Pyrosequencing and *de novo* assembly of Antarctic krill (*Euphausia superba)* transcriptome to study the adaptability of krill to climate-induced environmental changes

B. MEYER,*†[1] P. MARTINI,‡[1] A. BISCONTIN,‡ C. DE PITTÀ,‡ C. ROMUALDI,‡ M. TESCHKE,*
S. FRICKENHAUS,§¶ L. HARMS,§ U. FREIER,** S. JARMAN†† and S. KAWAGUCHI††

*Section Polar Biological Oceanography, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany, †Institute for Chemistry and Biology of the Marine Environment, Carl von Ossietzky University of Oldenburg, Carl-von-Ossietzky-Straße 9-11, 26111 Oldenburg, Germany, ‡Dipartimento di Biologia, Università degli Studi di Padova, via U. Bassi, 58/B, 35131 Padova, Italy, §Section Scientific Computing, Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, Germany, ¶Hochschule Bremerhaven, An der Karlstadt 8, 27568 Bremerhaven, Germany, **SC-Scientific Consulting, Münchener Str. 41a, D-41472 Neuss, Germany, ††Australian Antarctic Division, Kingston Tas., 7050, Australia

Abstract

The Antarctic krill, *Euphausia superba*, has a key position in the Southern Ocean food web by serving as direct link between primary producers and apex predators. The south-west Atlantic sector of the Southern Ocean, where the majority of the krill population is located, is experiencing one of the most profound environmental changes worldwide. Up to now, we have only cursory information about krill's genomic plasticity to cope with the ongoing environmental changes induced by anthropogenic CO_2 emission. The genome of krill is not yet available due to its large size (about 48 Gbp). Here, we present two cDNA normalized libraries from whole krill and krill heads sampled in different seasons that were combined with two data sets of krill transcriptome projects, already published, to produce the first knowledgebase krill 'master' transcriptome. The new library produced 25% more *E. superba* transcripts and now includes nearly all the enzymes involved in the primary oxidative metabolism (Glycolysis, Krebs cycle and oxidative phosphorylation) as well as all genes involved in glycogenesis, glycogen breakdown, gluconeogenesis, fatty acid synthesis and fatty acids β-oxidation. With these features, the 'master' transcriptome provides the most complete picture of metabolic pathways in Antarctic krill and will provide a major resource for future physiological and molecular studies. This will be particularly valuable for characterizing the molecular networks that respond to stressors caused by the anthropogenic CO_2 emissions and krill's capacity to cope with the ongoing environmental changes in the Atlantic sector of the Southern Ocean.

Keywords: 454 pyrosequencing, Antarctic Krill, *Euphausia superba*, transcriptome

Received 16 June 2014; revision received 13 March 2015; accepted 18 March 2015

Introduction

Despite 90 years of krill research, we have only limited knowledge of the adaptive capability of this keystone species in the Southern Ocean to a range of possible temperature and pCO_2 regimes because the main driver in krill research has been the fisheries' requirements for stock forecasting and conservation measures.

The Atlantic sector contains over half of the total krill stocks in the Southern Ocean (Atkinson *et al.* 2004, 2008)

Correspondence: Bettina Meyer, Fax: +49-471-4831-1149; E-mail: bettina.meyer@awi.de

[1]These authors contributed equally to this work.

and is, with the region of the west Antarctic Peninsula, one of the most rapidly warming regions on Earth (Meredith & King 2005; Ducklow *et al.* 2007). Long-term abundance data of krill for the Scotia Sea, starting in the 1920s, indicate a declining trend in krill biomass since the 1970s. Growth rates of adult krill from the Scotia Sea have been shown to decline at sea water temperatures as low as 3 to 4 °C (Atkinson *et al.* 2006), whereas laboratory experiments show that early larval stages seem to be most affected by increasing pCO_2 and temperature (Kawaguchi *et al.* 2011). The northern Weddell Sea is predicted to be one of the most affected regions by ocean acidification in the Southern Ocean (Kawaguchi *et al.* 2013).

Temperature is a very important environmental factor affecting all biological processes (Hochachka & Somero 2002). However, the thermal window defining optimal function of important physiological processes in krill's life cycle is unknown. The seasonal cycle of krill is closely synchronized with their highly seasonal environment in terms of sea ice extent and food availability (Meyer 2012). Adult individuals show a clear seasonal pattern in metabolic activity, body lipid content, maturation and growth, with high energy demands and growth rates from mid-spring to early autumn (Meyer *et al.* 2010). Field and laboratory studies have shown reduced oxygen uptake rates from late autumn to the following spring, with a minimum in mid-winter, are irrespective of food supply (Teschke *et al.* 2008; Meyer *et al.* 2010). Lipid storage in adult krill is highest in late autumn and will be utilized during winter (Ju *et al.* 2009). Krill's life cycle is characterized by a strong interplay between endogenous physiological functions and seasonal environmental factors (Teschke *et al.* 2011). Therefore, it is crucial to understand how these important physiological life cycle functions are affected by stressors such as sea water temperature rise, increasing ocean acidification and decreasing salinity due to glacier melt caused by anthropogenic CO_2 emission.

A powerful approach to examine organismal responses to environmental change is by combining physiology performance indicators with transcriptomic changes, as demonstrated by recent characterization of the optimal thermal window for Antarctic fish (Windisch *et al.* 2011, 2014). While there is considerable scientific knowledge about krill's biology, ecology and physiology, respectively (for review see Meyer 2012), its genome sequence is not yet available due to its large genome size of up to 48 Gbp (Jeffery 2012), which is an order of magnitude larger than the human genome. Therefore, a systematic sequencing of cDNA libraries is an efficient approach for identifying a large proportion of the transcribed regions of the krill genome (De Pittà *et al.* 2008; Seear *et al.* 2010). Comprehensive transcript characterization allows the identification of molecular networks that will respond to physiological conditions outside the optimal range (Windisch *et al.* 2011). *E. superba* has been the subject of only two large-scale transcriptome sequencing projects so far. The first transcriptome of krill based on 454 pyrosequencing technology was generated by Clark *et al.* (2011), focusing on chaperone genes. A further transcriptome sequencing project was performed by De Pittà *et al.* (2013), focusing on circadian clock genes and clock-controlled genes.

Recent investigations on thermal acclimation in Antarctic fish have shown that rising sea water temperature affect a network of metabolic pathways rather than single genes (Windisch *et al.* 2011). A shift in metabolic pathways such as an alteration from a lipid-based metabolic network to pathways associated with carbohydrate metabolism was observed as a response to thermal acclimation in the Antarctic eelpout (*Pachycara brachycephalum*) (Windisch *et al.* 2011, 2014). A similar response in *E. superba* would have profound consequences for krill's overwintering success, given its reliance on stored lipids. In this respect, to get a holistic view of thermal acclimation in Antarctic krill, the main focus in our transcriptome sequencing project was on genes involved in metabolic pathways. In addition, we focused on genes related to temperature and pCO_2 stress, which were not addressed by Clark *et al.* (2011).

The overall aim of our transcriptome sequencing project was threefold as follows: (i) to enhance the amount of new transcripts using whole krill and krill heads sampled in different seasons, (ii) to develop a krill 'master' transcriptome by combining the new 454 reads with the ones performed by Clark *et al.* (2011) and De Pittà *et al.* (2013) and with already published EST and (iii) to analyse the 'master' transcriptome by focusing on genes which are involved in important annual life cycle functions such as metabolic activity, biochemical pathways, maturation and growth. The newly developed 'master' transcriptome provides new opportunities for experimental work to identify and characterize the response of regulatory networks of genes in krill to environmental stressors induced by the anthropogenic CO_2 emission.

Material and methods

Krill sampling and RNA extraction

Krill were caught in the Indian sector (east Antarctica) and south-west Atlantic sector (Lazarev Sea) of the Southern Ocean by oblique hauls of a Rectangular Midwater Trawl (RMT 8) in the upper 100 m of the water column. Krill from east Antarctica were sampled during a voyage with the Australian research vessel *Aurora Australis* on 12th February 2009 (late austral summer) at position 64.01°S, 111.1212°E. The captured krill were placed in 200-L tanks of sea water located in a shipboard constant-temperature room at 0 °C and dim light on board (for detail see Teschke *et al.* 2007). After arriving in Hobart, Tasmania, krill were delivered directly to the Australian Antarctic Division (AAD) krill-aquarium and kept in a 1670-L holding tank, which was connected to a 5000-L chilled sea water recirculation system. The sea water was maintained at 0.5 °C and was recirculated every hour through an array of filtration devices. Fluorescent tubes provided lighting, and a controlled-timer system was used to set a natural photoperiod, corresponding to the Southern Ocean at 66°S and 30 m depth. Live krill from this holding tank were used for fresh

tissue RNA preparation. Enzymatic processes were stopped in RNA*later*® solution (Life Technologies), and RNA isolation was subsequently conducted with TRIzol® (Invitrogen) reagent according to a modified supplier's procedure at the AAD genetic laboratory. Sedimented pellets in the Eppendorf tubes were again carefully resuspended in TRIzol® reagent and repeatedly processed, and phase separation steps of the supplier's protocol were conducted multiple times.

In the Lazarev Sea, krill were caught on three expeditions in austral late spring, early summer (ANTXXIII-2, 19 November 2005 to 12 January 2006), austral autumn (ANTXXI-4, 27 March to 6 May 2004) and winter (ANTXXIII-6, 11 June to 27 August 2006) with the German research vessel *Polarstern* along parallel meridional transects from 60°S to the Antarctic continent at 70°S and between 4°E and 6°W (Meyer *et al.* 2010). The freshly caught krill were shock frozen in liquid nitrogen and stored at −80 °C for further RNA extraction at the Alfred Wegener Institute (AWI), Germany. Four single krill heads per season were dissected from the frozen Lazarev Sea krill and used for RNA extraction. Krill heads were immediately transferred from −80 °C to a mortar and preground in liquid nitrogen to a homogenous powder. The powder was then stored in 1 mL TRIzol® reagent (Life Technologies), and total RNA was extracted according to the supplier's instructions.

Quantity and purity of the RNA extracts were determined using the NanoDrop ND1000 (Peqlab Biotechnology, Erlangen, Germany), and integrity of the RNA was analysed by capillary electrophoresis using an Agilent Bioanalyzer (Agilent, Waldbronn, Germany). Before cDNA synthesis, the RNA samples from the four single late austral summer krill from east Antarctica and those from the single krill heads sampled at different seasons in the Lazarev Sea (austral autumn, winter, late spring and early summer) were combined into two RNA pools which were used to set up two separate cDNA libraries (whole krill and krill heads).

Construction of normalized cDNA libraries and 454 sequencing

Two separate cDNA libraries were sequenced by 454 pyrosequencing (Roche): a library exclusively based on whole late summer krill from east Antarctica and a library based on samples of krill heads dissected from krill caught in different seasons, in the Lazarev Sea. Both mixtures were used for library constructions by the Max Planck Institute for Molecular Genetics (Berlin, Germany). Total RNA of the two pools (whole krill and krill heads) was used for cDNA synthesis using the SMART protocol (Mint-Universal cDNA synthesis kit, Evrogen, Moscow, Russia). The cDNA was subsequently normal-

ized using duplex-specific nuclease and re-amplified thereafter following the instructions of the 'Trimmer Kit' (Evrogen, Moscow, Russia). Sequencing libraries were prepared from cDNA using the 'GS FLX Titanium General Library Preparation Kit' (Roche, Basel, Switzerland). Before sequencing, the libraries were amplified by polymerase chain reaction (PCR) using the 'GS FLX Titanium LV emPCR Kit' (Roche, Basel, Switzerland) (De Gregoris *et al.* 2011). Sequencing was performed by the Max Planck Institute for Molecular Genetics (Berlin, Germany) on a 454 Genome Sequencer FLX using the Titanium chemistry (Roche). Initial quality control and filtering of adapters and barcodes were performed at the Max Planck Institute for Molecular Genetics (Berlin, Germany). Raw data were archived at the European Nucleotide Archive (ENA) of the EBI under Accession PRJEB6147.

De novo sequence assembly and mapping of reads

In addition, to the 454 transcriptome libraries on whole krill and krill heads from different seasons described here (hereafter BM), two 454 libraries were recently published by Clark *et al.* 2011 (hereafter CK, SRA study: PRJNA79749, SRA sequences: SRP003407) and De Pittà *et al.* 2013 (hereafter DP, SRA study: PRJNA179348, SRA sequences: SRX205108). A total of 2.7 million raw reads (Fig. 1, step 1) were produced from BM, CK and DP. The adapter sequences and other artefacts of the pyrosequencing procedure were trimmed using SeqClean (https://sourceforge.net/projects/seqclean/) resulting in 2.6 million reads of good quality. All reads shorter than 70 bp were discarded. After the filtering process (Fig. 1, step 2), all the 454 sequences of BM, CK and DP were assembled using independently two different software packages (Fig. 1, step 3.1 and 3.2): MIRA 3.4 (Chevreux *et al.* 1999) and NEWBLER 2.6 (Roche) (www.454.com) (see Table S1 for more details). The results of these assemblies were clustered with CD-HIT 4.5 (Li & Godzik 2006) (Fig. 1, step 4). Two or more contigs were clustered when their similarity was higher than 85%. The longest contig was used to represent each cluster in the final assembly. To improve the quality of the annotation process, we filtered out all contigs smaller than 300 bp. The assembled sequences are available at EBI (Study PRJEB6147, Accession range HACF01000001-HACF01058581).

Functional annotation analysis

The annotation of the putative transcripts was performed according to De Pittà *et al.* (2013) (Fig. 1, step 5). Briefly, each of the selected consensus transcripts was searched locally against the NCBI nucleotide database and
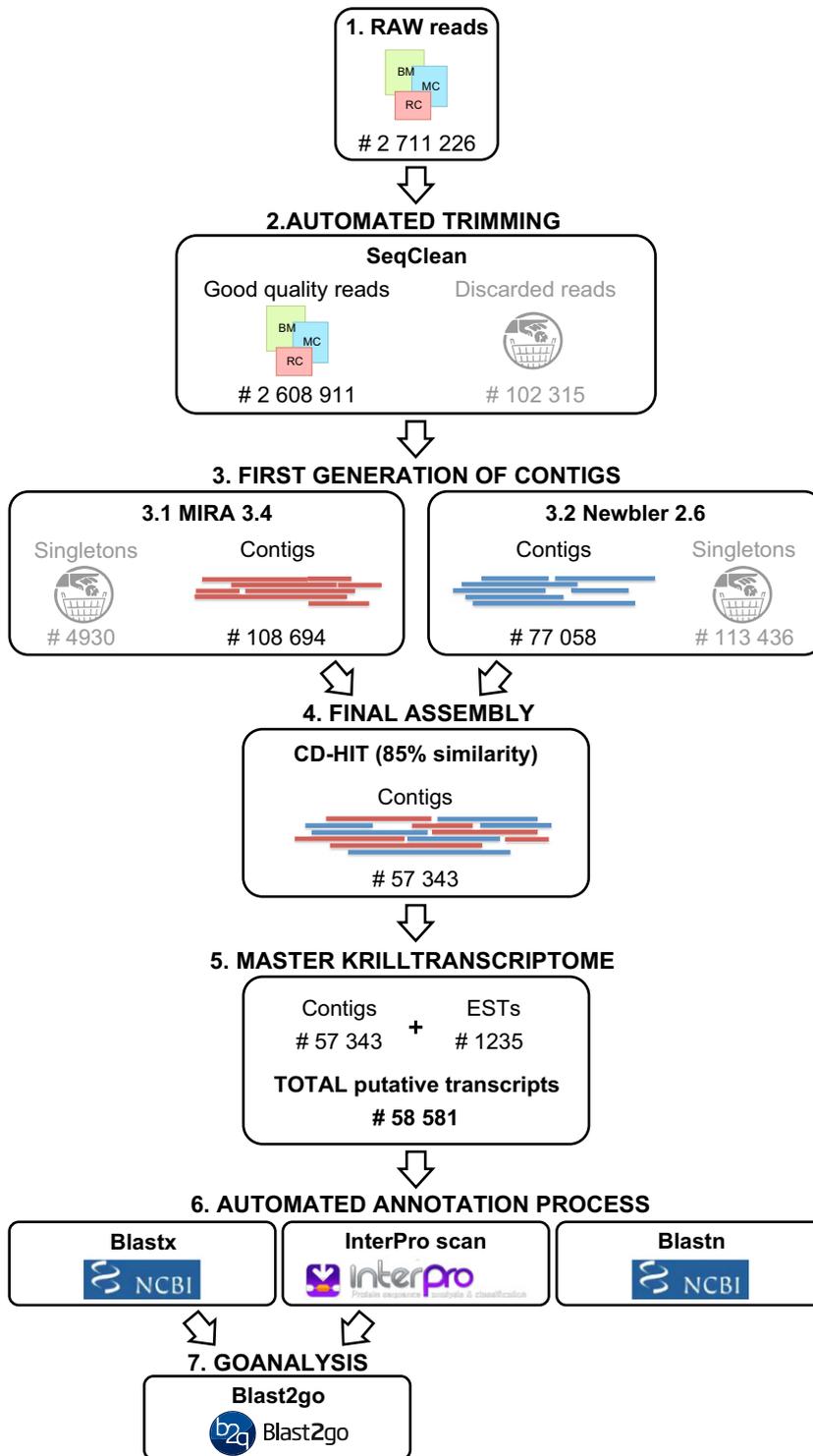
Fig. 1 Flow chart of the assembly and automated annotation of 454 reads. 1. Raw reads. Raw reads from three different 454 sequencing runs (BM, CK and DP) were grouped together. 2. Automated trimming. The adapter sequences and other artefacts were trimmed using Seq-Clean, and reads shorter than 70 bp were discarded. 3. First generation of contigs. 454 good-quality reads were assembled with MIRA 3.4 and NEWBLER 2.6 independently. 4. Final assembly. The results of two independent assemblies were clustered together with CD-HIT 4.5. 5. 'Master' krill transcriptome. A total of 58 581 putative krill transcripts were obtained adding the 1235 *E. superba* ESTs, available in the public databases. 6. Automated annotation process. Each consensus sequence was searched locally against the NCBI and UniProtKB databases. 6. GO Analysis. Functional annotation of the *E. superba* transcriptome was performed using the BLAST2GO software v.2.6.0. See Material and methods for more details.

UniProtKB database, using BLAST-X and BLAST-N, respectively. Results with expectation values $>e^{-6}$ for protein (Blast-X) and e^{-50} for nucleotide (BLAST-N) were discarded, as they were considered uninformative. Higher priority was given to the BLAST-N hits, while alignments characterized by <30% of coverage were discarded. Finally, among the five best hits, we selected the hit associated with the organism having the closest taxonomic relationship with *E. superba*. The taxonomic distribution of best hits in our transcriptome was then analysed with

Metagenome Analyzer (MEGAN – version 4.70.4) (Huson *et al.* 2011).

Functional annotation of the Antarctic krill transcriptome was performed with BLAST2GO software v.2.6.0. (Fig. 1, step 6) (Conesa *et al.* 2005; Götz *et al.* 2008). Homology searches were performed using BLAST-X against nonredundant protein database, and InterPro-Scan against protein domains in all available protein signature databases (Quevillon *et al.* 2005) using the parameters is described in Table S1.

We used TransDecoder (transdecoder.source-forge.net, Brian *et al.* 2013) using PFAM 27.0 (pfam-A database portions, Punta *et al.* 2012) and minimum open reading frame (ORF) length of 30 amino acids (AA) to find putative ORFs on those transcripts without annotation. Then, we search for transposable elements with RepeatMasker. The reference organism we used was *Drosophila melanogaster* because this is currently the genetic model organism most closely related to krill. We use Repbase 19.06 and RepeatMasker libraries 20140131 (Jurka *et al.* 2005).

Results and discussion

Krill ‘master’ transcriptome assembling

For developing a unique krill transcriptome database to set up the most comprehensive representation of the krill’s transcriptome, we assembled the 454 reads generated by the BM library together with all the 454 sequences available from public databases (Clark *et al.* 2011; De Pittà *et al.* 2013) for a total of 2.7 M raw reads. After the cleaning process (see Material & Methods for details), a total of 2.6 M (96.3%) high-quality reads were further processed. The assembly approach we adopted followed the strategy described in Kumar & Blaxter (2010). We combined two different assemblers

with different features to get a more robust final assembly. Specifically, we selected NEWBLER 2.6 which gave longer contigs while keeping the contig number smaller and MIRA 3.4 which is more suited for reads obtained from normalized libraries and maintains contigs shorter with few singletons (see Fig. 1, step 1, 2 and 3). As expected, 108 694 contigs and 4930 singletons were produced by MIRA 3.4 while NEWBLER 2.6 assembling provided 77 058 contigs and 113 436 singletons. The singletons were discarded and excluded from further analyses. Contigs obtained from both assemblies were clustered with 85% of similarity using CD-HIT 4.5 (Fig. 1, step 4). The longest contig was used to represent each cluster in the final assembly. This clustering process contributes to the robustness of the final assembly as it refines the final transcriptome, removing similar sequences. At the end of the assembling process, a total of 57 343 contigs longer than 300 bp were obtained: 26 378 contigs identified by MIRA 3.4 and 30 965 by NEWBLER 2.6.

Assembled contigs ranged in size from 300 bp to 11 127 bp, with an average size of 691 bp (median 521 bp). A total of 8289 (14.45%) contigs were larger or equal than 1 kb (Fig. 2A). In comparison with other libraries constructed from 454 sequences, the average length of our assembled contigs was longer than that previously reported for krill by Clark *et al.* (2011) (average of 492 bp) but similar to the length of contigs found in other decapod crustaceans (Jung *et al.* 2011; Mundry *et al.* 2012; Harms *et al.* 2013). However, our contigs showed an average length slightly shorter than that obtained by De Pittà *et al.* (2013) (average 890 bp), where the authors used expressed sequence tags (ESTs) previously produced by the Sanger method (De Pittà *et al.* 2008; Seear *et al.* 2010). Comparing our contigs with the ESTs deposited at the NCBI, we found that 5366 of 6884 ESTs (77.9%) have a similarity greater than 90%
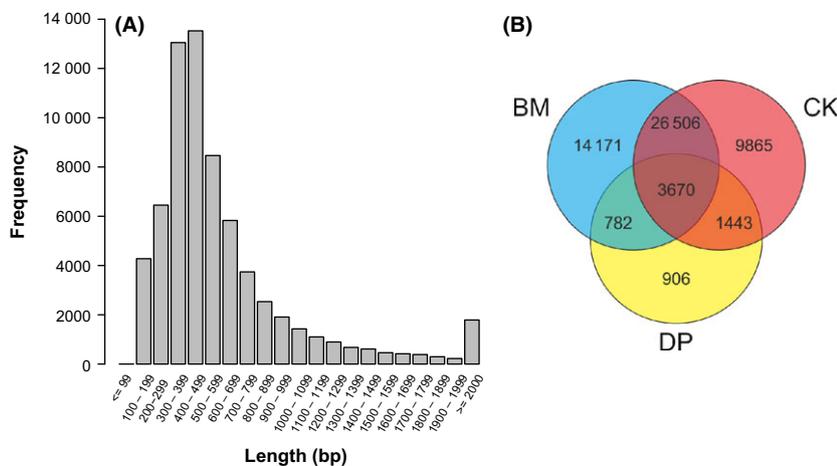
Fig. 2 (A). Size distribution of contigs from 454 pyrosequencing. Length distribution of contigs generated by the final assembling of 454 reads generated by BM, CK and DP. (B) Gene discovery rate of each cDNA library. The Venn diagram shows the contribution of each 454 sequencing projects to define the final assembly. Blue, red and yellow circles represent BM, CK and DP cDNA libraries, respectively. 454 pyrosequencing of BM, CK and DP cDNA libraries provided about 25%, 17% and 2% of new krill transcripts, respectively.

indicating that our assembly contains the majority of the information obtained by Sanger sequencing. Even so, the 1235 unique ESTs derived from Sanger sequencing that were not represented in our assembly were manually included in the 'master' transcriptome to produce the most comprehensive resource, representing the true Antarctic krill transcriptome.

The advantage of our approach is threefold: First, the combination of different sequencing efforts increases the overall coverage. Secondly, the combination of biologically distinct samples allows deep exploration of the complexity of transcriptomes (Yu *et al.* 2014). Thirdly, the combination of two assemblers yields longer contigs and favours the identification of new transcripts.

Comparison of BM library with available public krill databases

The krill 454 final assembly is the result of three different cDNA libraries created by *de novo* assembly (CK, DP and BM). Figure 2B shows the contribution of each library in the identification of new putative krill transcripts. The BM, CK and DP libraries contribute to an increase of new transcripts by 25% (14 171 of 57 343, corresponding to 31% of its own transcripts), 17% (9865 of 57 343, corresponding to 24% of its own transcripts) and 2% (906 of 57 343, corresponding to 13% of its own transcripts), respectively. The high gene discovery rate in the BM library but also in the DP library, despite the relatively small number of reads produced in the DP library, confirms the validity of the normalization procedures adopted to enhance gene discovery rate, which was not performed by the CK library.

Annotation process and construction of the krill 'master' transcript catalogue

To make an assessment for the identities of putative transcripts, each nonredundant sequence was searched, as described in De Pittà *et al.* (2013), in the NCBI nucleotide and UniProtKB databases (Fig. 1, step 5). Overall, 26% of the sequences (15 347 of 58 581) were successfully annotated (Table S2) while the remaining 74% (43 234 of 58 581) showed no or poor similarity matches representing presumably completely unknown Antarctic krill transcripts. In detail, 7942 (52%) and 7429 (48%) of putative transcripts were successfully annotated with BLAST-N (nonredundant nucleotide database) and BLAST-X (nonredundant protein database), respectively. The percentage of annotated transcripts might appear rather low compared to that reported by De Pittà *et al.* (2013), and this is probably due to a high proportion of novel genes and the lack of fully annotated transcriptomes in closely related species.

Regarding the nonannotated transcripts (74% corresponding to 43 234 transcripts), we used TransDecoder to identify potential ORFs. With a minimum length of predicted protein set to 30AA, TransDecoder predicted at least one ORF for 17 080 (29%) transcripts. These sequences may represent either specific *E. superba* transcripts or fragments that are too short to get a significant similarity on available databases. We compared the average lengths of the transcripts with a coding sequence with that of the annotated group and with that of the not annotated group. We found that the annotated group has an average length of ~900 nt; this length decreases to ~706 nt for those transcripts with predicted ORFs and even more (to ~536 nt) in the group of nonannotated transcripts. This suggests that nonannotated sequences may represent fragments of transcripts, noncoding RNA or transcripts with poor or no homology to annotated species. There is growing evidence that noncoding transcripts can provide an extra layer of regulation of gene expression and the proportion of noncoding transcripts is thought to broadly increase with developmental complexity because protein-based regulation seems to reach its limit with prokaryotes (Mattick 2004).

Gene duplication is postulated to have played a major role in the evolution of biological novelty (Roth *et al.* 2007). Based on this hypothesis and according to our results, we can speculate that the large krill genome size could be the result of an evolutionary adaptation to different environmental changes in terms of increasing plasticity under the control of noncoding transcripts rather than protein coding. The krill genome is not polyploid and has 17 chromosomes (2N karyotype) (Van Ngan 1989), suggesting that the abnormal genome size of *E superba* could be due to the activity of transposable elements rather than genome duplication (Jarman *et al.* 1999, 2000; Jeffery 2012). For testing this hypothesis, we ran RepeatMasker on our transcriptome and found that there are no particular evidences either of retroelements (0.26% of the assembled bases) or of DNA transposons (0.02% of the assembled bases) within the assembled sequences. Despite the bias towards the number of transposable elements discovered in arthropods that is considerably lower than in mammals, this analysis may suggest that transposable elements have minimal to negligible activity in krill. All these evidences seem to support the role and the presence of noncoding transcripts as major actors of the krill plasticity.

The taxonomic distribution of all *E. superba* putative transcripts was reported in Fig. 3. A large proportion of the reads have no clear similarity to reads characterized in other organisms (43 234 of 58 581). The largest annotated fraction of putative transcripts was similar to reads from *Daphnia pulex* (2453 of 15 347), which is one of the few crustaceans for which the characterized genome shows
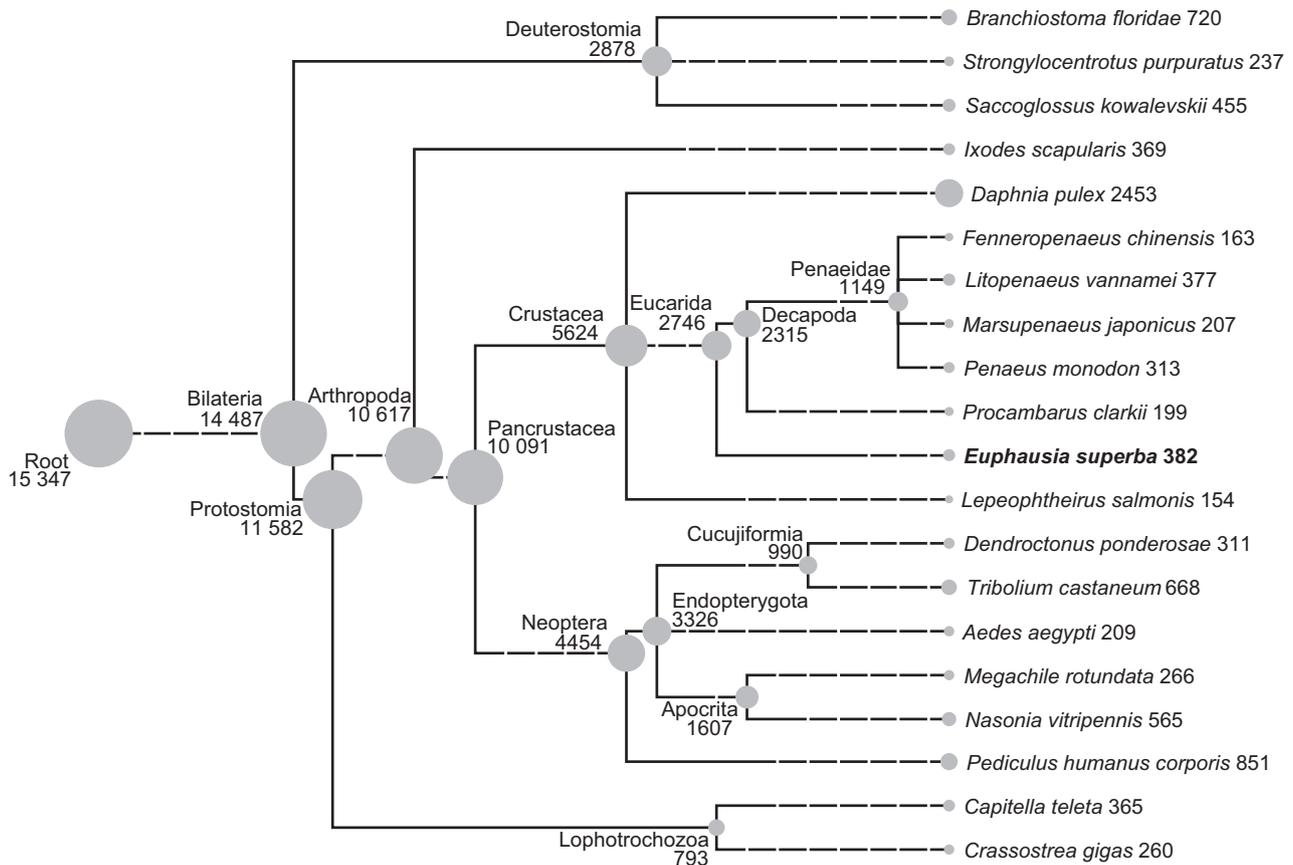
Fig. 3 Organisms most represented in the protein similarity searches with krill sequences. The taxonomic distribution of all *E. superba* putative transcripts (15 347) was plotted using the Metagenome Analyzer (MEGAN – version 4.70.4) based on the best hit for each putative transcript. Grey circles with different diameters represent the number of putative transcripts annotated with a given species. The diameter of each circle is proportional to the contribution of a given species in the transcriptome annotation of *E. superba*. See Table S2 for more details.

higher levels of transcriptome annotation (Colbourne *et al.* 2011). The great majority (about 66%) of the putative transcripts showed high similarity with nucleotide or amino acid sequences from Pancrustacea (10 091 of 15 347) falling nearly equally between crustaceans (about 37%) and insects (about 29%), but only 3% were similar to the known sequences of Euphausiacea (432 of 15 347), and in particular 382 putative transcripts out of 432 were similar to *E. superba*. In addition, our transcriptome was checked for contamination by microorganisms associated with the sampled krill. We found that 3.7% of the transcripts (571 of 15 347) have at least one BLAST-X hit (among the five best hits) of organism groups, which could be considered as contaminants: bacteria (1.53%), protists (1.08%), fungi (0.48%), algal (0.57%) and viruses (0.04%).

Functional analysis of the krill 'master' transcriptome

Gene Ontology (GO) has been widely used to perform gene classification and functional annotation (Bard &

Rhee 2004) using controlled vocabulary and hierarchy including molecular function, biological process and cellular components. GO analysis of the Antarctic krill transcriptome identifies 42 398 GO terms for 13 175 putative transcripts (about 22%; 13 175 of 58 861). Using generic GO slim, which groups GO terms giving a broad could overview of the ontology content without the details of the specific fine-grained terms, we obtained 19 277 terms on 'Molecular function' (45.4%), 7744 on 'Cellular components' (18.3%) and 15 377 on 'Biological process' (36.3%) shown in Fig. S1.

The distribution of GO terms assigned to *E. superba* transcripts was compared with that obtained from the *D. pulex* genome (Colbourne *et al.* 2011). We found high similarity between the GO category distribution in the two annotations (Fig. 4), suggesting that the major biological and functional categories are represented in our krill 'master' transcriptome.

We focused our attention on the 'Biological process' category for transcriptomic profiling of metabolic
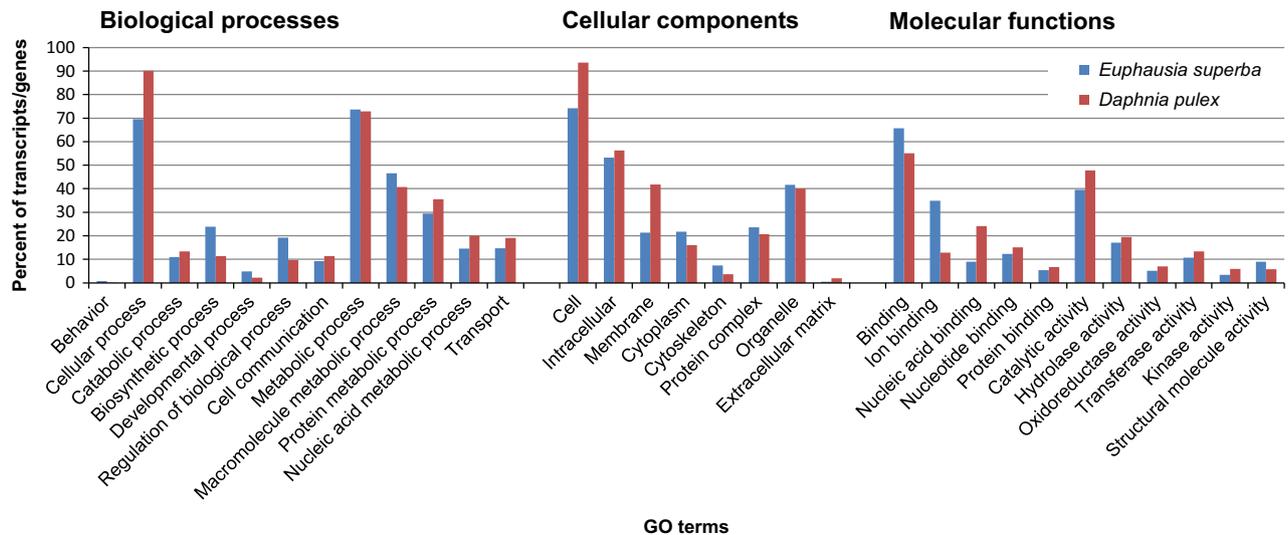
Fig. 4 Comparative distribution of gene ontology terms of *E. superba* 'master' transcriptome with respect to *D. pulex* genome. The most represented GO terms were divided in three main categories: biological processes, cellular components and molecular functions. The two distributions show a clear overlap and confirm the representation of main GO terms in our 'master' transcriptome.

pathways, deciding to manually examine the 15 377 GO terms of this category to further cluster the annotated putative transcripts in a total of twelve GO categories (Table S3). As shown in Fig. 5A, the most frequent categories were 'metabolic processes' (33.7%), 'protein metabolism' (23.1%) and 'transport' (13.0%), followed by 'nucleic acid metabolism' (8.7%), 'signal transduction' (6.5%), 'cellular processes' (4.5%), 'cellular component organization' (3.8%) and 'response to stress' (3.4%). Other 'Biological Processes' categories such as 'developmental processes', 'reproduction' and 'behaviour' are also present, albeit at lower percentages. All the transcripts were assigned to a GO term, but transcripts unambiguously annotated as contaminants from other kingdoms, and were manually added to the 'symbiosis, encompassing mutualism through parasitism' category (0.4%).

Furthermore, we compared the new transcriptome with the most complete krill transcriptome previously published (krill1.0; De Pittà *et al.* 2013 – resulting from the assembly of the two 454 libraries CK and DP together with the *E. superba* ESTs deposited at the NCBI). We were able to double the putative transcripts (from 32 217 to 58 581) and the number of sequences successfully assigned to biological process categories (from 3121 to 7471). In particular, we significantly increased the number of transcripts involved in metabolic processes (from 1120 to 2516 with 363 BM specific transcripts), protein metabolism (from 886 to 1725 with 247 BM specific transcripts), transport (from 368 to 974 with 156 BM specific transcripts), nucleic acid metabolism (from 203 to 652 with 136 BM specific transcripts), signal transduction

(from 110 to 484 with 117 BM specific transcripts), cellular processes (from 107 to 335 with 66 BM specific transcripts), stress response (from 74 to 253 with 50 BM specific transcripts), cellular component organization (from 136 to 286 with 38 BM specific transcripts) and developmental processes (from 57 to 177 with 34 BM specific transcripts). These results confirm the validity of the normalization procedure for increasing the gene discovery rate and the effectiveness of the assembly strategy adopted.

Transcripts involved in 'metabolic processes' and in the 'response to stress'

Recent investigations of the thermal acclimation in the Antarctic eelpout, *P. brachycephalum*, have shown a hepatic metabolic reorganization, indicating an alteration from a lipid-based metabolic network to pathways associated with carbohydrate metabolism (Windisch *et al.* 2011, 2014). This picture of cellular adjustments to the warmth by the Antarctic eelpout has illustrated that we have to take a holistic view by identifying molecular networks rather than single genes to understand marine ectotherms capacities to cope with environmental change caused by the anthropogenic CO_2 emission (e.g. elevated sea water temperature, ocean acidification or reduced salinity due to glacier melt).

In adult Antarctic krill, a shift in metabolic pathways as shown for the Antarctic eelpout would have profound implications for krill's overwintering and spawning activity in the forthcoming spring. Krill build up considerable amount of body lipid reserves during the austral

summer for their utilization during winter (Hagen *et al.* 2001). Increasing energy demands due to a warming environment (Pörtner & Farrell 2008) may impede the build-up of sufficient reserves during summer to allow survival of the winter season (Hagen *et al.* 2001) and to fulfil the external maturation process (Teschke *et al.* 2008). For this reason, our analysis focused on the presence of genes in the 'master' transcriptome involved in 'metabolic processes'.

The majority of transcripts grouped in 'metabolic processes' in our 'master' transcriptome are involved in carbohydrate metabolism (23.8%) as shown in Fig. 5B. We have identified nearly all the enzymes involved in primary oxidative metabolism (Glycolysis, Krebs cycle and Oxidative phosphorylation) (Fig. S2 and Table S4). All genes involved in glycogenesis, glycogen breakdown, gluconeogenesis, fatty acids synthesis and fatty acid β-oxidation were successfully identified. In this respect, our krill 'master' transcriptome provides the most updated picture of metabolic pathways in Antarctic krill (see also metabolic KEGG-pathway map in Fig. S3). Among the genes involved in the energy storage, we have identified, for the first time, (i) UDP-glucose pyrophosphorylase and the glycogen branching enzyme which promotes glucose conversion to glycogen (Fig. S2B and Table S4), (ii) acetyl-CoA carboxylase which catalyses the first step of fatty acids synthesis (Fig. S4A and Table S4) and (iii) Acyl-CoA synthetase that produce a Palmitoyl-CoA (Fig. S4A and Table S4). Moreover, we completed the molecular characterization of fatty acid β-oxidation and identified (i) the carnitine palmitoyltransferase I, which is part of a shuttle system to transport the long chain fatty acids to the mitochondrial matrix (Table S4) and (ii) the enoyl-CoA hydratase that is essential to

catalyse the second step in the breakdown of fatty acids (Fig. S4A). Finally, we have identified the pyruvate carboxylase and glucose-6-phosphatase that catalyse the first and the last step of gluconeogenesis, respectively (Fig. S2A and Table S4).

About four per cent of the annotated transcripts were assigned to 'Response to stress' (Table S4). A set of these genes was generated in marine organisms after exposure to warming sea water or increasing sea water pCO_2, and only partly described in Clark *et al.* (2011).

Several of these studies observed the expression of transcripts involved in response to oxidative stress in marine copepods (Lauritano *et al.* 2012), the coral *Acropora millepora* (Bellantuono *et al.* 2012) and the white shrimp *Litopenaeus vannamei* (Zhou *et al.* 2010). In the latter species, the enzymes superoxide dismutase, catalase, glutathione peroxidase and glutathione transferase were identified as biomarkers for temperature stress (Zhou *et al.* 2010). These transcripts and 18 further transcripts involved in response to oxidative stress are now included in our krill 'master' transcriptome (Table S4). One of the 18 transcripts is thioredoxin peroxidase, highly expressed in *L. vannamei* when exposed to pH stress (Wang *et al.* 2006), whereas another is ferritin, up-regulated in the stone coral *A. millepora* after thermal stress (Bellantuono *et al.* 2012). In addition, we found, for the first time, genes coding for the AP-1 transcription factor and the inhibitor of NF-kB, hypothesized to be involved in the thermal tolerance of *A. millepora* by regulating the thermal stress signalling and inhibiting the apoptotic cascade, respectively (Bellantuono *et al.* 2012). The DNA-binding activities of AP-1 and NF-kB transcription factors have been demonstrated to be induced by changes in the intracellular redox state due to expo-

(A) Biological processes

■ Protein metabolism (23.1%)
■ Nucleic acid metabolism (8.7%)
■ Metabolic process (33.7%)
■ Cellular Process (4.5%)
■ Developmental process (2.4%)
■ Cellular component organization (3.8%)
■ Transport (13.0%)
■ Signal transduction (6.5%)
■ Response to stress (3.4%)
■ Symbiosis, encompassing mutualism through parasitism (0.3%)
■ Reproduction (0.4%)
■ Behavior (0.2%)

(B) Metabolic processes

■ Carbohydrate metabolic process (23.8%)
■ Cellular amino acid metabolic process (14.1%)
■ Oxidation-reduction process (14.5%)
■ Lipid metabolic process (7.9%)
■ Nucleobase-containing compound metabolic process (7.5)
■ Generation of precursor metabolites and energy (7.3)
■ Nitrogen compound metabolic process (1.3%)
■ Other (23.6%)

Fig. 5 Classification of the annotated putative transcripts of *E. superba* into 12 functional categories. (A) Classification of the 7491 annotated putative transcripts into 12 different 'Biological process' GO categories (B) Subclassification of the 'Metabolic process' GO category (33.7%, 2516 contigs). Diagrams show the proportion of each GO term. See Table S3 for more details.

sure to environmental stress (Mattson *et al.* 2004). Finally, we identified several heat shock proteins (HSP70, HSC70, HSP90, HSP60, HSP40 and HSP10) (Table S4), which were already addressed in detail by Clark *et al.* (2011). However, the typical 'stress' genes of the heat shock protein (HSP) family such as HSP70 seem to be not ideal candidates for 'stress response biomarkers' due to their pluripotent nature of chaperone function (Gross 2004). They seem to be important intermodular elements of cellular networks, acting as multifunctional hubs (Korcsmáros *et al.* 2007).

Another stressor for marine invertebrates is increasing sea water pCO_2. Ocean acidification leads to reduction in the carbonate ion concentration, the essential component for shell and skeleton construction of marine organisms. However, the response of marine organism to different elevated sea water pCO_2 levels on their calcification rate is very variable between organism groups and species (Ries *et al.* 2009). An increase in the calcification rate was observed in the shrimp *Penaeus plebejus* exposed to low calcium carbonate saturation in sea water (Ries *et al.* 2009). Analysing the expression of key transcripts in biomineralization processes such as the carbonic anhydrase as well as structural component of cuticle could give us the opportunity to study the ocean acidification effect on synthesis rate of cuticle of krill from a molecular point of view. Recently, Seear *et al.* (2010) identified differentially expressed genes across the moult cycle of *E. superba* and defined gene expression signatures specific to known phenotypic structural changes. The authors focused their attention on cuticle genes, chitin metabolic enzymes, protease and several main players of immune response. All these transcripts are included in our krill 'master' transcriptome. We also identified all main enzymes involved in chitin synthesis, such as glucosamine-phosphate N-acetyltransferase and UDP-N-acetylglucosamine pyrophosphorylase (Fig. S4B). Several proteins responsible for chitin degradation such as beta-N-acetylhexosaminidase, glucosamine-6-phosphate deaminase and two chitinases were identified for the first time in the krill 'master' transcriptome. Moreover, we doubled the number of putative transcripts coding for cuticular proteins. Finally, we identified some transcripts involved in the hormonal control of moult such as ecdyson-induced protein 74EF isoform B and ecdyson receptor isoform 2a that are members of the ecdyson cascade and trigger ecdysis. Moreover, juvenile hormone esterase-binding protein, farnesoic acid O-methyltransferase and juvenile hormone epoxide hydrolase 1 are involved in the juvenile hormone metabolism, which plays a crucial role in the control of moult phases and the attainment of sexual maturity (Table S4).

Our krill 'master' transcriptome provides the most advanced transcripts catalogue of the nonmodel organism, *Euphausia superba*, and provides the most updated picture of metabolic pathways in krill. In combination with robust physiological and ecophysiological studies, the krill 'master' transcriptome is a stepping stone on the way to a holistic view of a better understanding how krill will be affected by environmental stressors induced by anthropogenic CO_2 emission.

Acknowledgements

References

Atkinson A, Siegel V, Pakhomov EA *et al.* (2004) Long-term decline in krill stock and increase in salps within the Southern Ocean. *Nature*, **432**, 100–103.

Atkinson A, Shreeve RS, Hirst AG *et al.* (2006) Natural growth rates in Antarctic krill (*Euphausia superba*): II: Predictive models based on food, temperature, body length, sex, and maturity stage. *Limnology & Oceanography*, **51**, 973–987.

Atkinson A, Siegel V, Pakhomov EA *et al.* (2008) Oceanic circumpolar habitats of Antarctic krill. *Marine Ecology Progress Series*, **362**, 1–23.

Bard JBL, Rhee SY (2004) Ontologies in biology: design, applications and future challenges. *Nature Reviews Genetics*, **5**, 213–222.

Bellantuono AJ, Granados-Cifuentes C, Miller DJ, Hoegh-Guldberg O, Rodriguez-Lanetty M (2012) Coral thermal tolerance: tuning gene expression to resist thermal stress. *PLoS One*, **7**, e50685.

Brian J, Haas AP, Yassour M *et al.* (2013) De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nature Protocols*, **8**, 1494–1512.

Chevreux B, Wetter T, Suhai S (1999) Genome sequence assembly using trace signals and additional sequence information. *Journal of Computer Science & Systems Biology*, **99**, 45–56.

Clark MS, Thorne MA, Toullec JY *et al.* (2011) Antarctic krill 454 pyrosequencing reveals chaperone and stress transcriptome. *PLoS One*, **6**, e15919.

Colbourne JK, Pfrender ME, Gilbert D *et al.* (2011) The ecoresponsive genome of *Daphnia pulex*. *Science*, **331**, 555–561.

Conesa A, Gotz S, García-Gómez JM *et al.* (2005) BLAST2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**, 3674–3676.

De Gregoris TB, Rupp O, Klages S *et al.* (2011) Deep sequencing of naupliar-, cyprid- and adult-specific normalised Expressed Sequence Tag (EST) libraries of the acorn barnacle *Balanus amphitrite*. *Biofouling: The Journal of Bioadhesion and Biofilm Research*, **27**, 367–374.

De Pittà C, Bertolucci C, Mazzotta GM *et al.* (2008) Systematic sequencing of mRNA from the Antarctic krill (*Euphausia superba*) and first tissue specific transcriptional signature. *BioMed Central Genomics*, **9**, 9–45.

De Pittà C, Biscontin A, Albiero A *et al.* (2013) The Antarctic krill *Euphausia superba* shows diurnal cycles of transcription under natural conditions. *PLoS One*, **8**, e68652.

Ducklow HW, Baker K, Martinson DG *et al.* (2007) Marine pelagic ecosystems: the West Antarctic Peninsula. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **362**, 67–94.

Götz S, García-Gómez JM, Terol J, *et al.* (2008) High-throughput functional annotation and data mining with the Blast2Go. *Nucleic Acids Research*, **36**, 3420–3435.

Gross M (2004) Emergency services: a bird's eye perspective on the many different functions of stress proteins. *Current Protein & Peptide Science*, **5**, 213–223.

Hagen W, Kattner G, Terbrüggen A, Van Vleet ES (2001) Lipid metabolism of the Antarctic krill *Euphausia superba* and its ecological implications. *Marine Biology*, **139**, 95–104.

Harms L, Frickenhaus S, Schiffer M *et al.* (2013) Characterization and analysis of a transcriptome from the boreal spider crab *Hyas araneus*. *Comparative Biochemistry and Physiology Part D Genomics Proteomics*, **8**, 344–351.

Hochachka PW, Somero GN (2002) *Biochemical Adaptation: Mechanisms and Process in Physiological Evolution*. Oxford University Press, New York.

Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Shuster SC (2011) Integrative analysis of environmental sequences using megan. *Genome Research*, **21**, 1552–1560.

Jarman S, Elliott N, Nicol S *et al.* (1999) The base composition of the krill genome and its potential susceptibility to damage by UV-B. *Antarctic Science*, **11**, 23–26.

Jarman SN, Elliott NG, Nicol S, McMinn A (2000) Molecular phylogenetics of circumglobal *Euphausia* species (Euphausiacea: Crustacea). *Canadian Journal of Fisheries and Aquatic Sciences*, **57**, 51–58.

Jeffery NW (2012) The first genome size estimates for six species of krill (Malacostraca, Euphausiidae): large genomes at the north and south poles. *Polar Biology*, **35**, 959–962.

Ju S-J, Kang H-K, Kim WS, Harvey HR (2009) Comparative lipid dynamics of euphausiids from the Antarctic and Northeast Pacifc Oceans. *Marine Biology*, **156**, 1459–1473.

Jung H, Lyons RE, Dinh H, Hurwood DA, McWilliam S, Mather PB (2011) Transcriptomics of a giant freshwater prawn (*Macrobrachium rosenbergii*): De novo assembly, annotation and marker discovery. *PLoS One*, **6**, e27938.

Jurka J, Kapitonov VV, Pavlicek A *et al.* (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, **110**, 462–467.

Kawaguchi S, Haruko K, King R *et al.* (2011) Will krill fare well under Southern Ocean acidification? *Biology Letters*, **7**, 288–291.

Kawaguchi S, Ishida A, King R *et al.* (2013) Risk maps for Antarctic krill under projected Southern Ocean acidification. *Nature Climate Change*, **3**, 843–847.

Korcsmáros T, Kováks IA, Szalay MS, Csermely P (2007) Molecular chaperones: the modular evolution of cellular networks. *Journal of Biosciences*, **32**, 441–446.

Kumar S, Blaxter ML (2010) Comparing de novo assemblers for 454 transcriptome data. *BioMed Central Genomics*, **11**, 571–583.

Lauritano C, Procaccini G, Ianora A (2012) Gene expression patterns and stress response in marine copepods. *Marine Environmental Research*, **76**, 22–31.

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.

Mattick J (2004) RNA regulation: a new genetics? *Nature Reviews Genetics*, **5**, 316–323.

Mattson D, Bradbury CM, Bisht KS, Curry HA, Spitz DR *et al.* (2004) Heat shock and the activation of AP-1 and inhibition of NF-kappa B DNA-binding activity: possible role of intracellular redox status. *International Journal of Hyperthermia*, **20**, 224–233.

Meredith MP, King JC (2005) Rapid climate change in the ocean west of the Antarctic Peninsula during the second half of the 20th century. *Geophysical Research Letters*, **32**, L19604.

Meyer B (2012) The overwintering of Antarctic krill, *Euphausia superba*, from an ecophysiological perspective. *Polar Biology*, **35**, 15–37.

Meyer B, Auerswald L, Siegel V *et al.* (2010) Seasonal variation in body composition, metabolic activity, feeding, and growth of adult krill *Euphausia superba* in the Lazarev Sea. *Marine Ecology Progress Series*, **398**, 1–18.

Mundry M, Bornberg-Bauer E, Sammeth M, Feulner PG (2012) Evaluating characteristics of de novo assembly software on 454 transcriptome data: a simulation approach. *PLoS One*, **7**, e31410.

Pörtner H-O, Farrell A (2008) Physiology and climate change. *Science*, **322**, 690–692.

Punta M, Coggill PC, Eberhardt RY *et al.* (2012) The Pfam protein families database. *Nucleic Acids Research*, **40**, D290–D301.

Quevillon E, Silventoinen V, Pillai S *et al.* (2005) InterProScan: protein domains identifier. *Nucleic Acids Research*, **33**, W116–W120.

Ries JB, Cohen AL, McCorkle DC (2009) Marine calcifiers exhibit mixed responses to CO_2-induced ocean acidification. *Geology*, **37**, 1131–1134.

Roth C, Rastogi S, Arvestad L *et al.* (2007) Evolution after gene duplication: models, mechanisms, sequences, systems, and organisms. *Journal of Experimental Zoology*, **308B**, 58–73.

Seear PJ, Tarling GA, Burns G *et al.* (2010) Differential gene expression during the moult cycle of antarctic krill (*Euphausia superba*). *BioMed Central Genomics*, **11**, 582–595.

Teschke M, Kawagushi S, Meyer B (2007) Simulated light regimes affect feeding and metabolism of Antarctic krill, *Euphausia superba*. *Limnology Oceanography*, **53**, 1046–1054.

Teschke M, Kawaguchi S, Meyer B (2008) Effects of simulated light regimes on maturity and body composition of Antarctic krill, *Euphausia superba*. *Marine Biology*, **154**, 315–324.

Teschke M, Wendt S, Kawaguchi S, Kramer A, Meyer B (2011) A circadian clock in Antarctic krill: an endogenous timing system governs metabolic output rhythms in the Euphausid species *Euphausia superba*. *PLoS One*, **6**, e26090 doi: 10.1371/journal.pone.0026090.

Van Ngan P (1989) Preliminary study on chromosomes of Antarctic krill, *Euphausia superba* Dana. *Polar Biology*, **10**, 149–150.

Wang W, Wang A, Liu Y, Xiu J, Liu Z *et al.* (2006) Effects of temperature on growth, adenosine phosphates, ATPase and cellular defense response of juvenile shrimp *Macrobrachium nipponense*. *Aquaculture*, **256**, 624–630.

Windisch HS, Kathöver R, Pörtner H-O *et al.* (2011) Thermal acclimation in Antarctic fish: transcriptomic profiling of metabolic pathways. *American Journal of Physiology – Regulatory, Integrative and Comparative Physiology*, **301**, R1453–R1466.

Windisch HS, Frickenhaus S, John U *et al.* (2014) Stress response or beneficial temperature acclimation: transcriptome signatures in Antarctic fish (*Pachycara brachycephalum*). *Molecular Ecology*, **23**, 3469–3482.

Yamada T, Letunic I, Okuda S *et al.* (2011) iPath2.0: interactive pathway explorer. *Nucleic Acids Research*, **39**, W412–W415.

Yu Y, Fuscoe JC, Zhao C *et al.* (2014) A rat RNA-Seq transcriptomic BodyMap across 11 organs and 4 developmental stages. *Nature Communications*, **5**, 3230.

Zhou J, Wang L, Xin Y, Wang W, He W *et al.* (2010) Effect of temperature on antioxidant enzyme gene expression and stress protein response in white shrimp, *Litopenaeus vannamei*. *Journal of Thermal Biology*, **35**, 284–289.

Data accessibility

BM raw data are archived at the European Nucleotide Archive (ENA) of the EBI under Accession PRJEB6147. DP and CK raw data are deposited at the Short Read Archive of the NCBI under the following Accession nos: (a) DP: study PRJNA179348, sequences SRX205108 and (b) CK: study PRJNA79749, sequences SRP003407). The assembled sequences are available at EBI (Study PRJEB6147, Accession range HACF01000001-HACF01058581).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 Gene Ontology categorization of *E. superba* annotated contigs.

Fig. S2 Integrated metabolic map of carbohydrate metabolism.

Fig. S3 Overview of metabolic pathways covered by annotated contigs of the *E. superba* 'master' transcriptome (Yamada *et al.* 2011).

Fig. S4 Integrated metabolic map of lipid and chitin metabolism.

Table S1 For reproducibility purposes, here we report the complete list of commands used in the assembly and annotation strategies.

Table S2 'Master' krill transcript catalogue.

Table S3 Classification of the annotated contigs into the 'Biological Process' GO class.

Table S4 Description of the putative transcripts grouped into the 'Metabolic process', 'Response to stress' and 'Biomineralization process' GO categories.