Accessing transcriptomic data for ecologically important genes in the goose barnacle (*Pollicipes pollicipes*), with particular focus on cement proteins

A. Perina[a], B.M. von Reumont[b], A. Martínez- Lage[a], A.M. González-Tizón[a]

[a] Department of Cell and Molecular Biology, Evolutionary Biology Group (GIBE), Universidade da Coruña, A Fraga 10, E-15008 A Coruña, Spain

[b] Department of Life Sciences, The Natural History Museum, London, United Kingdom

Marine Genomics Volume 15, pages 9–11, June 2014
Received 23 December 2013, Revised 19 February 2014, Accepted 19 February 2014, Available online 4 March 2014

How to cite:
Perina, A., von Reumont, B.M., Martínez- Lage, A., González-Tizón, A.M., 2014. Accessing transcriptomic data for ecologically important genes in the goose barnacle (*Pollicipes pollicipes*), with particular focus on cement proteins. Mar. Genom. 15, 9-11. https://doi.org/10.1016/j.margen.2014.02.003

Abstract

In this study 4310 expressed sequence tags (ESTs) were used to identify potentially useful transcripts for future studies in the gooseneck barnacle *Pollicipes pollicipes* (Gmelin, 1789). 119 ESTs were obtained in this work and 4191 were taken from Meusemann et al. (2010). The gooseneck barnacle is a sessile pedunculate cirripede of great economic importance that occurs in dense aggregations, and is harvested for human consumption. The assembly of these ESTs yielded 1805 unigenes (461 contigs and 1344 singlets). The identification of cement proteins in our data is particularly interesting for cirripedes. Only a small part of the assembled unigenes could be functionally annotated. However, our results greatly improve our understanding of the biological features of *P. pollicipes*. In addition to this, a large number of potentially interesting genes were identified in order to serve as the base for future evolutionary studies in *P. pollicipes*.

Keywords: Annotation; Cement proteins; Expressed sequence tags (ESTs); *Pollicipes pollicipes*

Short introduction

The gooseneck barnacle *Pollicipes pollicipes* (Gmelin, 1789) (Crustacea: Pedunculata) is a sessile pedunculate cirripede occurring in dense aggregations exposed to heavy swell on rocky intertidal sites on the north-eastern Atlantic coast from Dakar in Senegal (15°N) to the northern coast of Brittany in France (48°N) (Barnes, 1996). These barnacles represent an important economic resource in Spain, where they are considered a delicacy. They are harvested for human consumption by a specialized branch of local fishermen, named “percebeiros”. The consumption of goose barnacles is a tradition that reaches back to the Early Holocene, as evidence of it has been found in SW Europe from the Mesolithic (about 8000 BP), and

Early Neolithic (about 6000 BP) (Álvarez-Fernández et al., 2010). The evolution of the Class Thecostraca, in which cirripedes form one group, is still unclear (Pérez-Losada et al., 2009). Besides, the evolutionary origin of the cirripede crustacean lineage remains ambiguous (Meusemann et al., 2010), and the most recent phylogenomic studies on pancrustaceans add new transcriptome data on several crustaceans (von Reumont *et al*., 2012; Oakley *et al*., 2013). The number of sampled crustacean groups remains low, with the majority of EST data representing economically important species, such as Antarctic krill *Euphausia superba* (De Pittà et al., 2008), Pacific white shrimp *Litopenaeus vannamei* (Gorbach *et al*., 2009; Li *et al*., 2012), porcelain crab *Petrolisthes cinctipes* (Stillman *et al*., 2006; Tagmount *et al*., 2010) and, *Daphnia pulex* (Colbourne et al., 2012).

The main goal of the present study was to generate a preliminary EST databank of *P. pollicipes*. We constructed an in-house developed EST library which was merged with the EST data obtained by Meusemann et al. (2010). In our view, this databank will be highly useful in further studies focusing on specific genes, e.g. secreted proteins of the cement glands, or more generally to provide useful genomic information on *P. pollicipes*.

Data description

Adult *P. pollicipes* were collected from O Roncudo (43° 15′ 51″N, 8° 58′ 45″W) (Corme, Galicia, Spain) and stored in RNA*later*® (Life Technologies). Total RNA was extracted with Aurum™ Total RNA Mini Kit (BioRad) from part of the foot tissue. The Creator™ SMART™ cDNA Library Construction Kit (Clontech) was used with minor adaptations to create a cDNA library with full-length insertion. The cDNA was ligated to pDNR-LIB vector and then transformed into *Escherichia coli* (TOP10). Recombinant white colonies were randomly picked out and amplified by PCR using M13 primers. PCR products were visualized on 1% agarose gels to ensure quality of amplification. Amplicons were directly sequenced after being purified with ExoSAP-IT (USB). Sequencing reactions were carried out using both M13 Forward and M13 Reverse primers in a capillary DNA sequencer (3130xl Genetic Analysis System, Applied Biosystems).

Electropherograms were quality controlled using Geneious Pro 5.4.6 (Drummond et al., 2011). Finally 119 high-quality ESTs remained for analysis and were deposited in GenBank (accession nos. HG792878–HG792996). In addition to this, 4191 ESTs of *P. pollicipes* (Meusemann et al., 2010) were included into analysis (accession nos. FN243242–FN247432). Both EST sets were assembled into unigenes using the iAssembler software (Zheng et al., 2011) that employs MIRA (Chevreux et al., 2004) and CAP3 (Huang and Madan, 1999) to generate initial assemblies. The individual ESTs were assembled into unigenes including contigs and singlets with minimum overlap of 30 bp and minimum percent identity of 97. Resulting unigenes were translated into amino acids following the pipeline described in von Reumont et al. (2013). BLAST2GO software (Conesa et al., 2005) was used to conduct BLAST (nr database, blastP, e-value = 0.001), domain prediction (InterProScan), and GO-term annotation (Biological Process, Cellular Component and Molecular Functions, where each unigene may be annotated in any or all three categories). All analyses were performed in May 2013.

A total of 4310 ESTs were used and assembled in this study. Sequences ranged from 100 to 1068 bp in length (mean: 506 bp). The mean length of the 1805 assembled unigenes (461 contigs and 1344 singlets) was 922 bp.

The BLASTp search against the non-redundant protein database (nr) returned 4.66% of *with BLAST results* (Fig. S1). A 19.55% of the unigenes had at least one GO term assigned (see additional file, Figs. S1–S2: general data distribution). Besides, most sequences are found *without BLAST results or hits* (70%) (Fig. S1) and, therefore, we estimate that 2056 of the 2937 predicted proteins used in this study had not been previously described. Most top-BLAST matches represent a diversity of arthropod taxa. *P.*

pollicipes sequences were very similar to *D. pulex* Leydig, 1860, *Tribolium castaneum* (Herbst, 1797), *Nasonia vitripennis* (Walker, 1836), and *Pediculus humanus* Linnaeus, 1758 among others. However, there is a large variety of blast-matches that were grouped in "non-arthropod" species. This is probably due to the limited available information on the gene background of crustaceans and arthropods (Li et al., 2012).

In total, 3569 GO terms were allocated for sequences. Functional annotation of the genes from the *Pollicipes* library indicated that the highest percentage of GO terms was seen in the Biological Process category with 1743 GO terms (49%), 1068 terms (30%) corresponding to a Cellular Component, and 757 terms (21%) to a Molecular Function. GO terms assigned are shown in the additional file, Fig. S3.

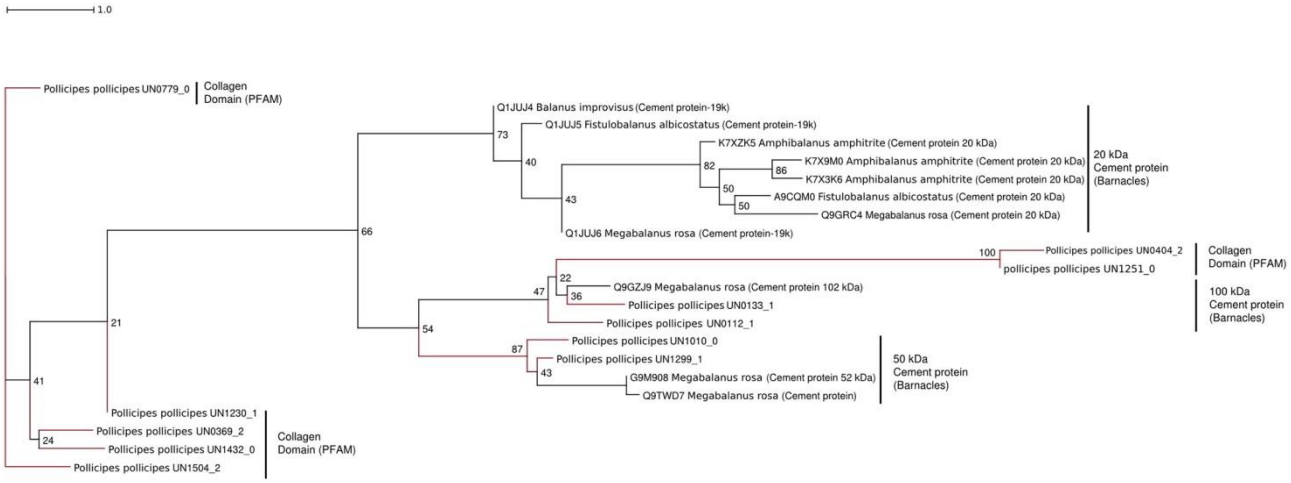

Fig. 1. Phylogenetic tree of putative cement proteins. Phylogenetic tree of putative cement proteins reconstructed in raxmlHPC-PTHREADS-SSE3, with 1000 BS replicates. Favored substitution model chosen by Prottest3 was PROTGAMMAIWAG (-f a). All available cement protein sequences of Crustacea in Uniprot were used in the pipeline to identify putative sequences (von Reumont et al., 2013) in our data and all significant sequence hits (e-value = 0.00001) were included in the alignment.

This study provides some of the first insights and represents a base for further studies on gene expression and protein pathways in goose barnacles. We used the databank developed in this study to investigate one particularly important adaptation for sessile living in *P. pollicipes*, cement gland proteins, only recently studied in goose barnacles. RNA used for this work was extracted from body tissue and foot tissue of adult individuals. We identified several protein transcripts of the cement glands, which are secreted in the foot tissue to attach to the substratum. Specifically, we have discovered two 100 kDa and 52 kDa cement protein transcripts in our data set which cluster with cement protein sequences of *Balanus amphitrite* and *Megabalanus rosa*, respectively (see phylogenetic tree, Fig. 1). Barnacle cement proteins are classified into two types, a primary cement protein that is produced while the barnacle attaches to the substratum, and the secondary cement protein that is secreted to aid barnacle's reattachment (Saroyan *et al*., 1970; Chen *et al*., 2011). We identified in our data at least three clades of transcripts, which partly share similarities to known cement proteins of crustaceans (He et al., 2013), but also different sequences that need further investigation (Fig. 1). Comparing transcriptomes of whole bodies and larvae of *P. pollicipes* could contribute to the understanding of the complexity of their ontogenetic adaptation to a sessile mode of life and the evolution of cement proteins in cirripeds.

EST generation and identification of specific genes of *P. pollicipes* provide a more general understanding of these crustaceans. The only small number of genes that could be functionally annotated in this study

indicates that our knowledge about goose barnacle physiology and biological processes is insufficient. The analysis of the fraction of identified unigenes already highlights a large number of genes that are of interest for future research concerning protein evolution (with focus on cement gland proteins) and physiology (involving adaptational and ontogenetic processes).

Acknowledgements and funding

We thank Iago F. Meilán for computer support. This work was funded by a CTM2007-62034 grant from the Spanish government (Ministerio de Educación y Ciencia) and, a 10MMA103008PR grant by Xunta de Galicia. A. Perina was supported by a scholarship from Ministerio de Economía y Competitividad, Subprograma de Formación de Personal Investigador (FPI) (Spain). B.M. von Reumont was funded by the German Science Foundation (DFG grants: RE 3454/1-1 and RE 3454/1-2).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.margen.2014.02.003.

References

Álvarez-Fernández, E., Ontañón-Peredor, Molares-Vila, J., 2010. Archaeological data on the exploitation of the goose barnacle Pollicipes pollicipes (Gmelin, 1790) in Europe. J. Archaeol. Sci. 37, 402–408.

Barnes, M., 1996. Pedunculate Cirripedes of the genus Pollicipes. Oceangr. Mar. Biol. Annu. Rev. 34, 303–394.

Chen, Z.F., Matsumura, K., Wang, H., Arellano, S.M., Yan, X., Alam, I., Archer, J.A.C., Bajic, V.B., Qian, P.Y., 2011. Toward an understanding of the molecular mechanisms of baranacle larval settlement: a comparative transcriptomic approach. PLoS One 6, e22013.

Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A.J., Muller, W.E.G., Wetter, T., Suhai, S., 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res. 14, 1147–1159.

Colbourne, J.K., Pfrender, M.E., Gilbert, D., Thomas, W.K., Tucker, A., Oakley, T.H., et al., 2012. The ecoresponsive genome of Daphnia pulex. Science 331, 555–561.

Conesa, A., Gotz, S., Garcia-Gómez, 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.

De Pittà, C., Bertolocci, C., Mazzotta, G.M., Bernante, F., Rizzo, G., De Nardi, B., Pallavicini, A., Lanfranchi, G., Costa, R., 2008. Systematic sequencing of mRNA from the Antarctic krill (Euphausia superba) and first tissue specific transcriptional signature. BMC Genomics 9, 45.

Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T., Wilson, A., 2011. Geneious v5.6.3, Created by Biomatters. Available from http://www.geneious.com/.

Gorbach, D.M., Hu, Z.L., Du, Z.Q., Rothschild, M.F., 2009. SNP discovery in Litopenaeus vannamei with a new computational pipeline. Anim. Genet. 40, 106–109.

He, L.-S., Zhang, G., Qian, P.-Y., 2013. Characterization of two 20 kDa-cement protein (cp20k) homologues in Amphibalanus amphitrite. PLoS ONE 8 (5), e64130. http://dx.doi.org/10.1371/journal.pone.0064130.

Huang, X., Madan, A., 1999. CAP3: a DNA sequence assembly program. Genome Res. 9, 868–877.

Li, C., Weng, S., Chen, Y., Yu, X., Lü, L., Zhang, H., He, J., Xu, X., 2012. Analysis of Litopenaeus vannamei transcriptome using the next-generation DNA sequencing technique. PLoS One 7, e47442.

Meusemann, K., von Reumont, B.M., Simon, S., Roeding, F., Strauss, S., Kuck, P., Ebersberger, I., Walz, M., Pass, G., Breuers, S., Achter, V., von Haeseler, A., Burmester, T., Hadrys, H., Wagele, J.W., Misof, B., 2010. A phylogenomic approach to resolve the arthropod tree of life. Mol. Biol. Evol. 27, 2451–2464.

Oakley, T.H., Wolfe, J.M., Lindgren, A.R., Zaharoff, A.K., 2013. Phylotranscriptomics to bring the understudied into the fold: monophyletic ostracoda, fossil placement, and pancrustacean phylogeny. Mol. Biol. Evol. 30, 215–233.

Pérez-Losada, M., Høeg, J.T., Crandall, K.A., 2009. Remarkable convergent evolution in specialized parasitic Thecostraca (Crustacea). BMC Biol. 7, 15.

Reumont von, B.M., Blanke, A., Richter, S., Alvarez, F., Bleidorn, C., Jenner, R.A., 2013. The first venomous crustacean revealed by transcriptomics and functional morphology: remipede venom glands express a unique toxin cocktail dominated by enzymes and a neurotoxin. Mol. Biol. Evol. http://dx.doi.org/10.1093/molbev/mst199.

Saroyan, J.R., Lindner, E., Dooley, C.A., 1970. Repair and reattachment in the balanidae as related to their cementing mechanism. Biol. Bull. 139, 333–350.

Stillman, J.H., Teranishi, K.S., Tagmount, A., Lindquist, E., Brokstein, P.B., 2006. Construction and characterization of EST libraries from the porcelain crab, Petrolisthes cinctipes. Integr. Comp. Biol. 46, 919–930.

Tagmount, A., Wang, M., Lindquist, E., Tanaka, Y., Teranishi, K.S., Sunagawa, S., Wong, M., Stillman, J.H., 2010. The porcelain crab transcriptome and PCAD, the porcelain crab microarray and sequence database. PLoS ONE 5, e9327.

von Reumont, B.M., Jenner, R.A., Wills, M.A., Dell'Ampio, E., Pass, G., Ebersberger, I., et al., 2012. Pancrustacean phylogeny in the light of new phylogenomic data: support for Remipedia as the possible sister group of Hexapoda. Mol. Biol. Evol. 29 (3), 1031–1045. http://dx.doi.org/10.1093/molbev/msr270.

Zheng, Y., Zhao, L., Gao, J., Fei, Z., 2011. iAssembler: a package for de novo assembly of Roche-454/Sanger transcriptome sequences. BMC Bioinforma. 12, 453.