Marine Genomics xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Marine Genomics

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

Genomics/technical resources

De novo assembly and annotation of the European abalone *Haliotis tuberculata* transcriptome

Ewan Harney [a,*], Bruno Dubief [a], Pierre Boudry [b], Olivier Basuyaux [c], Markus B. Schilhabel [d], Sylvain Huchette [e], Christine Paillard [a], Flavia L.D. Nunes [a]

[a] *Laboratoire des Sciences de l'Environnement Marin (LEMAR), UMR6539 CNRS/UBO/IRD/Ifremer, Institut Universitaire Européen de la Mer, University of Brest (UBO), Université Européenne de Bretagne (UEB), Place Nicolas Copernic, 29280, Plouzané, France*
[b] *Ifremer, Laboratoire des Sciences de l'Environnement Marin (LEMAR), UMR 6539 CNRS/UBO/IRD/Ifremer, Centre Bretagne Z.I. Pointe du Diable, 29280 Plouzané, France*
[c] *SMEL (Synergie Mer Et Littoral), Centre Expérimental, 50560 Blainville-sur-Mer, France*
[d] *Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, 24105 Kiel, Germany*
[e] *France Haliotis, Kerazan Lilia, 29880 Plouguerneau, France*

ARTICLE INFO

Article history:
Received 8 February 2016
Accepted 3 March 2016
Available online xxxx

Keywords:
RNA-Seq
Mollusca
Gastropoda
Development
Vibrio harveyi
Climate change

ABSTRACT

The European abalone *Haliotis tuberculata* is a delicacy and consequently a commercially valuable gastropod species. Aquaculture production and wild populations are subjected to multiple climate-associated stressors and anthropogenic pressures, including rising sea-surface temperatures, ocean acidification and an emerging pathogenic *Vibrio* infection. Transcript expression data provides a valuable resource for understanding abalone responses to variation in the biotic and abiotic environment. To generate an extensive transcriptome, we performed next-generation sequencing of RNA on larvae exposed to temperature and pH variation and on haemolymph of adults from two wild populations after experimental infection with *Vibrio harveyi*. We obtained more than 1.5 billion raw paired-end reads, which were assembled into 328,519 contigs. Filtration and clustering produced a transcriptome of 41,099 transcripts, of which 10,626 (25.85%) were annotated with Blast hits, and 7380 of these were annotated with Gene Ontology (GO) terms in Blast2Go. A differential expression analysis comparing all samples from the two life stages identified 5690 and 10,759 transcripts with significantly higher expression in larvae and adult haemolymph respectively. This is the greatest sequencing effort yet in the *Haliotis* genus, and provides the first high-throughput transcriptomic resource for *H. tuberculata*.

1. Introduction

Coastal marine organisms are increasingly being impacted by global climate change and anthropogenic pressures (Harley et al., 2006). The study of their responses to these stressors is vital for understanding how marine populations and species are able to acclimatise or adapt to future change and for devising management strategies for species of economic interest. The European abalone *Haliotis tuberculata* is a commercially valuable gastropod, with production based on fisheries of wild populations or hatchery-based aquaculture. As is the case with many calcifying species, rising pCO_2 and subsequent ocean acidification increase the energetic cost of calcification (Hofmann and Todgham, 2010). The concurrent effects of rising sea surface temperatures and declining pH have deleterious consequences for the development of calcifying organisms (Parker et al., 2010), with abalone being particularly sensitive to changes in pH and temperature (Byrne, 2011). In addition, the emergence of abalone disease due to *Vibrio harveyi* has been associated with the warming of coastal waters (Nicolas et al., 2002). In order to better understand how the European abalone responds to abiotic and biotic stressors, global transcript expression (RNA-Seq) was examined as part of two ongoing experiments: one that is investigating the effects of pH and temperature on larval development; and the other investigating the immune response of adults to infection with *V. harveyi*. The aim of this work was to provide a detailed transcriptomic data set for *H. tuberculata*. Rather than carrying out separate transcriptome assemblies for each life stage, Illumina sequences from both experiments were combined and assembled into a single transcriptome, allowing generation of a more complete transcriptomic resource, and direct comparison of larvae and adults. This data provides an invaluable resource for identifying transcripts involved in the response of *H. tuberculata* to global change and other anthropogenic stresses, and taken with other recent high throughput transcriptomic data for *Haliotis* species (De Wit and Palumbi, 2013; Franchini et al., 2011; Huang et al.,

* Corresponding author.
E-mail address: ewan.harney@univ-brest.fr (E. Harney).

http://dx.doi.org/10.1016/j.margen.2016.03.002
1874-7787/

Table 1
Comparison of molluscan transcriptomes assembled *de novo* from Illumina sequences using Trinity and other bioinformatic assembly tools.

Species	Raw reads (millions)	Assembler	Total contigs	Filtered and/or non-redundant contigs	Mean contig length (bp)*	N50 (bp)*	Number of annot. contigs (percent)*	Reference
Haliotis tuberculata	1517	Trinity	328,519	41,099	1033	1544	10,626 (25.9)	This study
H. laevigata	38	Trinity	222,172	97,420	–	1313	20,702 (21.3)	Shiel et al., 2015
H. midae	25	CLC	22,761	–	260	356	3841 (16.9)	Franchini et al., 2011
H. rufescens	355.68	CLC	162,928	–	522	653	48,004 (29.5)	De Wit and Palumbi, 2013
Anadara trapezia	27	Trinity	75,024	63,778	505	–	29,013 (38.7)	Prentis and Pavasovic, 2014
Clio pyramidata	400	Trinity	45,739	30,800	618	852	9280 (20.3)	Maas et al., 2015
Conus tribblei	33.54	Trinity	163,513	–	513	614	21,069 (12.9)	Barghi et al., 2014
Mytilus galloprovincialis	393	Trinity	–	151,320	570	–	50,998 (33.7)	Moreira et al., 2015
Nucella lapillus	42.08	Trinity	90,674	–	–	413	20,922 (23.1)	Chu et al., 2014
Patinopecten yessoensis	55.88	Trinity	135,963	86,521	733	1266	28,228 (20.8)	Sun et al., 2015
Reishia clavigera	110.77	Trinity	197,324	151,684	499	582	28,948 (14.7)	Ho et al., 2014
Trintonia diomedia	133.1	Trinity	185,546	123,154	74	1353	18,246 (14.8)	Senatore et al., 2015
Corbicula fluminea	62.25	Oases	–	134,684	791	1264	38,985 (28.9)	Chen et al., 2013
Crassostrea gigas	2204	Oases	–	55,651	1659	–	44,912 (80.7)	Riviere et al., 2015
C. virginica	52.86	Oases	48,562	–	874	–	11,237 (23.1)	Zhang et al., 2014
Elliptio complanata	361.03	CLC	136,000	–	–	982	26,689 (19.6)	Cornman et al., 2014
Pinctada fucata	26.51	SOAP	723,797	92,964	434	–	33,882 (36.4)	Huang et al., 2013

* Contig stats and annotation reported either from total or non-redundant contigs, depending on cited example.

2012; Picone et al., 2015; Shiel et al., 2015; Valenzuela-Muñoz et al., 2014) enriches the genomic resources available for this commercially important genus.

2. Data description

2.1. Sampling and sequencing

Biological samples used for generating the *H. tuberculata* transcriptome were collected as part of two ongoing studies in our group. Combining the sequences from two life stages (larvae and adults) and two tissues (whole organism and haemolymph) enables the assembly of a more complete transcriptome which reflects a broader range of functions. Here we present the bioinformatic analysis used to assemble the *de novo* transcriptome, and provide a differential expression analysis of whole larvae versus adult haemolymph. Detailed differential expression analysis dealing with the experimental treatments performed for each study will be examined in two companion papers.

H. tuberculata larvae were produced following controlled fertilizations carried out at a commercial hatchery (France Haliotis, Plouguerneau, Finistère, France). Larvae were transferred to experimental facilities at Ifremer (Plouzané, Finistère, France) and reared in one of four treatments (control = 20 °C pH 8.1; acidified = 20 °C pH 7.9; warmed = 22 °C pH 8.1; interaction = 22 °C pH 7.9), RNA samples deriving from approximately 10,000 pooled larvae were collected by filtration from two replicated experimental tanks for each treatment at 14, 24 and 38 h post-fertilisation. These 24 samples were flash frozen in liquid nitrogen and stored at −80 °C until RNA extraction. Adult haemolymph RNA was collected during successive infection experiments on abalone from two different natural populations in Brittany (Molène and Saint-Malo, average size of individuals = 74 ± 15 mm) with the bacteria *V. harveyi* (strain ORM4), carried out during the spawning period. Haemolymph was sampled 72 h post-infection, following the first exposure for abalone from Molène and after the third exposure for abalone from Saint-Malo. Uninfected control abalones were sampled at the same time point as infected individuals for both populations. Three replicates were sampled per treatment for a total of 12 samples. Two millilitres of haemolymph was withdrawn with a 5 mL syringe from each abalone, and spun at 200 rcf for 10 min. The serum was removed, 1 mL of TRIzol reagent (Life Technologies) was added to the pellet, and then the samples were frozen and stored at −80 °C until RNA extraction.

All samples were homogenised by bead beating and total RNA was extracted using TRIzol reagent (Life Technologies) according to the manufacturer's instructions. RNA quality and concentration were determined using an Agilent 2100 RNA Bioanalyser (Agilent, Santa Clara, CA, USA) and a NanoDrop 8000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) respectively. Library preparation and sequencing of larval samples were carried out at the Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Germany, with Illumina HiSeq (101 bp paired-end, non-strand-specific), with 8 samples multiplexed per lane (two samples were resequenced later with 126 bp paired-end reads). Library preparation and sequencing of adult haemolymph RNA samples were performed by the GenePool, University of Edinburgh, Scotland, on Illumina HiSeq (100 bp paired-end, strand-specific). For adult haemolymph RNA, two runs were performed with lanes of 12 multiplexed samples. Raw sequence data is available from the NCBI sequence read archive (SRA) under the accession SRA303338.

2.2. De novo transcriptome assembly

A total of 849,307,722 and 668,045,828 paired-end reads were generated for larvae and adult haemolymph respectively. Adapters and low quality bases were trimmed using the Trimmomatic tool, with a 4 bp sliding window, cutting when the average quality score dropped below Q = 20 (Bolger et al., 2014), leaving 790,015,978 paired-end reads for larvae (mean length = 99.6 bp) and 544,556,316 paired-end reads for adult haemolymph (mean length = 88.9 bp). Separate *in silico* normalisation of kmers was carried out for larvae and adult haemolymph (Haas et al., 2013), and the transcriptome was assembled from the combined normalised reads using Trinity ver. 2.0.6 (Grabherr et al., 2011) with default parameters. Initial assembly produced 514,943 contigs of between 224 and 32,662 bp length. Removal of short contigs (≤ 300 bp) reduced this to 328,519 contigs with an average length of 825 bp and an N50 of 1103 bp. This Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GEAU00000000. The version described in this paper is the first version, GEAU01000000. The relatively large number of contigs was in part a result of sequencing effort, which was an order of magnitude greater than in many comparable molluscan *de novo* transcriptome assemblies (Table 1). Therefore to maximise confidence in our transcriptome, we carried out a number of filtrations. Initially, lowly expressed contigs were removed by applying a FPKM cut-off of 1. This filtration produced 79,374 contigs, with an average length of 1022 bp and N50 of 1509 bp. Contigs with high similarity, which likely represent different spliced isoforms or duplicated genes were grouped by Trinity into 41,099 non-redundant 'unigenes' (hereafter referred to as transcripts), with an average length of 1015 bp and an N50 of 1529 bp. This filtered transcriptome is used in subsequent

Please cite this article as: Harney, E., et al., *De novo* assembly and annotation of the European abalone *Haliotis tuberculata* transcriptome, Mar. Genomics (2016), http://dx.doi.org/10.1016/j.margen.2016.03.002

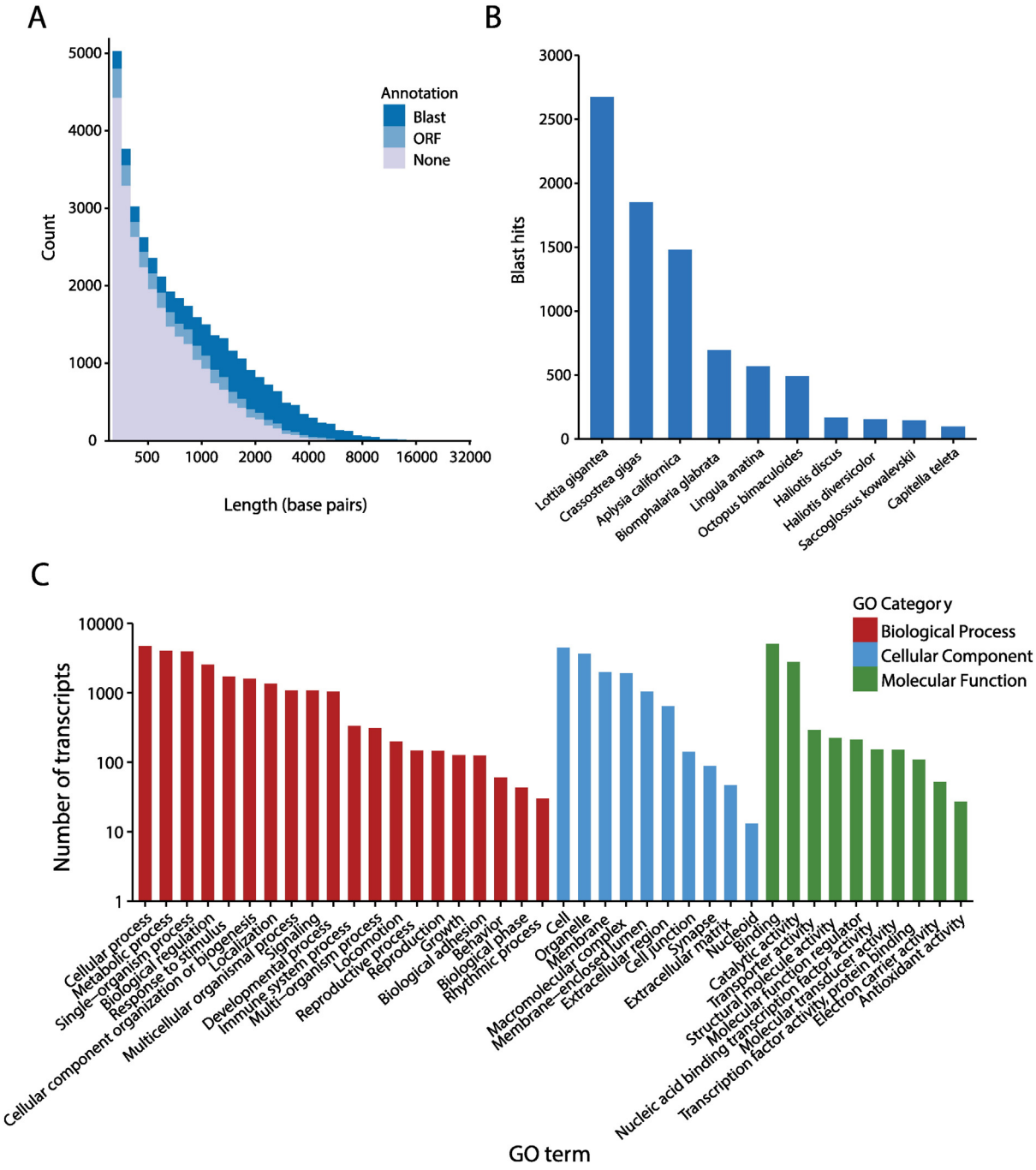

Fig. 1. A) Sequence length distribution of 41,099 filtered transcripts showing those with Open Reading Frames (ORFs) and Blastp hits in the NR database; B) species distribution of blast hits among the 10,626 annotated transcripts. Results from NR take precedence over Swiss-Prot; C) Gene Ontology (GO) classification of 7380 *H. tuberculata* Blast2GO annotated transcripts split by category. Results from Swiss-Prot take precedence over NR.

stages of annotation and analysis, and is available in the supplementary material (Appendix A).

2.3. Functional annotation

Coding sequence detection was carried out with TransDecoder (Haas et al., 2013). Among the 41,099 transcripts in the filtered transcriptome were 12,587 (30.6%) that contained at least one open reading frame (ORF) of ≥100 amino acids length. Translated ORFs were used as Blastp queries against the NR and Swiss-Prot databases, and non-translated nucleotide sequences from all transcripts were used as Blastx queries against the same databases. All blasts were carried out using an e-value cutoff of 1e−5.

A total of 10,626 transcripts (25.85%) were annotated following Blast analysis: 10,603 transcripts were associated with blast hits in the NR database, and 9347 transcripts were associated with blast hits in the Swiss-Prot database. Longer transcripts were more likely to contain an ORF and be associated with blast hits (Fig. 1A); however, our total of 10,626 annotated transcripts is comparable in terms of percentage annotation (25.9%) with other Illumina transcriptome assemblies of non-model marine molluscs without a reference genome (Table 1). When blast results from the NR database were given precedence over those from the Swiss-Prot database, the majority of hits among the 10,626 annotated transcripts derived from mollusc species, with four species accounting for over 63% of results: *Lottia gigantea* (25.22%), *Crassostrea gigas* (17.46%), *Aplysia californica* (13.96%), and

Please cite this article as: Harney, E., et al., *De novo* assembly and annotation of the European abalone *Haliotis tuberculata* transcriptome, Mar. Genomics (2016), http://dx.doi.org/10.1016/j.margen.2016.03.002

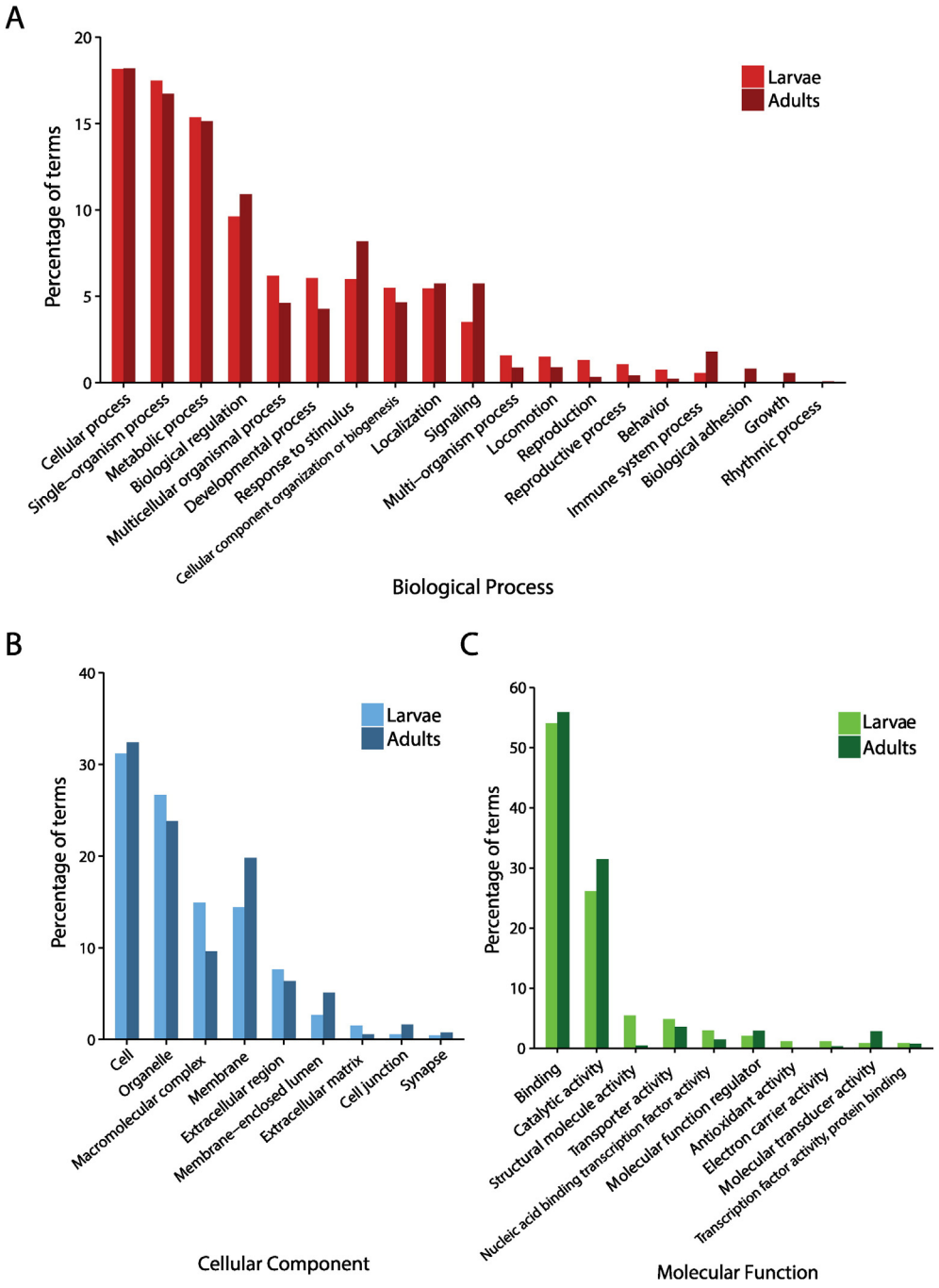

Fig. 2. Comparison of second level GO terms (percentage contribution towards the total) in differentially expressed larval and adult haemolymph transcripts, split according to GO categories of A) biological process; B) cellular component and c) molecular function.

Biomphalaria glabrata (6.55%). Furthermore, the congeners *H. discus* and *H. diversicolor* were also among the top 10 most represented species, accounting for 1.56% and 1.46% of hits respectively (Fig.1B). To estimate the influence of fragmentation on the transcriptome, non-duplicated accession numbers among blast annotated transcripts were counted (with NR results again given priority over Swiss-Prot results). A total of 9378 different accession numbers were present among the 10,626 annotated transcripts, suggesting that the large majority (88.25%) of annotated transcripts had unique identities.

Nevertheless, the number of transcripts presented here is higher than the number of genes commonly reported in marine mollusc genomes (*e.g.* 21,013 for *A. californica*, 24,676 for *L. gigantea* and 32,250 for *C. gigas*), and the majority remain unannotated. Sequencing of expressed transcripts is expected to identify many novel contigs that do not align to known gene models: up to 30–40% in organisms with reference genomes (Wang et al., 2009). Unidentified transcripts may derive from non-coding RNA (Eddy, 2001), which can represent a large part of the transcriptome of many organisms, and play an important functional role in the regulation of processes such as gene expression (Guttman and Rinn, 2012). A recent study on transcript expression in *C. gigas* by Riviere et al. (2015) demonstrates that *de novo* transcriptome assembly will often produce many novel contigs with unknown functions; in that case, over 56,000 contigs were identified including 11,817 contigs that did not match any known region of the reference genome or RefSeq protein records.

Finally, Blast2GO (Götz et al., 2008) was used to retrieve Gene Ontology (GO) terms among blast-annotated transcripts. For this analysis, blast results from the Swiss-Prot database took precedence over those

from the NR database, as they often include more functional information. As part of the annotation procedure, the InterProScan plugin of Blast2GO was used to search for functional domains (Jones et al., 2014). This step was applied to all 41,099 transcripts, revealing 16,305 transcripts that contained at least one InterPro domain (including 8265 that had not been annotated following blast analysis). For Blast2GO annotation, default parameters were applied, and subsequent GO terms were merged with InterProScan results; furthermore, the annex function of Blast2GO was used to augment GO term annotation among the transcripts. A total of 46,336 GO terms were annotated to 7380 transcripts (17.96% of total transcripts). The number of transcripts with second level GO terms is provided for each of the three GO categories of biological process, molecular function and cellular component (Fig. 1C).

2.4. Comparison of GO terms in larvae and adults

Broad differences in function between larvae and adult haemolymph were visualised by comparing relative contributions of second level GO terms in contigs differentially expressed between the life stages. Differential expression analysis was carried out in R using the package *edgeR* (McCarthy et al., 2012), and considered normalised expression values from all adult haemolymph libraries (across both populations and both treatments) against normalised expression from all larval libraries (across all stages and all treatments). Using all the libraries increased statistical replication for each life stage, and thus confidence in the identification of differentially expressed transcripts. Of the 41,099 transcripts considered, 16,449 were significantly differentially expressed between adult haemolymph and larvae (Benjamini–Hochberg FDR < 0.01, log-fold change > 5). Specifically, 5690 transcripts were significantly more expressed in larvae, and 10,759 were significantly more expressed in adult haemolymph. Among these transcripts, 1185 and 3537 were blast-annotated in larvae and adult haemolymph respectively, of which 566 and 2416 also had Blast2go annotation. Full lists of differentially expressed transcripts and their annotations are included in the supplementary material (Appendix A). For the three GO categories of biological process, cellular component and molecular function, second level GO terms for larvae and adult haemolymph were expressed as a percentage of the total number of GO terms in that category (Fig. 2). Among the biological processes, '*multicellular organismal process*' and '*developmental process*' had higher percentages in larvae compared to adults; while '*response to stimulus*', '*signalling*', '*biological regulation*' and '*immune system process*' had higher percentages in adults compared to larvae (Fig. 2A). Similarly, the cellular component '*macromolecular complex*' and molecular function '*structural molecule activity*' had higher percentages in larvae, while the cellular component '*membrane*' and molecular function '*catalytic activity*' had higher percentages in adults (Fig. 2B and C). These results provide a first glimpse into life stage specific transcript expression in *H. tuberculata*. Further analysis of function and differential expression between treatments in larvae and adult haemolymph will follow in companion papers.

Acknowledgements

This work was supported by a grant from the Regional Council of Brittany, from the European Funds (ERDF) and supported by LabexMER (ANR-10-LABX-19) and co-funded by a grant from the French Government under the programme "Investissements d'Avenir". Pierrick Le Souchu and Philippe Miner provided technical aide during larval experiments at Ifremer. Fabien Riera, Gaëlle Richard, Naïda Laisney and Stéphanie Pétinay assisted with experimental infections of adult abalones. Erwann Corre at Abims (Analysis bioinformatics for marine science, Roscoff Marine Station) provided invaluable advice on Trinity usage.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.margen.2016.03.002.

References

Barghi, N., Concepcion, G.P., Olivera, B.M., Lluisma, A.O., 2014. High conopeptide diversity in *Conus tribblei* revealed through analysis of venom duct transcriptome using two high-throughput sequencing platforms. Mar. Biotechnol. 17, 81–98. http://dx.doi.org/10.1007/s10126-014-9595-7.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. http://dx.doi.org/10.1093/bioinformatics/btu170.

Byrne, M., 2011. Impact of ocean warming and ocean acidification on marine invertebrate life history stages: vulnerabilities and potential for persistence in a changing ocean. Oceanogr. Mar. Biol. An Annu. Rev. 49, 1–42. http://dx.doi.org/10.1016/j.marenvres.2011.10.00.

Chen, H., Zha, J., Liang, X., Bu, J., Wang, M., Wang, Z., 2013. Sequencing and *de novo* assembly of the Asian clam (*Corbicula fluminea*) transcriptome using the Illumina GAIIx method. PLoS One 8, 1–12. http://dx.doi.org/10.1371/journal.pone.0079516.

Chu, N.D., Miller, L.P., Kaluziak, S.T., Trussell, G.C., Vollmer, S.V., 2014. Thermal stress and predation risk trigger distinct transcriptomic responses in the intertidal snail *Nucella lapillus*. Mol. Ecol. 23, 6104–6113. http://dx.doi.org/10.1111/mec.12994.

Cornman, R.S., Robertson, L.S., Galbraith, H., Blakeslee, C., 2014. Transcriptomic analysis of the mussel *Elliptio complanata* identifies candidate stress-response genes and an abundance of novel or noncoding transcripts. PLoS One 9, e112420. http://dx.doi.org/10.1371/journal.pone.0112420.

De Wit, P., Palumbi, S.R., 2013. Transcriptome-wide polymorphisms of red abalone (*Haliotis rufescens*) reveal patterns of gene flow and local adaptation. Mol. Ecol. 22, 2884–2897. http://dx.doi.org/10.1111/mec.12081.

Eddy, S.R., 2001. Non-coding RNA genes and the modern RNA world. Nat. Rev. Genet. 2, 919–929. http://dx.doi.org/10.1038/35103511.

Franchini, P., van der Merwe, M., Roodt-Wilding, R., 2011. Transcriptome characterization of the South African abalone *Haliotis midae* using sequencing-by-synthesis. BMC Res. Notes 4, 59. http://dx.doi.org/10.1186/1756-0500-4-59.

Götz, S., Garcia-Gomez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talon, M., Dopazo, J., Conesa, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36, 3420–3435. http://dx.doi.org/10.1093/nar/gkn176.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat. Biotechnol. 29, 644–652. http://dx.doi.org/10.1038/nbt.1883.

Guttman, M., Rinn, J.L., 2012. Modular regulatory principles of large non-coding RNAs. Nature 482, 339–346. http://dx.doi.org/10.1038/nature10887.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M.G., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A., 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494–1512. http://dx.doi.org/10.1038/nprot.2013.084.

Harley, C.D.G., Randall Hughes, A., Hultgren, K.M., Miner, B.G., Sorte, C.J.B., Thornber, C.S., Rodriguez, L.F., Tomanek, L., Williams, S.L., 2006. The impacts of climate change in coastal marine systems. Ecol. Lett. 9, 228–241. http://dx.doi.org/10.1111/j.1461-0248.2005.00871.x.

Ho, K.K.Y., Leung, P.T.Y., Ip, J.C.H., Qiu, J.W., Leung, K.M.Y., 2014. *De novo* transcriptomic profile in the gonadal tissues of the intertidal whelk *Reishia clavigera*. Mar. Pollut. Bull. 85, 499–504. http://dx.doi.org/10.1016/j.marpolbul.2014.02.023.

Hofmann, G.E., Todgham, A.E., 2010. Living in the now: physiological mechanisms to tolerate a rapidly changing environment. Annu. Rev. Physiol. 72, 127–145. http://dx.doi.org/10.1146/annurev-physiol-021909-135900.

Huang, Z.-X., Chen, Z.-S., Ke, C.-H., Zhao, J., You, W.-W., Zhang, J., Dong, W.-T., Chen, J., 2012. Pyrosequencing of *Haliotis diversicolor* transcriptomes: insights into early developmental molluscan gene expression. PLoS One 7, e51279. http://dx.doi.org/10.1371/journal.pone.0051279.

Huang, X. De, Zhao, M., Liu, W.G., Guan, Y.Y., Shi, Y., Wang, Q., Wu, S.Z., He, M.X., 2013. Gigabase-scale transcriptome analysis on four species of pearl oysters. Mar. Biotechnol. 15, 253–264. http://dx.doi.org/10.1007/s10126-012-9484-x.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236–1240. http://dx.doi.org/10.1093/bioinformatics/btu031.

Maas, A.E., Lawson, Gareth, L., Tarrant, A.M., 2015. Transcriptome-wide analysis of the response of the thecosome pteropod *Clio pyramidata* to short-term CO_2 exposure. Comp. Biochem. Physiol. Part D Genomics Proteomics http://dx.doi.org/10.1016/j.cbd.2015.06.002 (doi:doi:10.1016/j.cbd.2015.06.002).

McCarthy, D.J., Chen, Y., Smyth, G.K., 2012. Differential expression analysis of multifactor RNA-seq experiments with respect to biological variation. Nucleic Acids Res. 40, 4288–4297. http://dx.doi.org/10.1093/nar/gks042.

Moreira, R., Pereiro, P., Canchaya, C., Posada, D., Figueras, A., Novoa, B., 2015. RNA-seq in *Mytilus galloprovincialis*: comparative transcriptomics and expression profiles

among different tissues. BMC Genomics 1–18 http://dx.doi.org/10.1186/s12864-015-1817-5.

Nicolas, J.-L., Basuyaux, O., Mazurié, J., Thébault, A., 2002. *Vibrio carchariae*, a pathogen of the abalone *Haliotis tuberculata*. Dis. Aquat. Org. 50, 35–43. http://dx.doi.org/10.3354/dao050035.

Parker, L.M., Ross, P.M., O'Connor, W.A., 2010. Comparing the effect of elevated pCO_2 and temperature on the fertilization and early development of two species of oysters. Mar. Biol. 157, 2435–2452. http://dx.doi.org/10.1007/s00227-010-1508-3.

Picone, B., Rhode, C., Roodt-Wilding, R., 2015. Transcriptome profiles of wild and cultured South African abalone, *Haliotis midae*. Mar. Genomics 20, 1–4. http://dx.doi.org/10.1016/j.margen.2015.01.002.

Prentis, P.J., Pavasovic, A., 2014. The *Anadara trapezia* transcriptome: a resource for molluscan physiological genomics. Mar. Genomics 50, 8–10. http://dx.doi.org/10.1016/j.margen.2014.08.004.

Riviere, G., Klopp, C., Ibouniyamine, N., Huvet, A., Boudry, P., Favrel, P., 2015. GigaTON: an extensive publicly searchable database providing a new reference transcriptome in the pacific oyster *Crassostrea gigas*. BMC Bioinf. 16, 401. http://dx.doi.org/10.1186/s12859-015-0833-4.

Senatore, A., Edirisinghe, N., Katz, P.S., 2015. Deep mRNA sequencing of the *Tritonia diomedea* brain transcriptome provides access to gene homologues for neuronal excitability, synaptic transmission and peptidergic signalling. PLoS One 10, e0118321. http://dx.doi.org/10.1371/journal.pone.0118321.

Shiel, B.P., Hall, N.E., Cooke, I.R., Robinson, N.A., Strugnell, J.M., 2015. *De Novo* Characterisation of the greenlip abalone transcriptome (*Haliotis laevigata*) with a focus on the heat shock protein 70 (HSP70) family. Mar. Biotechnol. 17, 23–32. http://dx.doi.org/10.1007/s10126-014-9591-y.

Sun, X., Yang, A., Wu, B., Zhou, L., Liu, Z., 2015. Characterization of the mantle transcriptome of Yesso scallop (*Patinopecten yessoensis*): identification of genes potentially involved in biomineralization and pigmentation. PLoS One 10, e0122967. http://dx.doi.org/10.1371/journal.pone.0122967.

Valenzuela-Muñoz, V., Bueno-Ibarra, M.a., Escárate, C.G., 2014. Characterization of the transcriptomes of *Haliotis rufescens* reproductive tissues. Aquac. Res. 45, 1026–1040. http://dx.doi.org/10.1111/are.12044.

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–63. http://dx.doi.org/10.1038/nrg2484.

Zhang, L., Li, L., Zhu, Y., Zhang, G., Guo, X., 2014. Transcriptome analysis reveals a rich gene set related to innate immunity in the Eastern oyster (*Crassostrea virginica*). Mar. Biotechnol. 16, 17–33. http://dx.doi.org/10.1007/s10126-013-9526-z.

Please cite this article as: Harney, E., et al., *De novo* assembly and annotation of the European abalone *Haliotis tuberculata* transcriptome, Mar. Genomics (2016), http://dx.doi.org/10.1016/j.margen.2016.03.002