BMC Genomics

Research article

Systematic sequencing of mRNA from the Antarctic krill (*Euphausia superba*) and first tissue specific transcriptional signature

Cristiano De Pittà[†1,2], Cristiano Bertolucci[†3], Gabriella M Mazzotta[2], Filippo Bernante[1], Giorgia Rizzo[2], Barbara De Nardi[4], Alberto Pallavicini[4], Gerolamo Lanfranchi[1,2] and Rodolfo Costa*[2]

Address: [1]CRIBI Biotechnology Centre, Università degli Studi di Padova, Via U. Bassi, 58/B, 35121, Padova, Italy, [2]Dipartimento di Biologia, Università degli Studi di Padova, Via U. Bassi, 58/B, 35121, Padova, Italy, [3]Dipartimento di Biologia ed Evoluzione, Università degli Studi di Ferrara, Via L. Borsari, 46, 44100 Ferrara, Italy and [4]Dipartimento di Biologia, Università degli Studi di Trieste, P.le Valmaura, 9, 34148 Trieste, Italy

Email: Cristiano De Pittà - cristiano.depitta@unipd.it; Cristiano Bertolucci - bru@unife.it; Gabriella M Mazzotta - gabriella.mazzotta@unipd.it; Filippo Bernante - fbernante@cribi.unipd.it; Giorgia Rizzo - giorgia.rizzo@unipd.it; Barbara De Nardi - barbara.denardi@entecra.it; Alberto Pallavicini - pallavic@units.it; Gerolamo Lanfranchi - lanfra@cribi.unipd.it; Rodolfo Costa* - costa@mail.bio.unipd.it

* Corresponding author   †Equal contributors

Abstract

Background: Little is known about the genome sequences of Euphausiacea (krill) although these crustaceans are abundant components of the pelagic ecosystems in all oceans and used for aquaculture and pharmaceutical industry. This study reports the results of an expressed sequence tag (EST) sequencing project from different tissues of *Euphausia superba* (the Antarctic krill).

Results: We have constructed and sequenced five cDNA libraries from different Antarctic krill tissues: head, abdomen, thoracopods and photophores. We have identified 1.770 high-quality ESTs which were assembled into 216 overlapping clusters and 801 singletons resulting in a total of 1.017 non-redundant sequences. Quantitative RT-PCR analysis was performed to quantify and validate the expression levels of ten genes presenting different EST countings in krill tissues. In addition, bioinformatic screening of the non-redundant *E. superba* sequences identified 69 microsatellite containing ESTs. Clusters, consensuses and related similarity and gene ontology searches were organized in a dedicated *E. superba* database http://krill.cribi.unipd.it.

Conclusion: We defined the first tissue transcriptional signatures of *E. superba* based on functional categorization among the examined tissues. The analyses of annotated transcripts showed a higher similarity with genes from insects with respect to Malacostraca possibly as an effect of the limited number of Malacostraca sequences in the public databases. Our catalogue provides for the first time a genomic tool to investigate the biology of the Antarctic krill.

Background

Euphausiacea (krill) are small shrimplike crustaceans that are abundant in the pelagic ecosystems of all oceans. There are about 85 species of Euphausiacea, making this one of the smallest orders in the class of Malacostraca [1]. Phylogenetic analysis of the Eumalacostraca orders based on 28S rDNA sequences suggests that Euphausiacea are more closely related to Mysida than to the Decapoda [2].

In the Southern Ocean, krill is a critical link between primary productivity and most of the predators at higher trophic levels such as birds, fish, seals, squid and whales [3]. The krill biomass in the Southern Ocean has been estimated at 400–1550 million tons with sustainable annual harvest at around 70–200 million tons. Therefore, krill biomass that could be available for human food is comparable to the biomass of all the other aquatic species currently fished by humans, but only six species of krill are at present harvested commercially [4,5]. Commercial fishing of krill is done in the Southern Ocean and around Japan. The global annual production amounts to 150 – 200.000 tons, most of this from the Scotia Sea [6,7]. Most of the fished krill is used for aquaculture and aquarium feedings, as baits in sport fishing, or in the pharmaceutical industry.

The Antarctic krill (*Euphausia superba*, Dana 1852) has a circumpolar distribution with the highest concentrations in the Atlantic sector of the Southern Ocean. It is a key species of the Antarctic ecosystem and plays an important role both as feeder of algae, bacteria and micro-zooplancton and as a prey of vertebrates [8]. *E. superba* displays a large daily vertical migration that occurs generally within the upper 200 m water column making a significant amount of biomass available as food for predators near the surface at night and in deeper waters during the day [9]. Basic knowledge of crustacean biology is limited by the lack of information about their genomes. Considering all orders in the class of Malacostraca, no genome has yet been fully sequenced. At present Genbank carries 116,640 nucleotide and 11,932 protein sequences (Table 1), with a high rate of redundancy. Currently only 434 nucleotide and 310 protein sequences have been identified in Euphausiacea (GenBank source, release of November 2007). Specifically for *E. superba* only 69 nucleotide and 17 amino acid sequences have been obtained; they iden-

tify key proteins and enzymes of oxidative phosphorylation (*NADH dehydrogenase subunit 1, 2, 3, 4, 4L, 5; Cytochrome oxidase subunit I, II, III; ATP synthase subunit 6; cytochrome b; cytochrome b apoenzyme* and *cytochrome c oxidase subunit I*) and of phototransduction (*opsin*). In the subphylum Crustacea there are 33 complete (or nearly complete) mitochondrial DNA sequences: 4 Branchiopoda, 8 Maxillopoda, 18 Malacostraca, and one of Ostracoda, Cephalocarida and Remipedia (Table 2). In a previous investigation Machida *et al.* [10] determined the nearly complete DNA sequence of the mitochondrial genome of *E. superba* (14,606 bp).

The identification of novel shrimp genes by systematic sequencing of genomic DNA is hindered by the dispersion of the genes among large non-coding regions and by the presence of introns within genes. Current genomics technologies, like SAGE [11], differential display [12] and systematic sequencing of expressed sequence tags [13,14], are very useful approaches to identify protein coding genes rapidly on a large scale. Moreover, the frequency of a given sequence in the SAGE or cDNA libraries can be related to the relative abundance of the corresponding mRNA, giving an indication of the level of gene expression [15].

The aim of our study was to significantly increase the number of krill genes in the public database and to discover tissue specific genes. For this purpose we have produced and sequenced five cDNA libraries from different Antarctic krill tissues: head, abdomen, thoracopods and photophores. We have developed special cDNA libraries optimized to directionally cloning full-length cDNA in plasmid vectors without enzymatic digestion. We have identified 1,770 high-quality EST clones that have been grouped in 1,017 different clusters. Of these, 309 clusters

Table 1: Nucleotide and protein sequences belonging to all orders of the Malacostraca class available from public databases at November 2007.

Class	Subclass	Superorder	Order or Sub-order	Nucleotide sequences	Protein sequences
Malacostraca	Eumalacostraca	Eucarida	Decapoda	97,083	7,862
Malacostraca	Eumalacostraca	Eucarida	Euphausiacea	434	310
Malacostraca	Eumalacostraca	Hoplocarida	Stomatopoda	119	214
Malacostraca	Eumalacostraca	Peracarida	Amphipoda	14,732	1,745
Malacostraca	Eumalacostraca	Peracarida	Cumacea	93	82
Malacostraca	Eumalacostraca	Peracarida	Isopoda	3,355	1,155
Malacostraca	Eumalacostraca	Peracarida	Mictacea	1	0
Malacostraca	Eumalacostraca	Peracarida	Mysidacea	771	561
Malacostraca	Eumalacostraca	Peracarida	Spelaeogriphacea	1	0
Malacostraca	Eumalacostraca	Peracarida	Tanaidacea	18	1
Malacostraca	Eumalacostraca	Peracarida	Thermosbaenacea	4	0
Malacostraca	Eumalacostraca	Syncarida	Anaspidacea	27	2
Malacostraca	Eumalacostraca	Syncarida	Bathynellacea	2	0
Total				**116,640**	**11,932**

Table 2: List of complete or nearly complete mitochondrial DNA sequences of the subphylum Crustacea available from public databases at November 2007.

#	Class	Subclass	Organism	Refseq	Reference
1	Branchiopoda	Phyllopoda	*Triops longicaudatus*	NC_006079	[50]
2	Branchiopoda	Sarsostraca	*Artemia franciscana*	NC_001620	Unpublished
3	Branchiopoda	Phyllopoda	*Daphnia pulex*	NC_000844	[51]
4	Branchiopoda	Phyllopoda	*Triops cancriformis*	NC_004465	[52]
5	Cephalocarida	Brachypoda	*Hutchinsoniella macracantha*	NC_005937	[53]
6	Malacostraca	Eumalacostraca	*Halocaridina rubra*	NC_008413	[54]
7	Malacostraca	Eumalacostraca	*Geothelphusa dehaani*	NC_007379	[55]
8	Malacostraca	Eumalacostraca	*Ligia oceanica*	NC_008412	[56]
9	Malacostraca	Eumalacostraca	*Squilla mantis*	NC_006081	[50]
10	Malacostraca	Eumalacostraca	*Harpiosquilla harpax*	NC_006916	[57]
11	Malacostraca	Eumalacostraca	*Pseudocarcinus gigas*	NC_006891	[58]
12	Malacostraca	Eumalacostraca	*Macrobrachium rosenbergii*	NC_006880	[58]
13	Malacostraca	Eumalacostraca	*Callinectes sapidus*	NC_006281	Unpublished
14	Malacostraca	Eumalacostraca	*Squilla empusa*	NC_007444	Unpublished
15	Malacostraca	Eumalacostraca	*Lysiosquillina maculata*	NC_007443	Unpublished
16	Malacostraca	Eumalacostraca	*Gonodactylus chiragra*	NC_007442	Unpublished
17	Malacostraca	Eumalacostraca	*Marsupenaeus japonicus*	NC_007010	[59]
18	Malacostraca	Eumalacostraca	*Eriocheir sinensis*	NC_006992	[60]
19	Malacostraca	Eumalacostraca	*Cherax destructor*	NC_011243	[61]
20	Malacostraca	Eumalacostraca	*Panulirus japonicus*	NC_004251	[62]
21	Malacostraca	Eumalacostraca	*Portunus trituberculatus*	NC_005037	[63]
22	Malacostraca	Eumalacostraca	*Pagurus longicarpus*	NC_003058	[64]
23	Malacostraca	Eumalacostraca	*Penaeus monodon*	NC_002184	Unpublished
24	Maxillopoda	Copepoda	*Tigriopus californicus*	NC_008831	Unpublished
25	Maxillopoda	Cirripedia	*Pollicipes mitella*	NC_008742	Unpublished
26	Maxillopoda	Branchiura	*Argulus americanus*	NC_005935	[50]
27	Maxillopoda	Copepoda	*Tigriopus japonicus*	NC_003979	[65]
28	Maxillopoda	Copepoda	*Lepeophtheirus salmonis*	NC_007215	[66]
29	Maxillopoda	Cirripedia	*Megabalanus volcano*	NC_006293	Unpublished
30	Maxillopoda	Cirripedia	*Pollicipes polymerus*	NC_005936	[53]
31	Maxillopoda	Cirripedia	*Tetraclita japonica*	NC_008974	Unpublished
32	Ostracoda	Myodocopa	*Vargula hilgendorfii*	NC_005306	[67]
33	Remipedia	Nectiopoda	*Speleonectes tulumensis*	NC_005938	[53]

were successfully annotated while 708 did not show a significant similarity with known genes from other organisms. Clusters, consensus and related similarity and gene ontology searches were organized in a dedicated *E. superba* database [16].

Results and Discussion

Construction of cDNA libraries and EST analysis

Total RNA was independently isolated from tissues (head, abdomen, thoracopods and photophores) dissected from specimens of *E. superba* collected at five different time points during 24 hours. The analysis of total RNA samples, performed by capillary electrophoresis, showed absence of genomic DNA contamination and a peculiar electropherogram as shown in Fig. 1. In particular, total RNAs show low molecular weight from 200 bp to 1 kb, perhaps as result of a partial RNA degradation.

Five independent tissue-specific cDNA libraries, named K01 and K05 (head), K06 (abdomen), K07 (photo-

phores) and K09 (thoracopods), were produced from total RNA pools. For head we have sequenced only K05 cDNA library because it presented more recombinant clones compared to K01. Recombinant bacterial clones from each library were randomly picked and the EST were sequenced from the 5'-end. The average insert size for all libraries was estimated to be 412 bp.

EST assembly and construction of an Antarctic krill transcript catalogue

A total of 2,046 ESTs were initially analyzed for sequence quality and vector sequences were recognized and deleted. Two-hundred-seventy-six low quality ESTs were removed and 1,770 (86.5%) high-quality ESTs were further processed. These ESTs assembled by similarity into 216 clusters and 801 singletons, resulting in a total of 1,017 non-redundant (consensus) sequences. A list of the sequencing trend for each cDNA library is presented in Table 3. Interestingly, we obtained a low percentage of clusters composed by ≥ 2 ESTs from cDNA libraries prepared from the
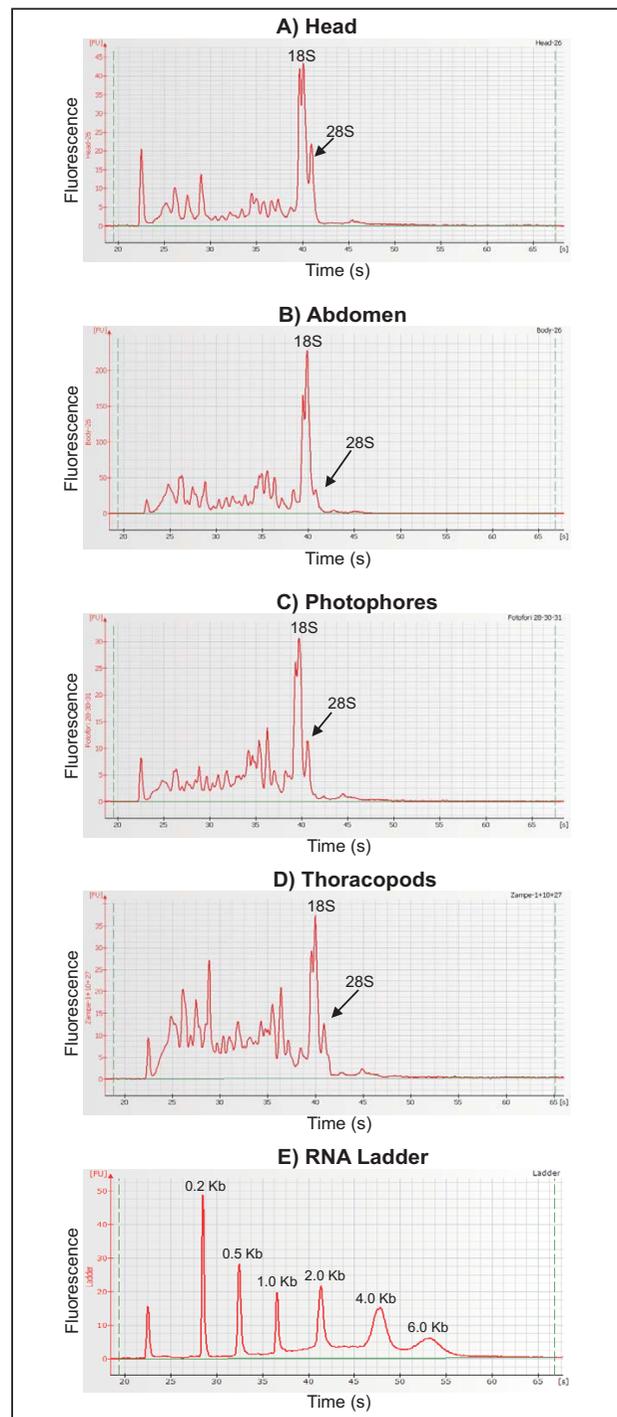
Figure 1
Electropherograms of *E. superba* tissue-specific total RNAs. (**A-D**) Electropherograms resulting from Agilent 2100 bioanalyzer analysis on total RNA extracted from head, abdomen, photophores and thoracopods. X-axis: time of ribosomal RNA peak appearance, corresponding to the size of the fragment; Y-axis: fluorescence of the peak, corresponding to its concentration. The size and the concentration of the sample peaks are calculated by the software via comparison with a RNA ladder at known concentration (**E**). *E. superba* RNA samples showed some products with a migration time between 22 and 35 seconds (from 200 bp to 1.000 bp) indicating a partial RNA degradation.

Table 3: Results of EST assembly for each krill cDNA library.

Library	Tissue	Total ESTs	Discarded ESTs	Analyzed ESTs	# GenBank	# EST in cluster	# cluster	# singletons	# consensus	% redundancy	% discovery
K01	Head	35	9	26	ES542703–ES542728	12	10	14	24	38.46	61.54
K05	Head	579	124	455	ES542729–ES543183	225	108	230	338	35.16	64.84
K06	Abdomen	508	47	461	ES543184–ES543644	250	110	211	321	42.73	57.27
K07	Photophores	506	52	454	ES543645–ES544098	255	114	199	313	49.78	50.22
K09	Thoracopods	418	44	374	ES544099–ES544472	227	108	147	255	55.61	44.39
Overall		2046	276	1770		969	450	801	1251	44.35	55.65

Total ESTs = number of produced chromatograms; Discarded ESTs = number of low quality ESTs; Analyzed ESTs = number of sequences processed for clustering; # GenBank = accession numbers of ESTs deposited in GenBank; # EST in cluster = number of sequences in cluster; # clusters = total number of clusters; # singletons = number of putative transcripts identified by one EST; # consensus = number of non-redundant sequences; % redundancy = percentage of sequences not identified in an exclusive manner; % discovery = percentage of ESTs identifying putative new transcripts in the total EST analyzed.

krill head (K01, K05) showing that no specific transcripts are particularly enriched in this tissue. As expected from the above data, the head reveals a higher percentage (about 65%) of new putative transcripts, with respect to abdomen (52.3%), photophores (50.2%) and thoracopods (44.4%) in which the presence of enriched transcripts was instead revealed by EST sequencing.

The number of ESTs in each cluster varies from 2 (93 clusters) to 88 (1 cluster). The average length of a cluster is 436 bp with the longest assembled sequence being 1,417 bp and the shortest 153 bp. The 91% of cluster consensus contains the 3'-end region of mRNAs as demonstrated by the presence of a poly-adenylation signal.

Each non redundant sequence was searched in the nucleotides database and UniProtKB database using Blast-N and Blast-X with an *e*-value cut off of < e^{-40} and < e^{-10}, respectively. These values were empirically chosen considering the low amount of sequences data available for *Euphausiacea* and similar shrimp species and the need of stringency in providing a reliable catalogue of Antarctic krill genes. All annotations were further manually examined, in order to assign the best describing text to the correspondent cluster.

Overall, 70% of non-redundant sequences (708 out of 1,017), identified by about 50% of total produced ESTs, showed no or poor similarity matches, and they probably represent completely unknown Antarctic krill transcripts that could be characterized in future studies. Additional file 1 lists the 309 (30%) non-redundant sequences identifying known *Euphausia* genes or sequences showing significant similarity to genes from arthropods (63.4%) and other species (36.6%), such as *Homarus americanus* (3.2%), *Aedes aegypti* (12.0%), *Drosophila melanogaster* (3.9%), *Bombyx mori* (3.6%), *Brachydanio rerio* (4.9%), *Mus musculus* (3.9%), *Rattus norvegicus* (3.9%), *Homo sapiens* (4.5%). Antarctic krill sequences generally show a greater similarity to genes of insects (about 36%) than to genes of Malacostraca (about 23%) and only 7% were similar to the known sequences of Euphausiacea. This result could be due to the limited number of Malacostraca gene and protein sequences available in the public databases (*gene sequences: 116,666 and protein sequences: 11,941 at November 2007*) respect to insects (*gene sequences: 1,869,511 and protein sequences: 336,246 at November 2007*). The total collection of 1,770 *E. superba* 3'-EST has been deposited in the EBI-GenBank-DBJ database (GenBank accession numbers from ES542703 to ES544472).

One of the most noticeable features of our EST catalogue is that mitochondrial transcripts are quite abundant (about 10% of the total ESTs): 84 ESTs (about 4.7%) matched with nine different *E. superba* mitochondrial transcripts (*ATPase6*, *COI*, *COII*, *COIII*, *cyt b*, *ND1*, *ND2*, *ND3*, *ND6*) and about 5% ESTs (88 out of 1,770) identified the large mitochondrial 16S rRNA gene. In future experiments, to avoid the repetitive sequencing of this abundant mRNA, we plan to introduce, during cDNA insert amplification, interference primers specifically designed for the 16S rRNA [17]. Moreover, we found a small percentage of ESTs (1.6%), showing similarity with 18S and 28S rRNA. About 2% of ribosomal RNAs contamination is common to other systematic sequencing projects.

The *E. superba* mitochondrial gene sequences are very similar to those commonly found in the mitochondrial genomes of other arthropods, including 13 protein-coding genes, 19 tRNA and 2 rRNA genes [10]. Machida *et al.* [10] have demonstrated that mitochondrial protein-coding genes are transcribed from the same DNA strand, left to right, except for *ND1*, *ND4L*, *ND4* and *ND5* genes, while two ribosomal RNA genes are encoded by L strands.

We did not find highly represented tissue-specific mRNAs, while some ribosomal proteins like L22 (31 ESTs), S25 (31), P1 (16), L37A (15), L24 (13), P2 (13), S14 (9), S14 (10), L34 (9) were expressed at the same level in all analyzed krill tissues. The only genuine tissue-specific mRNA appears to be *myosin light chain* (8 ESTs, cluster KRC00032) that is highly represented in a strictly commit-

ted tissue such as skeletal muscle localized in abdomen (library K06). Moreover, we found 3 unknown transcripts among the 20 most expressed genes: KRC00118 (10 ESTs), KRC00407 (9 ESTs) and KRC00101 (7 ESTs) that could be interesting for future functional studies.

Since very few abundant transcripts were found that could hamper the identification of rare transcripts, it seems plausible that random sequencing of our Antarctic krill libraries would continue to represent an effective strategy for identifying novel *E. superba* mRNAs.

Functional categorization of E. superba *ESTs*

In order to facilitate functional genomic studies in Antarctic krill, 309 consensus sequences showing similarity with known genes or proteins were grouped into 14 functional categories (Table 4) according to Gene Ontology [18] and other resources developed for gene functional annotation [19]. A list of all annotated transcripts is shown in Additional file 1. A large majority of ESTs (20.51%), displaying putative identity with ribosomal sequences and genes for the translation machinery, were grouped in the transla-

tion functional category, characterized by 8% of all known transcripts. We found genes with regulative functions in the translational initiation, like *translation factor SUI1*, *initiation factor 4A and 3*, in the translational elongation like *elongation factor 1α, 1β, 2* and a specific *tail muscle elongation factor 1γ*. Other abundant *E. superba* sequences fall into gene categories related to cell structure, cell motility and functional homeostasis. For instance, genes involved in the mechanisms of DNA transcription (1.47%), transport (1.97%), skeletal muscle contraction (3.4%), and in amino acid, fatty acid and carbohydrate metabolism (4.03 %) are comprised in this category. This class includes also SEC61 β-subunit, an important transport protein that plays a crucial role in the insertion of secretory and membrane polypeptides into the endoplasmic reticulum and cellular retinoic acid/retinol binding protein (RBP1), involved in the transport of retinol from the digestive gland to peripheral tissues [20]. A transcript included in this class (ID: KRC00589) shows a good similarity with hemocyanin, the main oxygen carrier molecule in arthropods and molluscs [21].

Table 4: Classification of the 1,017 krill unique consensus sequences in functional categories.

#	Functional categories	Biological process	# Clusters	% Cluster	# ESTs	% ESTs
I	**DNA replication & binding**	DNA replication	4	**0.79**	**8**	**0.45**
		DNA binding	4			
2	**Transcription**	Transcription process	2	**1.47**	**16**	**0.90**
		Regulation of transcription	8			
		Chromatine assembly and remodelling	5			
3	**Translation**	Translation	70	**8.06**	**363**	**20.51**
		Translation initiation	4			
		Translation elongation	7			
		Translation termination	I			
4	**Transport**	Electron transport	8	**1.97**	**32**	**1.81**
		Calcium ion transport	2			
		Intracellular protein transport	4			
		Transport	6			
5	**Metabolic processes**	Amino acid metabolic process	2	**4.03**	**73**	**4.12**
		Carbohydrate metabolic process	13			
		Fatty acid biosynthetic process	3			
		Various metabolic process	23			
6	**Proteolysis, protein folding and modification**	Proteolysis	16	**3.24**	**48**	**2.71**
		Protein modification and binding	9			
		Protein folding	8			
7	**Striated muscle contraction**	Contraction	26	**3.54**	**95**	**5.37**
		Calcium binding	10			
8	**Signal transduction**		8	**0.79**	**9**	**0.51**
9	**Structural constituent of cuticle**		14	**1.38**	**30**	**1.69**
10	**Cell growth, proliferation and adhesion**		9	**0.88**	**12**	**0.68**
II	**Ion Binding**		5	**0.49**	**5**	**0.28**
12	**Hypotetical protein**		10	**0.98**	**16**	**0.90**
13	**Mitochondrial genes**		14	**1.38**	**174**	**9.83**
14	**Ribosomal RNA gene (18S and 28S)**		14	**1.38**	**29**	**1.64**
15	**Without similarity**		708	**69.62**	**860**	**48.59**
	TOTAL		**1017**	**100.00**	**1770**	**100.00**

The last four columns of the Table show respectively number and percentage of clusters and ESTs belonging to each functional category.

Other interesting krill transcripts that we were able to annotate are those involved in stress responses, proteolysis and immunoresponse (3.24%) like *Hsp90*, *chaperones*, *cathepsine L-like cysteine protease*, a lysosomal cysteine proteinase [22], *peptidyl-prolyl cis-trans isomerase A1* (*cyclophilin* A1) and *peptidylprolyl isomerase B* (*cyclophilin* B). Cyclophilines are members of the immunophilin protein family, which play a role in immunoregulation and basic cellular processes involving protein folding and trafficking. ESTs with good similarity to hemocyanin are present in our collection: this protein has been recently reported to have antifungal and antiviral activities [23,24].

Some krill ESTs identify histone 2A (KRC00431) and histone 3.3A (KRC00024) indicating the presence of unexpected polyadenylated histone transcripts displaying the polyadenylation signal and tail. In vertebrates, these evolutionary conserved housekeeping mRNAs are not polyadenylated, and this has been related to the high turnover of these transcripts in the dividing cells. Interestingly, polyadenylated H2A and H3 histone sequences were detected also in the systematic sequencing of 3'-end cDNA libraries obtained from brain and kidney of channel catfish *Ictarulus punctatus* [25,26] and from various tissues (haemolymph, gills, digestive glands, mantles and adductor muscles) of the mussel *Mytilus galloprovincialis* [27]. The presence of polyadenylation signals in *E. superba* histone transcripts deserves a more detailed analysis. In fact, Eirin-Lopez *et al.* [28] have recently shown that all histone genes in the repetitive unit are characterized by two different mRNA termination signals in their 3' UTR: the typical stem-loop or hairpin-loop signal followed by a purine-rich element and a polyadenylation signal AATAAA located downstream to this last element. The presence of a double mRNA termination signal is unique to histone genes and common for other invertebrates such as *Chironomus thummi* [29], *D. melanogaster* [30], *Chaetopterus variopedatus* [31], *M. galloprovincialis* [32] and Crustacea [33]. Although in some invertebrates core histone transcripts (H2A, H2B, H3 and H4) include polyA tails, these sequences are among the most evolutionary conserved eukaryotic proteins [34].

Transcriptional signature of E. superba *tissues*
Gene expression profiling depends on the functional specificity of cells composing different tissues. So, the systematic sequencing of EST from unbiased cDNA libraries is a suitable approach for analyzing the gene expression profile of a given tissue [35]. In fact, the frequency of a given EST in the cDNA library can be related to the relative abundance of the corresponding mRNA in the source tissue.

To define tissue transcriptional signatures of *E. superba*, annotated ESTs obtained from the four tissue-specific cDNA libraries (head, abdomen, photophores, thoracopods) were separately grouped in 13 functional categories (Table 5) a further abundant category was created for those ESTs to which no function may be yet associated. Fig. 2 shows four different diagrams standing for ESTs distribution among functional categories in each cDNA library. The presence of highly represented functional categories is peculiar of strictly committed tissues such as abdomen and thoracopods in which transcripts involved in striated muscle contraction are very abundant (about 26% in abdomen and 7% in thoracopods). In the abdomen library, we were able to recognize the principal structural components of the sarcomeric contractile machinery (*myosin heavy chain*, *myosin light chain 1*, *myosin-2*, *actin*, *alpha-tubulin*, *tropomyosin*) and two subunits of the troponin complex (*troponin T*, *troponin I*), a key regulator of muscle contraction. About 10% of sequences produced from head and thoracopods libraries fall in functional categories related to metabolic processes (amino acid, fatty acid and carbohydrate metabolism). Interestingly, about 6% and 4% of ESTs respectively sequenced from head and thoracopods libraries identified structural constituents of cuticle (*arthrodial cuticle protein AMP16.3*, *arthrodial cuticle protein AMP1A*, *calcification-associated peptide-1 precursor*). This reflects the presence of cuticle traces in the head and thoracopods samples. In photophores and thoracopods transcripts displaying putative identity with ribosomal sequences are more abundant compared to other tissues (55% and 46%, respectively), indicating a relevant activity of the translation machinery.

We have also identified from the head cDNA library a novel opsin sequence (ID ESTs: KRC00735, KRC00802), a light-sensitive membrane-bound G protein-coupled receptors mediating the conversion of a photon of light into an electrochemical signal in the visual transduction cascade. In insects there are at least four main spectral classes: long-wavelenght-sensitive (LWS), middle-wavelenght-sensitive (MWS) and two short-wavelenght-sensitive (SWS) groups. The opsin sequences available for *E. superba* (GenBank accession no. DQ852576–DQ852580) show a spectral sensitivity with short wavelength (496–501 nm, λ_{max} = 487) and cannot be aligned with our consensus [36].

Quantitative RT-PCR analysis was performed to quantify and validate the expression level of some genes presenting different EST countings in krill tissues. We selected ten genes (*compound eye opsin BCRH1*, *myosin light chain*, *myosin heavy chain*, *arthrodial cuticle protein AMP16.3*, *tail muscle elongation factor 1 gamma*, *cellular retinoic acid/retinol binding protein*, *eukaryotic initiation factor 4A*, *transport protein SEC61 subunit gamma*, *chromodomain helicase DNA binding protein* and *voltage-dependent calcium channel*) representative of different levels of transcript abundance.

Table 5: Classification of the annotated ESTs for each library.

#	Process	Head		Abdomen		Photophores		Thoracopods	
		# ESTs	%	# ESTs	%	# ESTs	%	# ESTs	%
1	DNA replication	3	1.43	2	0.90	2	0.88	1	0.45
2	Transcription	3	1.43	6	2.71	5	2.19	2	0.90
3	Translation	69	32.86	65	29.41	126	55.26	103	46.40
4	Transport	5	2.38	12	5.43	7	3.07	8	3.60
5	Metabolic process	22	10.48	16	7.24	14	6.14	21	9.46
6	Proteolysis, protein folding and modification	12	5.71	6	2.71	14	6.14	16	7.21
7	Striated muscle contraction	11	5.24	57	25.79	12	5.26	15	6.76
8	Signal transduction	4	1.90	1	0.45	3	1.32	1	0.45
9	Structural constituent of cuticle	13	6.19	1	0.45	7	3.07	9	4.05
10	Cell growth, proliferation and adhesion	2	0.95	3	1.36	2	0.88	5	2.25
11	Ion Binding	4	1.90	0	0.00	1	0.44	0	0.00
12	Hypotetical protein	4	1.90	6	2.71	6	2.63	0	0.00
13	Mitochondrial genes	58	27.62	46	20.81	29	12.72	41	18.47
	Total	**210**	100.00	**221**	100.00	**228**	100.00	**222**	100.0

Table showing the number and the percentage of ESTs in each functional category for each tissue specific library

The housekeeping gene 18S rRNA was used as endogenous control. As reported in the Additional file 2, the expression values obtained with the quantitative RT-PCR for the tested transcripts were in agreement with the EST counting in the four libraries. In particular, we have demonstrated that the *compound eye opsin* is strongly expressed in head compared to other tissues and *myosin light chain* and *myosin heavy chain* are highly expressed in abdomen and thoracopods confirming their key role in the contractile machinery. Instead, the *eukaryotic initiation factor 4A* is expressed at about the same level in all tested tissues.

Identification of microsatellite-containing ESTs
Among the 1,017 non-redundant sequences examined in this study, 41 (4%) consensus sequences containing ESTs were identified by using MISA software. Twelve of these consensus present 2 distinct simple sequence repeats interrupted by more than 100 bp for a total of 69 identified microsatellites (SSR). The majority of these sequences (72%) fall into the 3 bp repeat type class with a preponderance of GAA and GAT. After a manual inspection of redundancy, raw sequence, data quality and the presence of sufficient flanking sequences we designed 9 pairs of specific PCR primers. We obtained successful amplifications for 6 of these 9 pairs of primers. Assessment of polymorphism information content (PIC), observed and expected heterozygosity and other population genetics analysis will be performed in the near future. These markers will increase the currently available Euphausiacea SSR markers. In fact, only five microsatellite loci isolated from the northern krill *Meganyctiphanes norvegica* have been reported so far [37]. Since our novel microsatellite markers were developed on the basis of expressed sequences

and they are presumably conserved across other Euphausiacea species, they could also be useful for comparative mapping and for a molecular approach to Antarctic krill ecology.

Conclusion
Since genome sequencing and BAC libraries of Antarctic krill are not yet available, EST sequencing from randomly selected cDNA clones represents a powerful approach to identify large numbers of transcripts that could be used in gene expression and functional genomics studies [38]. The systematic sequencing of four cDNA libraries prepared from different *E. superba* tissues has allowed us to establish an EST database containing 1,017 unique sequences. Over 65% of the Antarctic krill sequences resulted in no BLAST matches with published sequences and they probably represent novel genes that could be functionally characterized. We have defined the transcriptional signatures of krill tissues and performed qRT-PCR to validate the level of expression of ten representative genes. All sequencing data have been deposited on the *E. superba* EST database available from our web site [16]. In addition, the EST collection is a potential source for the development of genetic markers including microsatellite and single nucleotide polymorphisms. Among the 1,017 unique sequences, 41 (4%) unique microsatellite containing ESTs were identified by using MISA software. Moreover, we have designed and successfully tested 6 pairs of specific PCR primers for microsatellite loci.

Our EST catalogue could provide a source for the design of microarray platform that will allow the study of the
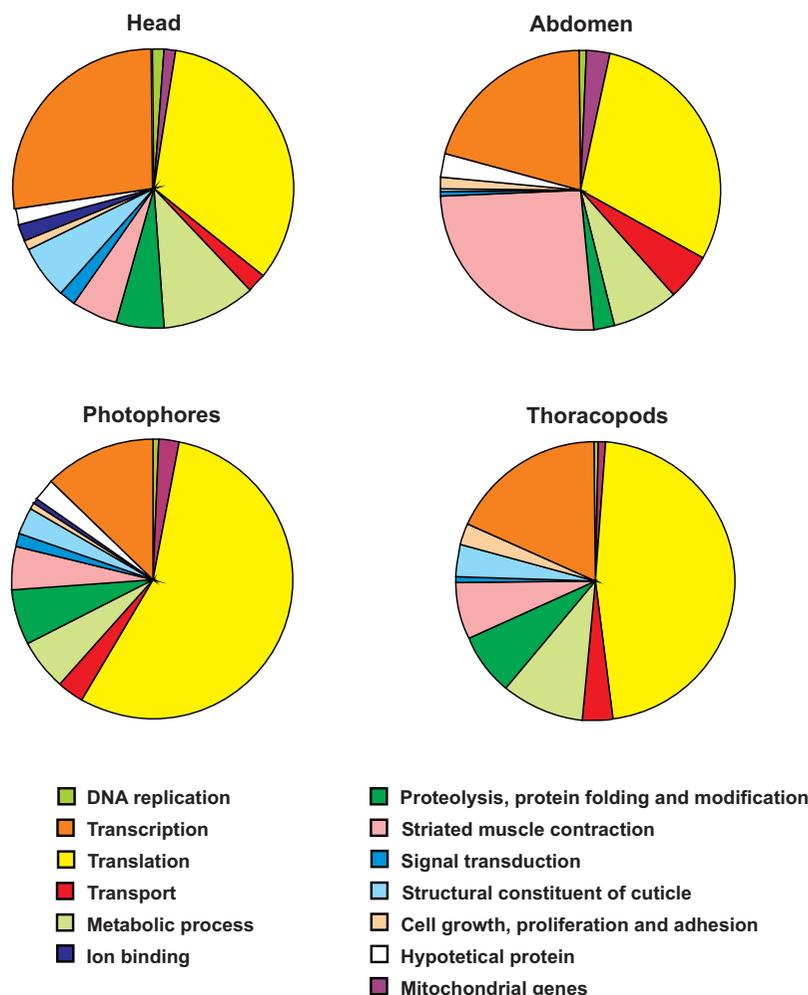
Head

Abdomen

Photophores

Thoracopods

- ☐ (green) DNA replication
- ☐ (orange) Transcription
- ☐ (yellow) Translation
- ☐ (red) Transport
- ☐ (light green) Metabolic process
- ☐ (dark blue) Ion binding
- ☐ (dark green) Proteolysis, protein folding and modification
- ☐ (pink) Striated muscle contraction
- ☐ (blue) Signal transduction
- ☐ (light blue) Structural constituent of cuticle
- ☐ (tan) Cell growth, proliferation and adhesion
- ☐ (white) Hypotetical protein
- ☐ (purple) Mitochondrial genes

Figure 2
Classification of the annotated ESTs for each library into different 13 functional categories. Diagrams showing the proportion of each functional category in all four tissues. See Table 5 for more details.

transcriptional responses of this abundant marine organism to environmental challenges [39].

Methods
Tissues samples, RNA extraction and quality control
Antarctic krill (*Euphausia superba*) were fished from the Ross Sea (longitude: 167°28'81" E – 179°54'68" W, latitude 68°40'54" S – 77°01"81" S) in the January 2004 during the XIX Italian Antarctic Expedition. Specimens were collected at different time of the day (01:00, 06:00, 10:00, 15:00, 18:00), over a complete 24-hour cycle. Samples were frozen at -40°C in RNA stabilization solution (RNAlater, Ambion). For each fishing, selected tissues (head including compound eyes and brain, abdomen, thoracopods and photophores) from five animals were dissected individually in RNA later ice solution (Ambion). After dissection, tissues were rapidly rinsed in sterile water,

weighed, frozen in Trizol reagent (Invitrogen) and stored at -80°C. A large excess of Trizol (15 ml for 0.5–1.5 g. of sample) was used in order to prevent RNA degradation by endogenous RNAse. Frozen tissues were minced and homogenized for 3–5 min using an ultra-turrax-T8.01 blender (IKA-Werke). Total RNA was isolated using the Trizol reagent (Invitrogen) following the manufacturer's instruction and further purified with LiCl in order to remove glucidic contaminants. All RNA samples were checked for quality by capillary electrophoresis (RNA 6000 Nano LabChip, Agilent Bioanalyzer 2100, Agilent Technologies). For each tissue (head, abdomen, thoracopods and photophores), equal amounts of total RNA (2 µg) extracted from every collection were pooled.

Construction of cDNA libraries

Five independent cDNA libraries, named K01 and K05 (head), K06 (abdomen), K07 (photophores) and K09 (thoracopods), were constructed.

We have developed a new method using a combination of SMART protocol (Clontech), ensuring almost full-length cDNA, and Gateway technology (Invitrogen), allowing unidirectional cloning without enzymatic digestion. In this protocol, only fully-transcribed first strand cDNA (ss cDNA) is tagged with a short sequence complementary to a modified SMART oligo (Fig. 3). The SMART oligo sequence (SMART-16attB1-T3: 5'-TACAAAAAAGCAG-GCTAATTAACCCTCACTAAAGGG-3') and the overhang of the oligo(dT) primer (5'-GGGGACCACTTTGTACAA-GAAAGCTGGGCGGCCGC [dT]$_{20}$VN-3') used for first strand synthesis include an attB1 and attB2 recombination site respectively.

First strand cDNA synthesis was performed from 1.5 µg of total RNA in a 15 µl reaction. Then, the reaction was then diluted 1:5 ratio and incubated at 72°C for 2 min. Second strand reaction mix was added to 1 µl of diluted first strand cDNA to give a final concentration of 1× BD Advantage 2 PCR reaction buffer (Clontech), 0.2 mM dNTPs, 120 nM primers (attB1-8T3: 5'-GGGGACAAGTTT-GTACAAAAAAGCAGGCTAATTAACC-3' and attB2: 5'-GGGGACCACTTTGTACAAGAAAGCTGGG-3') and 1× of Advantage2 DNA polymerase mix (Clontech) in a volume of 50 µl. This second strand reaction mixture was incubated for 21 cycles of 15 sec 95°C, 30 sec 66°C and 3 min 68°C. Only those ss cDNAs having a SMART anchor sequence at the 5' end were used as template and exponentially amplified. The second strand reaction was glass fibre column purified and cDNA was size selected by Sepharose CL-4B SPUN COLUMN (GE Healthcare). The cDNA was inserted in the cloning vector by a recombination reaction performed at 25°C for 18 h with about 35 ng of attB-cDNA, 150 ng of pDONR221 (Invitrogen) and 2 µl of BP Clonase II (Invitrogen) in 10 µl final volume. 1/5 of the purified reaction was used to transform electrocompetent DH10B *E. coli* cells. Recombinant colonies were selected on agar SOB medium plus kanamicin. Individual library colonies were arrayed by manual picking on 96 well plates in liquid selective SOB medium plus 7.5% of glycerol for independent growth [40].

DNA sequencing

After lysis of the bacterial colonies, cDNA inserts were directly amplified with universal primers M13 forward (5'-TGTAAAACGACGGCCAGTCTTA-3') and M13 reverse (5'-CAGGAAACAGCTATGACCATGT-3'). The polymerase chain reaction (PCR) profile consisted of: (1) initial denaturation for 5 min at 95°C, (2) 35 cycles of 40-s denaturation at 95°C, 40-s annealing at 60°C and 1-min

elongation at 72°C and (3) final extension for 5 min at 72°C; samples with a size over 0.5 Kb were selected for sequencing. Single pass DNA sequencing from plasmids was performed by using the vector specific primer attB1_seq (5'-CTTTGTACAAAAAAGCAGGCT-3') and a modified Sanger dideoxy terminator cycle sequencing chemistry, the ABI BigDye kit version 3.1, on a ABI 3730 48-capillary sequencer and 36 cm capillaries (Sequencing Service of Max-Plank Institute for Molecular Genetics, Berlin, Germany).

Computer Management of Data

Trace2dbest and Partigene [41] were used to process chromatograms, clusterize sequences, and build an annotation database. Trace2dbest extracts sequences and quality information from traces (Phred algorithm), removes vector contamination and poly(A), and performs the trimming of low quality sequences. Sequences shorter than 100 bp were discarded. Partigene reads all sequence files and performs an assembling process in two step: 1) CLOBB software [42] clusterizes sequences on the basis of BLAST similarity; 2) Phrap [43] makes a consensus from each cluster.

Each consensus, converted in FASTA format, was searched locally in nucleotides database, downloaded from NCBI [44] and UniProtKB database [45], using Blast-N and Blast-X, respectively. First 10 HSPs (High Scoring Pair) from each blast result were collected and stored in a local PostgreSQL table, as a collection of automatic annotations.

Each cluster annotation in our database was further manually examined to assign the best describing text to the correspondent cluster: matches with expectations values greater than e^{-10} for protein (Blast-X) and e^{-40} for nucleotide (Blast-N) were considered as poorly informative. Moreover, for each UniProt ID, taken from Blast-X description field, we associated specific Gene Ontology annotation, that integrates information about process, function, and component. Clusters, consensus and related similarity and gene ontology searches were electronically organized and stored in a dedicated PostgreSQL database.

Identification of microsatellite containing ESTs

The unique sequences were screened for microsatellites by using the MISA software [46]. Only di-, tri-, tetra-, penta- and esanucleotide repeats were targeted, since mononucleotide repeats are not useful for mapping or population genetics due to difficulties in their genotyping. Strings of oligonucleotide sequences were used to search for microsatellites: 6 repeats for dinucleotide; 5 repeats for trinucleotide; 5 repeats for tetranucleotide, pentanucleotide and esanucleotide. Primers were designed for the flanking regions of the SSR using a web-based software, "Primer3"
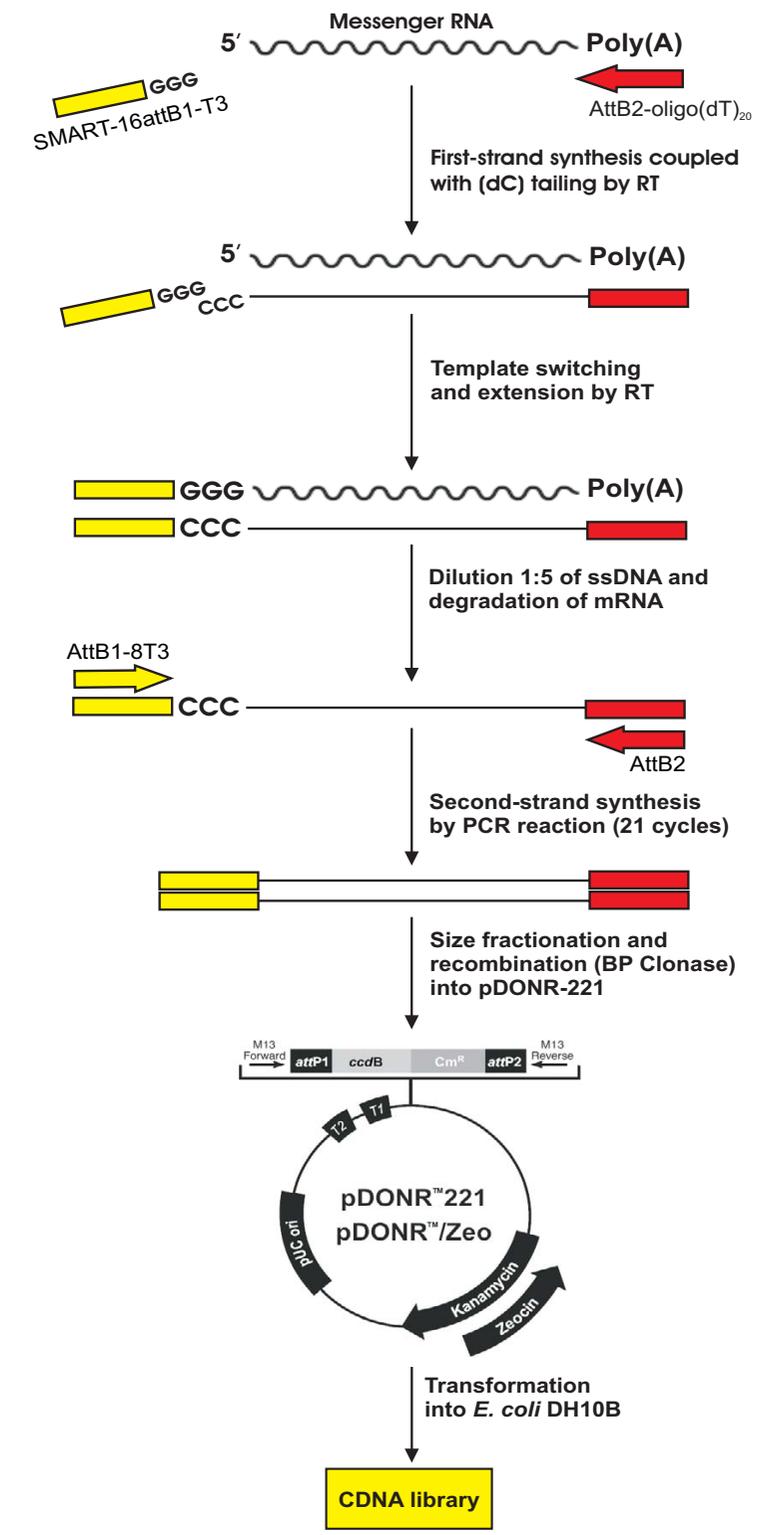
Figure 3
Schematic diagram of the method used for the construction of the cDNA libraries from different krill tissues. See the Methods for more details.

[47], and based on the criteria of 50% GC content, a minimum melting temperature of 55°C, and absence of secondary structure. Primers ranged from 18–27 nucleotides in length and amplified products of 150–390 bp. The primers were synthesized with a 5'-KS-tail (KS sequence: 5'-cgaggtcgacggtatcg-3') allowing to amplify microsatellite alleles in combination with a 5'-fluorescent-labeled KS primer [48].

Quantitative RT-PCR

Quantitative RT-PCR was conducted for some genes using the same tissues tested (head, abdomen, thoracopods and photophores) to confirm the integrity and robustness of EST sequencing.

Three μg of total RNA from each tissue was used to perform three independent cDNA syntheses in a final volume of 10 μl, using random decamers and SuperScript II reverse transcriptase (Invitrogen). 1 μl aliquot of diluted first-strand cDNA was PCR amplified in 10 μl volume using SYBR Green chemistry, according to the manufacturer's recommendations (Finnenzymes). Gene-specific primers were designed using Primer Express® Software (Applera) to amplify fragments of 120–180 bp in length, close to the 3'-end of the transcript. To avoid the amplification of contaminant genomic DNA, we treated total RNA samples with DNase I (Qiagen). The dissociation curve was used to confirm the specificity of the amplicon. PCR reactions were performed in a 7500 Real-Time PCR System (Applied Biosystems). Thermal cycling conditions were as follows: 15 min denaturation at 95°C; followed by 40 cycles of 30 sec denaturation step at 95°C, annealing and elongation steps for 1 min each at 60°C and a final 3 min elongation at 72°C. To evaluate differences in gene expression a relative quantification method was chosen where the expression of the target gene is standardized by a non-regulated reference gene; consequently, three replicates of each sample and endogenous control were amplified. 18S rRNA was used as an endogenous control because the level of rRNA remains essentially constant from sample to sample (QuantumRNA™ 18S Internal Standards, Ambion). To calculate the relative expression ratio, the $2^{-\Delta\Delta Ct}$ (RQ, relative quantification) method implemented in the 7500 Real Time PCR System software [49] was used. This method determines the change in expression of a nucleic acid sequence (target) in a test sample relative to the same sequence in a calibrator sample. In our experiments, the expression of ten targets was tested (*compound eye opsin BCRH1, myosin light chain, myosin heavy chain, arthrodial cuticle protein AMP16.3, tail muscle elongation factor 1 gamma, cellular retinoic acid/retinol binding protein, eukaryotic initiation factor 4A, transport protein SEC61 subunit gamma, chromodomain helicase DNA binding protein and voltage-dependent calcium channel*)

which displayed differential expression in the head, thoracopods and photophores, compared with the abdomen.

Availability and requirements

Project name: Systematic sequencing of mRNA from the Antarctic krill (*Euphausia superba*);

Project home page: http://krill.cribi.unipd.it;

Operating system(s): Debian GNU/Linux;

Programming language: PHP;

Licence: none;

Any restrictions to use by non-academics: none.

Authors' contributions

CDP performed total RNA sample preparation, construction and systematic sequencing of the cDNA libraries, annotation of ESTs, qRT-PCR and drafted the manuscript. CB conceived the study, carried out ESTs analysis and drafted the manuscript. GMM and GR participated in systematic sequencing of cDNA libraries, in design of the study and revision of the manuscript. FB performed bioinformatic analysis of cDNA libraries sequence data, clustering of ESTs and annotation of ESTs and identification of microsatellite containing ESTs. BDN and AP participated in development of cDNA libraries production method and identification of microsatellite containing ESTs. GL supervised the study, participating in the design and coordination of the work, the interpretation of the results and revision of the manuscript. RC supervised the study, participating in the design and coordination of the work, the interpretation of data and manuscript writing. All Authors read and approved the final version of the manuscript declaring that they have no potential conflicts of interests.

Additional material

Additional file 1
This table lists 309 non-redundant sequences identifying known E. superba *genes or sequences showing significant similarity with genes from arthropods and other species. These transcripts have been grouped into 13 different functional categories.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-9-45-S1.PDF]

Additional file 2

Quantitative RT-PCR validation. Forward and reverse primer pairs for ten tested transcripts are detailed at the top. Relative expression values resulting from the quantitative real-time PCR are in the middle and related data plots are also reported. For each transcriptional profiling we have associated the EST counting obtained from the cDNA libraries systematic sequencing.

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-9-45-S2.PDF]

References

1. Baker ADC, Boden BP, Brinton E: **A practical guide to Euphausiids of the world.** London: Natural history Museum Pub; 1990.
2. Jarman SN, Nicol S, Elliot NG, McMinn A: **28S rDNA evolution in the Eumalacostraca and the phylogenetic position of krill.** *Mol Phylogenet Evol* 2000, **17**:26-36.
3. Miller DGM, Hampton I: **Biology and ecology of the Antarctic krill.** *Biomass Scientific Series* 1989, **9**:1-166.
4. Nicol S, Endo Y: **Krill fisheries: Development, management and ecosystem implications.** *Aquat Living Resour* 1999, **12**:105-120.
5. Chen YC, Jaczynski J: **Gelation of protein recovered from whole Antarctic krill (*Euphausia superba*) by isoelectric solubilisation/precipitation as affected by functional additives.** *J Agric Food Chem* 2007, **55**:1814-1822.
6. Hewitt RP, Linen Low EH: **The fishery on Antarctic krill: defining an ecosystem approach to management.** *Rev Fish Sci* 2000, **8**:235-298.
7. Taki K, Hayashi T, Naganobu M: **Characteristics of seasonal variation in diurnal vertical migration and aggregation of Antarctic krill (*Euphausia superba*) in the Scotia Sea, using Japanese fishery data.** *CCAMLR Sci* 2005, **12**:.
8. Mauchline J: **The biology of mysids and euphausiids.** *Adv Mar Biol* 1980, **18**:1-681.
9. Godlewska M: **Vertical migrations of krill (*Euphausia superba*, Dana).** *Pol Arch Hydrobiol* 1996, **43**:9-63.
10. Machida RJ, Miya MU, Yamauchi MM, Nishida M, Nishida S: **Organization of the mitochondrial genome of Antarctic krill *Euphausia superba* (Crustacea: Malacostraca).** *Mar Biotechnol* 2004, **6**:238-50.
11. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: **Serial analysis of gene expression.** *Science* 1995, **270**:484-487.
12. Liang P, Pardee AB: **Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction.** *Science* 1992, **257**:967-971.
13. Adams MD, Kerlavage AR, Fields C, Venter JC: **3400 new expressed sequence tags identify diversity of transcripts in human brain.** *Nat Genet* 1993, **4**:256-267.
14. Tassanakajon A, Klinbunga S, Paunglarp N, Rimphanitchayakit V, Udomkit A, Jitrapakdee S, Sritunyalucksana K, Phongdara A, Pongsomboon S, Supungul P, Tang S, Kuphanumart K, Pichyangkura R, Lursinsap C: ***Penaeus monodon* gene discovery project: the generation of an EST collection and establishment of a database.** *Gene* 2006, **384**:104-112.
15. Okubo K, Hori N, Matoba R, Nuyama T, Fukushima A, Kojima Y, Matsubara K: **Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression.** *Nat Genet* 1992, **2**:173-179.
16. *E. superba* database [http://krill.cribi.unipd.it]
17. Pacchioni B, Trevisan S, Gomirato S, Toppo S, Valle G, Lanfranchi G: **Semi-multiplex PCR technique for screening of abundant transcripts during systematic sequencing of cDNA libraries.** *Biotechniques* 1996, **21**:644-646.
18. Bard JBL, Rhee SY: **Ontologies in biology: design, applications and future challenges.** *Nat Rev Genet* 2004, **5**:213-222.
19. Rapp BA, Wheeler DL: **Bioinformatics resources from the national center for biotechnology information: an integrated foundation for discovery.** *J Am Inform Sci Technol* 2005, **56**:538-550.
20. Toikkanen JH, Miller KJ, Soderlund H, Jantti J, Keranen S: **The beta subunit of the Sec61p endoplasmic reticulum translocon interacts with the exocyst complex in *Saccharomyces cerevisiae*.** *J Biol Chem* 2003, **278**:20946-20953.
21. Beltramini M, Colangelo N, Giomi F, Bubacco L, Di Muro P, Hellmann N, Jaenicke E, Decker H: **Quaternary structure and functional properties of *Penaeus monodon* hemocyanin.** *FEBS J* 2005, **272**:2060-2075.
22. Golstein P, Kroemer G: **Cell death by necrosis: towards a molecular definition.** *Trends Biochem Sci* 2007, **32**:37-43.
23. Zhang X, Huang C, Qin Q: **Antiviral properties of hemocyanin isolated from shrimp *Penaeus monodon*.** *Antivir Res* 2004, **61**:93-99.
24. Destoumieux-Garzon D, Saulnier D, Garnier J, Jouffrey C, Bulet P, Bachere E: **Crustacean immunity. Antifungal peptides are generated from the C terminus of shrimp hemocyanin in response to microbial challenge.** *J Biol Chem* 2001, **276**:47070-47077.
25. Ju Z, Karsi A, Kocabas A, Patterson A, Li P, Cao D, Dunham R, Liu Z: **Transcriptome analysis of channel catfish (*Ictalurus punctatus*): genes and expression profile from the brain.** *Gene* 2000, **261**:373-382.
26. Cao D, Kocabas A, Ju Z, Karsi A, Li P, Patterson A, Liu Z: **Transcriptome of channel catfish (*Ictalurus punctatus*): initial analysis of genes and expression profiles of the head kidney.** *Animal Genet* 2001, **32**:169-188.
27. Venier P, Pallavicini A, De Nardi B, Lanfranchi G: **Towards a catalogue of genes transcribed in multiple tissues of *Mytilus galloprovincialis*.** *Gene* 2003, **314**:29-40.
28. Eirin-Lopez JM, Gonzalez-Tizon AM, Martinez A, Mendez J: **Birth-and-death evolution with strong purifying selection in the histone H1 multigene family and the origin of orphon H1 genes.** *Mol Biol Evol* 2004, **21**:1992-2003.
29. Hankeln T, Schmidt ER: **The organization, localization and nucleotide sequence of the histone genes of the midge *Chironomus thummi*.** *Chromosoma* 1991, **101**:25-31.
30. Akhmanova A, Miedema K, Kremer H, Henning W: **Two types of polyadenilated mRNAs are synthesized from *Drosophila* replication-dependent histones.** *Eur J Biochem* 1997, **244**:294-300.
31. del Gaudio R, Potenza N, Stefanoni P, Chiusano ML, Geraci G: **Organization and nucleotide sequence of the cluster of five histone genes in the polychaete worm *Chaetopterus variopedatus*: First record of a H1 histone gene in the phylum Annelida.** *J Mol Evol* 1998, **46**:64-73.
32. Eirýn-Lopez JM, Gonzalez-Tizon AM, Martýnez A, Mendez J: **Molecular and evolutionary analysis of mussel histone genes (*Mytilus spp.*): Possible evidence of an "orphon" origin for H1 histone genes.** *J Mol Evol* 2002, **55**:272-283.
33. Barzotti R, Pelliccia F, Bucciarelli E, Rocchi A: **Organization, nucleotide sequence, and chromosomal mapping of a tandemly repeated unit containing the four core histones genes and a 5S rRNA gene in an isopod crustacean species.** *Genome* 2000, **43**:341-345.
34. Marino-Ramirez L, Jordan IK, Landsman D: **Multiple independent evolutionary solutions to core histone gene regulation.** *Genome Biol* 2006, **7**:R122.
35. Bortoluzzi S, Danieli GA: **Towards an in silico analysis of transcription patterns.** *Trends Genet* 1999, **15**:118-119.
36. Porter ML, Cronin TW, McClellan DA, Crandall KA: **Molecular characterization of crustacean visual pigments and the evolution of pancrustacean opsins.** *Mol Biol Evol* 2007, **24**:253-268.

37. Ostellari L, Zane L, Maccatrozzo L, Bargelloni L, Patarnello T: **Novel microsatellite loci isolated from the northern krill, *Meganyctiphanes norvegica* (Crustacea, Euphausiacea).** *Mol Ecol* 2000, **9:**377-8.

38. Rudd S: **Expressed sequence tags: alternative or complement to whole genome sequences?** *Trends Plant Sci* 2003, **7:**321-329.

39. Jackson RB, Linder CR, Lynch M, Purugganan M, Somerville S, Thayer SS: **Linking molecular insight and ecological research.** *Trends Ecol Evol* 2002, **17:**409-414.

40. Pallavicini A, Del Mar Costa M, Gestal C, Dreos R, Figueras A, Venier P, Novoa B: **High sequence variability of myticin transcripts in hemocytes of immune-stimulated mussels suggests ancient host-pathogen interactions.** *Dev Comp Immunol* 2008, **32:**213-26.

41. Parkinson J, Anthony A, Wasmuth J, Schmid J, Hedley A, Blaxter M: **PartiGene – constructing partial genomes.** *Bioinformatics* 2004, **20:**1398-1404.

42. Parkinson J, Guiliano DB, Blaxter M: **Making sense of EST sequences by CLOBBing them.** *BMC Bioinformatics* 2002, **3:**31.

43. **Phrap software** [http://www.phrap.org]

44. **The National Center for Biotechnology Information** [ftp://ftp.ncbi.nih.gov/blast/db]

45. **UniProtKB database** [http://www.pir.uniprot.org/database/download.shtml]

46. **MISA software** [http://pgrc.ipk-gatersleben.de/misa/]

47. Rozen S, Skaletsky H: **Primer3 on the WWW for general users and for biologist programmers.** *Methods Mol Biol* 2000, **132:**365-86.

48. Schuelke M: **An economic method for the fluorescent labeling of PCR fragments.** *Nature Biotechnology* 2000, **18:**233-234.

49. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using Real-Time Quantitative PCR and the 2$^{-\Delta\Delta Ct}$ method.** *Methods* 2001, **25:**402-408.

50. Cook CE, Yue Q, Akam M: **Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic.** *Proc Biol Sci* 2005, **272:**1295-1304.

51. Crease TJ: **The complete sequence of the mitochondrial genome of *Daphnia pulex* (Cladocera: Crustacea).** *Gene* 1999, **233:**89-99.

52. Umetsu K, Iwabuchi N, Yuasa I, Saitou N, Clark PF, Boxshall G, Osawa M, Igarashi K: **Complete mitochondrial DNA sequence of a tadpole shrimp (*Triops cancriformis*) and analysis of museum samples.** *Electrophoresis* 2002, **23:**4080-4084.

53. Lavrov DV, Brown WM, Boore JL: **Phylogenetic position of the Pentastomida and (pan)crustacean relationships.** *Proc Biol Sci* 2004, **271:**537-544.

54. Ivey JL, Santos SR: **The complete mitochondrial genome of the Hawaiian anchialine shrimp *Halocaridina rubra* Holthuis, 1963 (Crustacea: Decapoda: Atyidae).** *Gene* 2007, **394:**35-44.

55. Segawa RD, Aotsuka T: **The mitochondrial genome of the Japanese freshwater crab, *Geothelphusa dehaani* (Crustacea: Brachyura): evidence for its evolution via gene duplication.** *Gene* 2005, **355:**28-39.

56. Kilpert F, Podsiadlowski L: **The complete mitochondrial genome of the common sea slater, *Ligia oceanica* (Crustacea, Isopoda) bears a novel gene order and unusual control region features.** *BMC Genomics* 2006, **7:**241.

57. Miller AD, Austin CM: **The complete mitochondrial genome of the mantid shrimp *Harpiosquilla harpax*, and a phylogenetic investigation of the Decapoda using mitochondrial sequences.** *Mol Phylogenet Evol* 2006, **38:**565-574.

58. Miller AD, Murphy NP, Burridge CP, Austin CM: **Complete mitochondrial DNA sequences of the decapod crustaceans *Pseudocarcinus gigas* (Menippidae) and *Macrobrachium rosenbergii* (Palaemonidae).** *Mar Biotechnol (NY)* 2005, **7:**339-349.

59. Yamauchi MM, Miya MU, Machida RJ, Nishida M: **PCR-based approach for sequencing mitochondrial genomes of decapod crustaceans, with a practical example from kuruma prawn (*Marsupenaeus japonicus*).** *Mar Biotechnol (NY)* 2004, **6:**419-429.

60. Sun H, Zhou K, Song D: **Mitochondrial genome of the Chinese mitten crab *Eriocheir japonica* sinensis (Brachyura: Thoracotremata: Grapsoidea) reveals a novel gene order and two target regions of gene rearrangements.** *Gene* 2005, **349:**207-217.

61. Miller AD, Nguyen TT, Burridge CP, Austin CM: **Complete mitochondrial DNA sequence of the Australian freshwater crayfish, *Cherax destructor* (Crustacea: Decapoda: Parastacidae): a novel gene order revealed.** *Gene* 2004, **331:**65-72.

62. Yamauchi M, Miya M, Nishida M: **Complete mitochondrial DNA sequence of the Japanese spiny lobster, *Panulirus japonicus* (Crustacea: Decapoda).** *Gene* 2002, **295:**89-96.

63. Yamauchi M, Miya M, Nishida M: **Complete mitochondrial DNA sequence of the swimming crab, *Portunus trituberculatus* (Crustacea: Decapoda: Brachyura).** *Gene* 2003, **311:**129-135.

64. Hickerson MJ, Cunningham CW: **Dramatic mitochondrial gene rearrangements in the hermit crab *Pagurus longicarpus* (Crustacea, Anomura).** *Mol Biol Evol* 2000, **17:**639-644.

65. Machida RJ, Miya MU, Nishida M, Nishida S: **Complete mitochondrial DNA sequence of *Tigriopus japonicus* (Crustacea: Copepoda).** *Mar Biotechnol (NY)* 2002, **4:**406-417.

66. Tjensvoll K, Hodneland K, Nilsen F, Nylund A: **Genetic characterization of the mitochondrial DNA from *Lepeophtheirus salmonis* (Crustacea; Copepoda). A new gene organization revealed.** *Gene* 2005, **353:**218-230.

67. Ogoh K, Ohmiya Y: **Complete mitochondrial DNA sequence of the sea-firefly, *Vargula hilgendorfii* (Crustacea, Ostracoda) with duplicate control regions.** *Gene* 2004, **327:**131-139.