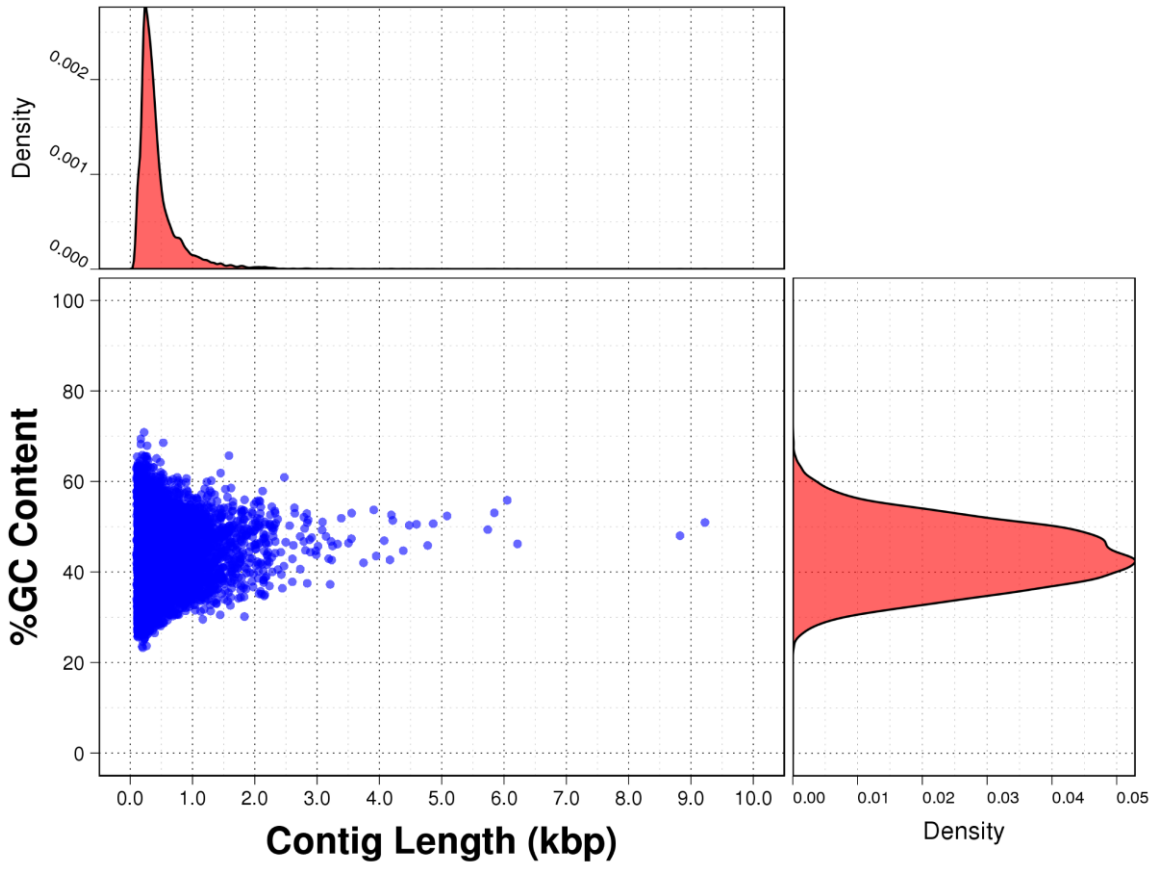

Figure 2. Assembly statistics: top table summarizes main assembly features, while the bottom figures present a scatterplot with marginal density plots that illustrate the distribution of contig lengths (in base pairs) , with values ranging from 100 to 9,225 bp, and the average GC content of each contig sequences.

2.3 Functional annotation

By running the contigs against the UniProt and NCBI non-redundant databases, using NCBI-BLASTX with an E-value cut-off of 10^{-25}, a total of 923 and 1,252 sequences were

found to have an homolog, respectively. Functional annotation was performed by mapping the Gene Ontology (GO) features of the homolog proteins (further information available in Supplementary Data S1). Figure 3 shows the most abundant terms found on the category "Molecular Function" ontology (same bar plots for "Biological Process" and "Cellular Compartment" are available on Supplementary Figs. S2 and S3 respectively). The assembled contigs were also compared, using NCBI-BLASTN and TBLASTX with identical cut-offs, against the transcriptome of the sea urchin *Strongylocentrotus purpuratus*, a closely-related sequenced organism (Sodergren et al. 2006). Furthermore, the sea urchin sequences were also functionally annotated following the same protocol resulting in 19,343 hits to UniProt and 3,449 hits to NCBI-nr. Analysis of relative abundances of GO annotations between the two transcriptomes was also performed (data not shown, Supplementary Fig. S4).

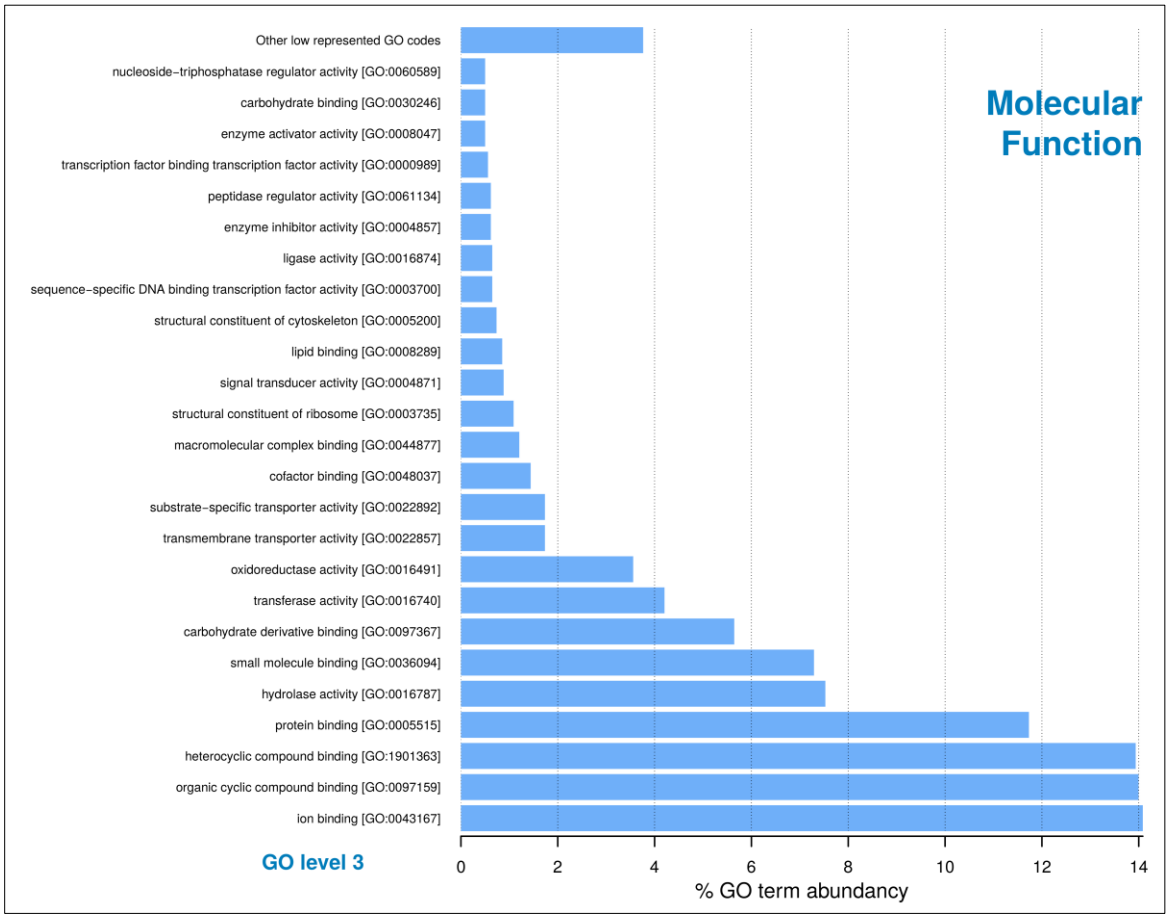

Figure 3. GO term abundancy frequencies for the Molecular Function (MF) ontology at level three, after summing up all leaves counts. Topmost 25 terms were shown, while the remaining terms, having frequencies below, were collapsed into the "Other" category.

2.4 Sequence database

Raw data sequences were uploaded into NCBI Sequence Read Archive (BioSample: SAMN03371637). Further information on assembly, supplementary data and computational protocols are available at: compgen.bio.ub.edu/Coscinasterias+transcriptome. A GBrowse has been set up to interactively browse all the annotations on the transcript sequences, available at: compgen.bio.ub.edu/gbrowse/starfish_transcriptome/.

2.5 Supplementary files

Supplementary Data S1: Excel file divided into three sheets containing **S1.a**) "BLASTX NCBInr" contains BLASTX hits when comparing against NCBI non-redundant database, together with annotations and characterization of each hit based on the description provided by the target sequence. **S1.b**) "BLASTX UniProt" contains BLASTX hits from UniProt. **S1.c**) "Assembly Coverage" contains statistics about the reads that map to the contigs, as well as the annotation of the best BLAST hits for those contigs.

Supplementary Figure S2: Bar plot showing the topmost frequent terms for the Biological Process (BP) functional ontology, which were annotated by homology search over the current sea star transcriptome.

Supplementary Figure S3: Bar plot showing the topmost frequent terms for the Cellular Component (CC) functional ontology, which were annotated by homology search over the current sea star transcriptome.

Supplementary Figure S4: Bar plot summarising the set of ontology terms that are significantly over-represented (hypergeometric test, p-value<10^{-5}) when comparing the transcriptomes of *C. muricata* and sea urchin (*S. purpuratus*).

Acknowledgements

The study was financially supported by Ministerio de Ciencia e Innovación "BFU2009-09102" (JFA); Generalitat de Catalunya "2009-SGR-001018" (JFA); Swedish Royal Academy of Sciences, Swedish Research Council (VR-NT), "621-2005-4605" (HS); ERASMUS and Lars Hiertas Minne (JG). The visit of PM to the Sven Lovén Centre was funded by the EU through the program ASSEMBLE. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors want to emphasize that the collection and rearing of animals were done following the standards of ethical behavior. They also declare that they do not have conflicts of interest.

References

Hernroth, B., F. Farahani, G. Brunborg, S. Dupont, A. Dejmek and H. N. Skold (2010). "Possibility of mixed progenitor cells in sea star arm regeneration." J Exp Zool B Mol Dev Evol **314**(6): 457-468.

Holm, K., S. Dupont, H. Skold, A. Stenius, M. Thorndyke and B. Hernroth (2008). "Induced cell proliferation in putative haematopoietic tissues of the sea star, *Asterias rubens* (L.)." J Exp Biol **211**(Pt 16): 2551-2558.

Sköld, M., S. R. Wing and P. V. Mladenov (2003). "Genetic subdivision of a sea star with high dispersal capability in relation to physical barriers in a fjordic seascape." Marine Ecology Progress Series **250**: 163-174.

Sodergren, E., G. M. Weinstock, E. H. Davidson, R. A. Cameron, R. A. Gibbs, R. C. Angerer, et al [Sea Urchin Genome Sequencing Consortium] (2006). "The genome of the sea urchin *Strongylocentrotus purpuratus*." Science **314**(5801): 941-952.

Waters, J. and M. Roy (2003). "Global phylogeography of the fissiparous sea-star genus *Coscinasterias*." Marine Biology **142**(1): 185-191.