# The most comprehensive annotation of the Krill transcriptome provides new insights for the study of physiological processes and environmental adaptation

Ilenia Urso[1], Alberto Biscontin[1], Davide Corso[1], Cristiano Bertolucci[2,3], Chiara Romualdi[1], Cristiano De Pittà[1], Bettina Meyer[4,5,6]*, Gabriele Sales[1]*

[1]University of Padova, Department of Biology Via U. Bassi 58/B, 35131 Padova (PD) – Italy

[2]University of Ferrara, Department of Life Sciences and Biotechnology Via Luigi Borsari 46 44121 Ferrara (FE) – Italy

[3]Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn Napoli, Villa Comunale, 80121 Naples, Italy

[4]Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research Am Handelshafen 12 D-27570 Bremerhaven – Germany

[5]Institute for Chemistry and Biology of the Marine Environment (ICBM) Carl-von-Ossietzky University Carl-von-Ossietzky-Straße 9-11 D-26111 Oldenburg – Germany

[6]Helmholtz Institute for Functional Marine Biodiversity (HIFMB) at the University of Oldenburg, Oldenburg, 26111, Germany.

*** Co-corresponding authors**

## Abstract

The krill species *Euphausia superba* plays a critical role in the food chain of the Antarctic ecosystem, as the abundance of its biomass affects trophic levels both below it and above. Major changes in climate conditions observed in the Antarctic Peninsula region in the last decades have the potential to alter the distribution of the krill population and its reproductive dynamics. A deeper understanding of the adaptation capabilities of this species, and of the molecular mechanisms behind them, are urgently needed. The availability of a large body of RNA-seq assays gave us the

opportunity to extend the current knowledge of the krill transcriptome, considerably reducing errors and redundancies. The study covered the entire developmental process, from larval stages to adult individuals, information which are of central relevance for ecological studies. We describe KrillDB$^2$ database, which combines the latest annotation of the krill transcriptome with a series of analyses specifically targeting genes and molecular processes relevant to krill physiology. KrillDB$^2$ provides in a single resource the most complete collection of experimental data and bioinformatic annotations: it includes an extended catalog of krill genes; an atlas of their expression profiles over all RNA-seq datasets publicly available; a study of differential expression across multiple conditions such as developmental stages, geographical regions, seasons, and sexes. Finally, it provides information about non-coding RNAs, a class of molecules whose contribute to krill physiology, which have never been reported before.

# Introduction

Antarctic krill *Euphausia superba* (hereafter krill) represents a widely distributed crustacean of the Southern Ocean and one of the world's most abundant species with a total biomass between 100 and 500 million tonnes (Nicol & Endo, 1997). Due to its crucial ecological role in the Antarctic ecosystem, where it represents a link between apex predators and primary producers, several studies have been carried out over the years in order to characterize krill distribution (Atkinson et al. 2008; Hofmann & Murphy 2004; Marr 1962; Siegel 2005), population dynamics and structuring (Bortolotto, Bucklin, Mezzavilla, Zane & Patarnello, 2011; Valentine & Ayala 1976) and above all to understand its complex genetics (Batta-Lona, Bucklin, Wiebe, Patarnello, & Copley, 2011; Bortolotto et al. 2011; Goodall-Copestake et al. 2010; Zane et al. 1998). A

52 sizable fraction of these studies focused on the DNA, specifically on mtDNA variation;
53 however, the information available about krill genetics remains relatively modest. The
54 difficulty in the progression of this kind of study mainly depends on the extraordinarily
55 large krill genome size (Jeffery, 2011), which is more than 15 times larger than the
56 human genome. This aspect largely complicates DNA sequencing, which is the reason
57 why in recent years - together with the advances in high-throughput RNA-sequencing
58 techniques - different krill transcriptome resources have been developed (Clark et al.
59 2011; De Pittà et al., 2008; De Pittà et al. 2013; Martins et al., 2015; Meyer et al., 2015;
60 Seear et al., 2010). However, it was with the KrillDB project (Sales et al., 2017) that a
61 detailed and advanced genetic resource was produced and made available to the
62 community as an organized database. KrillDB is a web-graphical interface with
63 annotation results coming from the *de novo* reconstruction of krill transcriptome,
64 assembling more than 360 million Illumina sequence reads.

65 Here we describe an update to KrillDB, now renamed KrillDB$^2$ (available at the address
66 https://krilldb2.bio.unipd.it/), specifically focusing on two aspects: the improvement of
67 the quality and breadth of the krill transcriptome sequences previously reconstructed,
68 thanks to the addition of an unprecedented amount of RNA-sequencing data; and,
69 correspondingly, an increase in the amount of annotation information associated to
70 each transcript, which is made available through interactive graphs, images and
71 downloadable files.

72

73

# Material and methods

## Krill collection

This study aims at covering the entire developmental process of krill. Therefore, we used samples coming from different developmental stages to cover the entire *E. superba* transcriptome, from larval to adult specimens. Specifically, adults included both male and female specimens, as well as summer and winter individuals and they also came from 3 different geographical regions: Lazarev Sea, South Georgia, and Bransfield Strait/South Orkney.

## Transcriptome assembly strategy

### Multiple independent de novo assemblies

The assembly of short (Illumina) reads to reconstruct the transcriptomes of non-model organisms has been subject to a considerable amount of research. Out of the many tools developed for this task, we have selected five which are arguably the most popular in the field: Trinity (Grabherr et al. 2011), BinPacker (Liu et al. 2016), rnaSPAdes (Bushmanova, Antipov, Lapidus & Prjibelski, 2019), TransABySS (Zhao et al. 2011) and IDBA-tran (Peng et al. 2013). We summarized all the steps of the assembly reconstruction strategy, annotation process and downstream analyses in **Figure 1a**.

At first, we performed a separate transcriptome reconstruction with each of the tools listed above. We evaluated their respective advantages through a series of independent measures, such as: the total number of transcripts; %GC content; the average fragment length; the total number of bases; the N50 value; and finally, the

results of the BUSCO analysis, which provides a measure of transcriptome completeness based on evolutionarily informed expectations of gene content from near-universal single-copy orthologs (Simão, Waterhouse, Ioannidis, Kriventseva & Zdobnov, 2015).

## Assembly filtering and optimization

The raw sequencing data we used for the assemblies was obtained from different experiments and included both stranded and unstranded libraries. As mixing these two types of libraries in a single assembly is not well supported, we decided to run each software twice: we thus generated a total of ten different *de novo* assemblies.

A combination of two filtering steps was then applied to the newly reconstructed transcriptomes in order to discard artifacts and improve the assembly quality.

First, we estimated the abundances of all the transcripts reconstructed by each assembler using the Salmon software (v. 1.4.0, Patro, Duggal, Love, Irizarry & Kingsford, 2017). Samples were grouped according to the main experimental conditions: (1) sex, with female and male levels; (2) geographical area, covering Bransfield Strait, South Georgia, South Orkney and Lazarev Sea; and (3) season, with summer and winter levels. Abundance estimates were imported in the R statistical software using the *tximport* package (Love et al. 2017) and we implemented a filter to keep only those transcripts showing an expression level of at least 1 transcript per million (TPM) within each of the three experimental conditions.

In a second step, we considered the results of all assemblers jointly, and we ran the “cd-hit-est” program (Li & Godzik, 2006) in order to cluster similar sequences and to produce a set of non-redundant representative transcripts. Specifically, we collapsed

119 all sequences sharing 95% or more of their content, thus reducing the number of
120 transcripts from 1,650,404 to 551,110.

## 121 Meta-assembly

122 The procedure described above was designed to identify near-duplicate sequences
123 deriving from different software, but likely corresponding to the same biological
124 transcript. As a further refinement, we were also interested in grouping resulting
125 transcripts into units corresponding to genes. To this end, we relied on the
126 EvidentialGene pipeline (Gilbert 2013, 2019). We applied the "tr2aacds" tool which
127 clusters transcripts and classifies them to identify the most likely coding sequence
128 representing each gene. The software subdivides sequences into different categories,
129 including primary transcript with alternates (main), primary without alternates
130 (noclass), alternates with high and medium alignment to primary (althi1, althi, altmid)
131 and partial (part) incomplete transcripts. A "coding potential" flag is also added,
132 separating coding from non-coding sequences (see section "KrillDB$^2$ Web Interface").
133 The meta-assembly thus obtained consisted in 274,840 putative transcripts,
134 subdivided into 173,549 genes.

135 As these figures remained unrealistically high, we performed another round of
136 analyses to identify redundant or mis-assembled sequences still appearing in our
137 transcriptome. Here we used a combination of BLAST searches against known protein
138 and nucleotide databases (NR, NT, TREMBL) and information deriving from full-
139 length, experimentally validated transcripts from a previous study (Biscontin et al.,
140 2017). Results confirmed that the newly reconstructed transcriptome fully represented
141 krill RNAs, but the large amount of input reads, together with the number of
142 independent de novo assemblers, likely led to an inflation in the number of alternative

143 splicing variants being reconstructed. Moreover, transcript alignments against BUSCO
144 genes (Simão et al., 2015) and the *doubletime*, *cry1, shaggy* and *vrille* full-length
145 transcripts from Biscontin *et al*. (2017) highlighted the fact that multiple fragments of
146 the same gene were incorrectly assembled as separate transfrags. To remove these
147 artifacts, first we aligned all transcript sequences in our dataset against each other
148 using the *blastn* tool. We discarded all sequences already included in a longer
149 transcript for more than the 90% of their length. This filter helped us remove 78,731
150 redundant sequences (29% of transcripts, overall). Then, we ran a new abundance
151 quantification using Salmon and we discarded all transcripts with an average
152 abundance below 0.1.

153 The combination of all the filters discussed above allowed us to reduce the number of
154 transcripts to 151,585 and, correspondingly, that of genes to 85,905. Our approach
155 retained alternative and paralog transcripts with sufficient level of uniqueness in their
156 sequence, as confirmed by the fact that although we removed almost 45% of the
157 initially assembled transcripts, the drop barely affected the medium mapping rate
158 value, reducing it from 89% (EvidentialGene only) to 88% (full filtering).

159 In order to enhance the interpretability of the transcriptome reconstruction, we also
160 performed a SuperTranscripts analysis, on the basis of the workflow proposed by
161 Davidson, Hawkins & Oshlack (2017). Starting from the clustering information (groups
162 of transcripts belonging to different genes) provided by EvidentialGene, we ran the
163 Lace software (https://github.com/Oshlack/Lace) to get a representation of the block
164 structure of each gene (see section "KrillDB$^2$ Web Interface").

165

166 ## Expression Atlas

167 We estimated the abundances of all the transcripts in our final assembly using Salmon
168 over a wide range of RNA-seq datasets, including:

169 - Larval krill at two different stages of development exposed to different $CO_2$
170 conditions, coming from Sales et al. (2017)
171 - Adult krill (48 samples) coming from different geographical areas (Bransfield
172 Strait, Lazarev Sea, South Georgia, South Orkney) and different seasons
173 (summer and winter), divided into male and female specimens (Höring et al.,
174 2020)
175 - Adult krill divided into male and female specimens (Suter et al., 2019)
176 - Adult krill exposed to three different temperatures – Low Temperature, Mid
177 temperature, High Temperature (data downloaded from NCBI BioProject:
178 PRJNA640244)

179 Overall, the experimental design included six factors: geographical area, season,
180 developmental stage, $pCO_2$ exposure condition, sex and temperature. Abundances
181 and raw counts were imported using R (v. 4.0.5) and the package *tximport* (v. 1.18.0).
182 Batch effect removal was performed using the *removeBatchEffect* function
183 implemented in the *limma* package (v. 3.46.0). The resulting count matrix of transcripts
184 (rows) across samples (columns) was then converted to the transcripts per million
185 (TPM) scale. Finally, results were summarized to the gene level using the
186 *isoformToGeneExp* function (IsoformSwitchAnalyzeR v. 1.12.0 package). The
187 expression levels for each experimental condition are displayed in KrillDB[2] as a
188 barplot, as part of the webpage for each gene or transcript (see section "KrillDB[2] Web
189 Interface").

## Functional Annotation

Assembled fragments were aligned against the NCBI NR (non-redundant) UniProtKB/TrEMBL protein databases and also against the NCBI NT nucleotide collection (data downloaded on 22/04/2021). We also ran InterproScan (v. 5.51-85.0) in order to search for known functional domains and predict protein family membership.

In addition, all krill transcripts were searched against the RNAcentral database (https://rnacentral.org/; https://doi.org/10.1093/nar/gkw1008) in order to identify and characterize non-coding RNAs within the reconstructed sequences. Contextually, an orthology inference was performed using the OMA standalone package (https://omabrowser.org/standalone/; Altenhoff et al., 2019) which relies on a complete catalog of orthologous genes among more than 2,300 genomes covering the entire tree of life. This analysis helped us identify, based on protein sequences, those krill transcripts showing an orthology relationship with genes from other species and which sets of genes derived from a single common ancestral gene at a given taxonomic range (Altenhoff, Gil, Gonnet & Dessimoz, 2013).

## Differential Expression Analysis

Transcript-level abundances and estimated counts were summarized at the gene-level using the package *tximport*; resulting counts were normalized to remove unwanted variation by means of the RUVg method (Risso & Course, 2015). Specifically, we performed a preliminary between-sample normalization (EDASeq package, v. 2.24.0) to adjust for sequencing depth. We defined a set of empirical negative control genes, by means of a differential expression (DE) analysis with the edgeR software (v.

213 3.32.1). Our design matrix for the model included all the independent factors (*season*,
214 *area* and *sex*) and, in addition, the interaction between *area* and *season*, *sex* and *area*,
215 *sex* and *season*. We selected the 30,000 less differentially expressed genes as
216 negative controls and we applied the RUVg method. We finally re-ran the differential
217 expression analysis extending the original model to include two estimated factors to
218 capture unwanted variation.

## 219 Non-coding RNAs

220 We also investigated the possibility that the new transcriptome included sequences
221 corresponding to non-coding transcripts, specifically microRNAs.

222 To this aim, we ran the HHMMiR software (Kadri, Hinman & Benos, 2009), which
223 combines structural and sequence information to train a Hierarchical Hidden Markov
224 Model for the identification of microRNA genes. We also performed a *blastn* search of
225 all our assembled transcripts against the collection of miRBase
226 (http://www.mirbase.org/) mature sequences. Results from these two analyses were
227 combined: we collected all transcripts with a HHMMiR score below or equal to 0.71
228 and an alignment to a known mature microRNA with at most two mismatches. We then
229 used the QuickGO tool (https://www.ebi.ac.uk/QuickGO/) to identify any potential
230 association among our putatively identified microRNAs and GO categories.

## 231 Opsin phylogeny

232 To identify novel opsin genes in krill, we examined manually the list of transcripts that
233 were annotated as “opsin” by our automated pipeline. Furthermore, the entire krill
234 transcriptome was aligned against a curated opsin dataset (including 996 visual and
235 non-visual opsins, Henze & Oakley, 2015) using Blast+ (version 2.11.0). The longest

236 transcript was selected as a representative sequence of those genes that possess
237 many alternative transcripts. Secondary structure was assessed by NCBI Conserved
238 Domain Search (CDD database, May 2021). Phylogenetic tree was generated with
239 MUSCLE alignment tool and Maximum Likelihood method (Dayhof substitution matrix
240 and Nearest-Neighbor-Interchange method) as implemented in MEGA X version
241 10.2.6 (https://www.megasoftware.net/). New opsins were aligned against a curated
242 invertebrate-only opsin data set (DeLeo & Bracken-Grissom, 2020), the previously
243 cloned krill opsins (Biscontin et al., 2016), and the full-length onychopsin and
244 arthropsin sequences available on NCBI Protein database (May 2021,
245 ncbi.nlm.nih.gov/protein). The tree was rooted using the human G protein-coupled
246 receptor VIPR1 as an outgroup.

247

# 248 Results

## 249 Transcriptome Quality

250 We checked for transcriptome reconstruction quality step by step, starting from the 5
251 independent de novo assemblies, going through the first filter applied (**Figure 1b**),
252 then testing the potential of merging all the assemblies in a unique meta-assembly
253 through EvidentialGene and finally filtering the transcriptome for redundancy
254 reduction. As previously mentioned, the assembly quality was evaluated using
255 different measures. **Figure 1c** reports result of this approach in terms of contig N50
256 statistics, which highlighted an increase of the assembly quality at each step. Recent
257 benchmarks, such as Hölzer & Marz (2019), have shown that, while reconstructing the
258 transcriptome of a species, no single approach is uniformly superior at reconstructing

259 the transcriptome of a species: the quality of each result is influenced by a number of
260 factors, both technical (*k*-mer size, strategy for duplicate resolution) and biological
261 (genome size, presence of contaminants). In our study we observed that, although a
262 consistent number of sequences was removed through each step of assembly,
263 merging and filtering, we didn't find any decline in the quality described by the basic
264 statistics of the reconstructed transcripts.

265 We then explored the completeness of the krill transcriptome according to conserved
266 ortholog content using BUSCO. **Figure 1d** shows the results of the analyses
267 comparing the EvidentialGene reconstructed transcriptome to the redundancy-filtered
268 assembly, where we observed a minimal reduction of complete matches to the
269 BUSCO profile.

## 270 Functional Classification

271 Results from the functional annotation analyses showed that 63,633 contigs matched
272 at least one protein from the NR collection, corresponding to about 42s% of the total
273 krill transcriptome, while 62,249 contigs found a match among TREMBL protein
274 sequences (41% of the total). A search against the NT nucleotide sequences produced
275 22,071 krill matches (15% of the total). To classify transcripts by putative function, we
276 performed a GO assignment. Specifically, 2,612 GO terms were assigned, with 1,128
277 of those terms representing multiple molecular functions, 1,099 corresponding to
278 multiple biological processes and 385 to cellular components.

## 279 A case study on the discovery of opsin genes

280 To evaluate the gene discovery potential of the new version of assembly, we searched
281 transcriptome for new members of the opsin family. Opsins are a group of light

282 sensitive G protein-coupled receptors with seven transmembrane domains. 14 genes
283 were annotated as putative opsins and the conserved domains analysis revealed that
284 all of them possess the distinctive 7 α-helix transmembrane domains structure. The 8
285 previously cloned opsins (Biscontin et al. 2016) were represented by as many genes
286 (sequence identity >90%; **Table S1**). The other 6 identified genes can therefore be
287 considered new putative opsins. Among these, we found 4 putative rhabdomeric
288 opsins: *Es*Rh7 and *Es*Rh8, with 70% and 59 % of amino acid identity to *Es*Rh1a and
289 *Es*Rh4, respectively; *Es*Rh9 and *Es*Rh10 showing high sequence identity (87% and
290 74%, respectively) to *Es*Rh5. Further, we identified 2 putative ancestral opsins: a non-
291 visual arthropsin (*Es*Arthropsin), and an onychopsin (*Es*Onychopsin) with 70% and
292 49% of sequence identity with crustacean and onychophoran orthologous,
293 respectively. Phylogenetic analysis (**Figure 2**) suggested that *Es*Rh7-10 are middle-
294 wavelength-sensitive (MWS) rhabdomeric opsins, and further confirmed *Es*Arthropsin
295 and *Es*Onychopsin annotation.

## 296 Differential Expression

297 The availability of a new assembly of the krill transcriptome, reconstructed collecting
298 the largest amount of experimental data available thus far, suggested the possibility
299 of performing a more detailed investigation of differential expression patterns. In
300 particular, we decided to reanalyze the dataset from (Höring et al., 2020) in order to
301 assess the possibility of identifying differentially expressed genes which were not
302 detected in the original study due to the use of an older reference (Meyer et al., 2015).

303 In total 1,741 genes were differentially expressed (DEG) among experimental
304 conditions, corresponding to around 2% of the total reconstructed genes. **Table 2**

305 summarizes the list of contrasts that were performed, each one with the number of
306 differentially expressed up and down regulated genes.
307 1,195 DEGs were identified in the comparison between summer and winter
308 specimens: 1,078 were up-regulated and 117 down-regulated. 396 of such DEGs had
309 some form of functional annotation. Comparison between males and females
310 produced 14 DEGs (7 up-regulated and 7 down-regulated); none had some form of
311 annotation deriving from sequence homology. In general, these results are in
312 accordance with the discussion by Höring et al. (2020), which found that seasonal
313 differences are predominant in comparison to regional ones.

## 314 Summer vs Winter

315 We selected a series of genes among seasonal DEGs according to what has been
316 already described in literature. Höring et al. (2020) previously identified and described
317 35 relevant DEGs involved in seasonal physiology and behavior: we recovered the
318 same gene signature in our own analysis by comparing summer to winter samples.
319 The majority of these DEGs appear to be involved in the development of cuticles
320 (*chitine synthase*, *carbohydrate sulfotransferase 11*), lipid metabolism (*fatty acid*
321 *synthase 2*, *enoyl-CoA ligase*), reproduction (*vitellogenin*, *hematopoietic prostaglandin*
322 *D synthase*), metabolism of different hormones (*type 1 iodothyronine deiodinase*) and
323 in the circadian clock (*cryptochrome*). Our results also include DEGs that were found
324 to be involved in the moult cycle of krill in other studies (Seear et al., 2010).
325 Specifically, we identified a larger group of genes involved in the different stages of
326 cuticle developmental process (*peritrophin-A domain*, *calcified cuticle protein*,
327 *glycosyltransferase 9-domain containing protein*, *collagen alpha 1*, *glutamine-fructose*
328 *6 phosphate*), including proteins such as *cuticle protein-3,6,19.8, early cuticle protein,*
329 *pupal cuticle protein, endocuticle structural glycoprotein, chitinase-3* and *chitinase-4*,

330 the latter representing a group of chitinase which have been shown to be expressed
331 predominantly in gut tissue during larval and/or adult stages in other arthropods and
332 are proposed to be involved in digestion of chitin-containing substrates (Khajuria,
333 Buschman, Chen, Muthukrishnan & Zhu, 2010). Finally, in addition to *trypsin* and
334 *crustin 4* (immune-related gene, essential in early pre-moult stage when krill still have
335 a soft cuticle to protect them from pathogen attack, as seen by Seear et al., 2010) we
336 also identified *crustin-1,2,3,5* and *7*. All the reported genes were up-regulated in
337 summer, the period in which growth take place and krill moult regularly.
338 Cuticle development genes were also identified as differentially expressed in the
339 analysis of the interaction of multiple factors, in particular between male samples
340 coming from South Georgia and female specimens coming from the area of Bransfield
341 Strait-South Orkney (considered as a unique area since they are placed at similar
342 latitudes). Strikingly, we also identified a pro-resilin gene, whose role in many insects
343 consists in providing efficient energy storage, being upregulated in South Georgia
344 male specimens.

## 345 Interaction Effects

346 A number of relevant DEGs were found among specific interactions of regional and
347 seasonal changes. In the comparison between krill samples in South Georgia in
348 summer and individuals sampled in Bransfield Strait-South Orkney in winter we found
349 genes, up-regulated in summer in South Georgia, that are related to reproductive
350 activities, such as *doublesex* and *mab-3 related transcription factor*. The latter, in
351 particular, is a transcription factor crucial for sex determination and sexual
352 differentiation which has already described in other arthropods (Yi Jia et al., 2018)
353 where the male-specific isoform represses its targets. Since no differentially expressed
354 gene related to reproduction was found by Höring et al. (2020) in the same

355 comparisons, this suggests that the new krill transcriptome significantly improves the
356 interpretability of expression studies and the characterizations of krill samples.
357 Finally, comparison between male individuals from Lazarev Sea and female
358 specimens from Bransfield Strait-South Orkney showed additional DEGs involved in
359 reproduction, such as *ovochymase 2*, usually highly expressed in female adults or
360 eggs, *serine protease* and a *trypsin-like gene*. In particular, *trypsin-like genes* are
361 usually thought to be digestive serine proteases, but previous works suggested that
362 they can play other roles (Bao et al., 2014); many trypsins show female or male-
363 specific expression patterns and have been found exclusively expressed in males, as
364 in our analysis, suggesting that they play a role in the reproductive processes.
365 The simultaneous presence of genes involved in different steps of the krill moulting
366 cycle and in the reproductive process and sexual maturation that appear to be
367 differentially expressed in same comparisons is in accordance with what was already
368 observed in krill (Buchholz, Watkins, Priddle, Morris & Ricketts, 1996) and other krill
369 species (Tarling & Cuzin-Roudy, 2003). In particular, there are already evidences of a
370 strong relation between the krill moulting process and its growth and sexual maturation
371 during the year, which support and confirm the reliability of our results in terms of
372 genes involved in such krill life cycle steps.
373 All tables of differentially expressed genes tables are available and downloadable on
374 KrillDB[2] (**Figure 3c**; https://krilldb2.bio.unipd.it/, Section "Differentially Expressed
375 Genes (DEGs)").

## 376 MicroRNA Identification

377 In total we identified 261 krill transcripts with sequence homology to 644 known
378 microRNAs from other species. 306 sequences were linked to at least one GO term,

379 matching 54 krill transcripts (**Table S2**). Among them, we identified 5 putative
380 microRNAs involved with changes in cellular metabolism (age-dependent general
381 metabolic decline - GO:0001321, GO:0001323), as well as changes in the state or
382 activity of cells (age-dependent response to oxidative stress - GO:0001306,
383 GO:0001322, GO:0001324), 35 microRNAs involved in interleukin activity and
384 production regulation. Moreover, we found 26 putative microRNAs likely involved in
385 *ecdysteroidogenesis* (specifically GO:0042768), a process resulting in the production
386 of ecdysteroids, moulting and sex hormones found in many arthropods. In addition,
387 we found a microRNA involved in fused antrum stage (GO:0048165) which appears
388 to be related, in other species, to the oogenesis process. We also identified 27
389 microRNAs related to *rhombomere* morphogenesis, formation and development
390 (GO:0021661, GO:0021663, GO:0021570). These functions are well described as
391 involved in the development of portions of the central nervous system in vertebrates,
392 but they are also homologous to a portion of the arthropod brain. Moreover, 26 krill
393 sequences showed high similarity with 2 mature microRNA related to the formation of
394 tectum (GO:0043676), which represents in arthropods and, specifically, crustaceans,
395 the part of the brain acting as visual center.

## 396 KrillDB$^2$ Web Interface

397 The KrillDB website has been re-designed to include the new version of the
398 transcriptome assembly. **Figures 3**, **4** and **5** collect images taken from the main new
399 sections of the updated version of KrillDB. The integrated full-text search engine
400 allows the user to search for a transcript ID, gene ID, GO term, a microRNA ID or any
401 other free-form query. In particular, results of full-text searches are now organized into
402 a number of separate tables, each representing a different data source or biological

403 aspect (**Figure 3b**). Results of GO term searches are summarized in a table reporting
404 the related genes with corresponding domain (**Figure 4a**) or microRNA (**Figure 4b**)
405 match and associated description. Both gene and transcript-centric pages have been
406 extended with two new sections: “Orthology" and “Expression” (**Figure 5a**). The
407 Orthology section summarizes the list of orthologous sequences coming from OMA
408 analysis, each one with the species it belongs to and the identity score.

409 The “Expression” section shows a barplot, implemented using the Seaborn Python
410 library (v. 0.11.1), representing abundances estimates obtained from Salmon (Patro
411 et al., 2017). An additional section, called “Gene Structure” (**Figure 5a**), was added to
412 the gene page on the basis of the results coming from the SuperTranscript analysis.
413 Specifically, we modified the STViewer.py Python script (from Lace), optimizing and
414 adapting it to our own data and database structure, in order to produce a visualization
415 of each gene with its transcripts. Since Lace relies on the construction of a single
416 directed splice graph and it is not able to compute it for complex clusters with more
417 than 30 splicing variants, this section is available for a selection of genes only.

418 The new KrillDB$^2$ release includes completely updated transcript and gene identifiers.
419 However, the user searching for a retired ID is automatically redirected to the page
420 describing the newest definition of the appropriate transcript or gene.

421 The KrillDB$^2$ homepage now includes two additional sections (**Figure 3a**): one is
422 represented by the possibility to perform a BLAST search. Any nucleotide or protein
423 sequence (*query*) can be aligned against krill sequences stored in the database;
424 results are summarized in a table containing information about the krill transcripts
425 (*target*) that matched with the user’s query, and the e-value corresponding to the
426 alignment. The other new section, called “Differentially Expressed Genes”, allows the

427 user to browse all the tables listing the genes that were found to be differentially
428 expressed among the conditions we have described above. A drop-down menu gives
429 access to the different comparisons; DEG tables (**Figure 3c**) list for each gene its log
430 fold-change, p- and FDR values as estimated by edgeR. Moreover, each gene is linked
431 to a functional description (if available) inferred from sequence homology searches.

432 Information about krill transcripts that showed homology with an annotated microRNA
433 is available in the section "Predicted Hairpin" (**Figure 5b**). It contains a summary table
434 with details about the hairpin length and the similarity score (as estimated by
435 HHMMiR), followed by full listing of all the corresponding mature microRNAs (including
436 links to their miRBase page). In addition, an image displaying the predicted secondary
437 structure of the hairpin is included (computed by the "fornac" visualization software
438 from the ViennaRNA suite).

439

# 440 Conclusions

441 The availability of a large amount of public RNA-seq data capturing krill transcripts has
442 given us the possibility to re-assemble its transcriptome and to significantly extend its
443 annotation. We have now covered the entire developmental process of this species
444 and included in our analysis individuals belonging to different seasons and affected by
445 different environmental conditions. KrillDB[2] provides the most complete source of
446 information about the krill transcriptome and will offer a reliable starting point
447 development of novel ecological studies.

448 The differential expression analysis we have performed highlights the importance of
449 specific processes in the complex krill life cycle and in its adaptation capability to the
450 harsh Antarctic environment.

451 The identification of 6 novel putative opsin sequences, in addition to the eight that were
452 previously cloned, demonstrates a significant improvement in the gene discovery
453 potential of this new version of krill transcriptome. The finding of four novel MWS
454 rhabdomeric opsins, an onychopsin, and a non-visual arthropsin further enrich the
455 opsin repertoire of *E. superba* shedding light on a complex photoreception system able
456 to coordinate the physiological and behavioral responses to the extreme daily (diel
457 vertical migration) and seasonal changes in photoperiod and spectral composition.
458 Arthropsins are rhabdomeric non-visual opsins and its clade is the sister group of the
459 bilaterian rhabdomeric opsins (Hering et al., 2012; Hering and Mayer, 2014). It was
460 first discovered in the crustacean *Daphnia pulex* and subsequently in other arthropods,
461 onychophoran, molluscs, anallids and flatworms (Colbourne et al., 2011; Hering et al.,
462 2012; Eriksson et al., 2013; Hering and Mayer, 2014; Futahashi et al., 2015). Of
463 relevance is the identification of an onychopsin which has been suggested to be the
464 common ancestor of *Panarthropoda* visual opsins (Hering et al., 2012), and possibly
465 sensitive to wavelength from UV to green light (Beckmann et al., 2015). *Es*Onychopsin
466 could represent the short-wavelength sensitive opsin (SWS/UV) which we have long
467 been searching for. Indeed, the absence of a SWS/UV opsin was truly unexpected in
468 an organism that shows daily vertical migration reaching depth beyond the 30 m,
469 where only short wavelength light can penetrate.

470 Finally, KrillDB[2] includes the first evidence of the role of non-coding RNAs in krill.
471 Although this is just a preliminary analysis, the results we have described already hint
472 at a role of microRNAs in defining the adaptive capabilities of this species to the
473 Antarctic environment. This represents a promising starting point for the study of non-
474 coding RNAs in the Antarctic krill and in other species belonging to the same family.

475

# Acknowledgments

The position of Illenia Urso was supported by the Helmholtz Vitual Institute "PolarTime": Biological timing in a changing marine environment - clocks and rhythms in polar pelagic organisms (VH-VI-500), headed by Bettina Meyer. Alberto Biscontin was funded by the “Programma Nazionale di Ricerche in Antartide – PNRA” (grant 2016_00225) and by the Promega Corporation 2019 Real-Time PCR Grant Program. We would also like to acknowledge the CAPRI initiative (Calcolo ad Alte Prestazioni per la Ricerca e l’Innovazione”, University of Padova Strategic Research Infrastructure Grant 2017) for the technical support and the HPC resources we have used for the analyses. Cristiano Bertolucci was supported by the “Programma Nazionale di Ricerche in Antartide – PNRA” (grant 2016_00225) and by the University of Ferrara research grant (FIR2020 and FAR2021).

# References

Altenhoff, A. M., Gil, M., Gonnet, G. H., & Dessimoz, C. (2013). Inferring hierarchical orthologous groups from orthologous gene pairs. PloS one, 8(1), e53786. doi: 10.1371/journal.pone.0053786

Altenhoff, A. M., Levy, J., Zarowiecki, M., Tomiczek, B., Vesztrocy, A. W., Dalquen, D. A., ... & Dessimoz, C. (2019). OMA standalone: orthology inference among public and custom genomes and transcriptomes. Genome research, 29(7), 1152-1163. doi: 10.1101/gr.243212.118

Atkinson, A., Siegel, V., Pakhomov, E. A., Rothery, P., Loeb, V., Ross, R. M., ... & Fleming, A. H. Atkinson, A., Siegel, V., Pakhomov, E. A., Rothery, P., Loeb, V., Ross, R. M., ... & Fleming, A. H. (2008). Oceanic circumpolar habitats of Antarctic krill. Marine Ecology Progress Series, 362, 1-23., 2008. doi: org/10.3354/meps07498

Bao, Y. Y., Qin, X., Yu, B., Chen, L. B., Wang, Z. C., & Zhang, C. X. (2014). Genomic insights into the serine protease gene family and expression profile analysis in the planthopper, Nilaparvata lugens. BMC genomics, 15(1), 1-17. doi: 10.1186/1471-2164-15-507

Batta-Lona, P. G., Bucklin, A., Wiebe, P. H., Patarnello, T., & Copley, N. J. (2011). Population genetic variation of the Southern Ocean krill, Euphausia superba, in the

Western Antarctic Peninsula region based on mitochondrial single nucleotide polymorphisms (SNPs). Deep Sea Research Part II: Topical Studies in Oceanography, 58(13-16), 1652-1661. doi: 10.1016/j.dsr2.2010.11.017

Beckmann, H., Hering, L., Henze, M. J., Kelber, A., Stevenson, P. A., & Mayer, G. (2015). Spectral sensitivity in Onychophora (velvet worms) revealed by electroretinograms, phototactic behaviour and opsin gene expression. Journal of Experimental Biology, 218(6), 915-922. doi: 10.1242/jeb.116780

Biscontin, A., Frigato, E., Sales, G., Mazzotta, G. M., Teschke, M., De Pittà, C., ... & Bertolucci, C. (2016). The opsin repertoire of the Antarctic krill Euphausia superba. Marine genomics, 29, 61-68. doi: 10.1016/j.margen.2016.04.010

Bortolotto, E., Bucklin, A., Mezzavilla, M., Zane, L., & Patarnello, T. (2011). Gone with the currents: lack of genetic differentiation at the circum-continental scale in the Antarctic krill Euphausia superba. BMC genetics, 12(1), 1-18. doi: 10.1186/1471-2156-12-32

Buchholz, F., Watkins, J. L., Priddle, J., Morris, D. J., & Ricketts, C. (1996). Moult in relation to some aspects of reproduction and growth in swarms of Antarctic krill, Euphausia superba. Marine Biology, 127(2), 201-208. doi: 10.1007/BF00942104

Bushmanova, E., Antipov, D., Lapidus, A., & Prjibelski, A. D. (2019). rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. GigaScience, 8(9), giz100. doi: 10.1093/gigascience/giz100

Clark, M. S., Thorne, M. A., Toullec, J. Y., Meng, Y., Peck, L. S., & Moore, S. (2011). Antarctic krill 454 pyrosequencing reveals chaperone and stress transcriptome. PLos one, 6(1), e15919. doi: 10.1371/journal.pone.0015919

Colbourne, J. K., Pfrender, M. E., Gilbert, D., Thomas, W. K., Tucker, A., Oakley, T. H., et al. (2011). The ecoresponsive genome of *Daphnia pulex*. Science 331, 555–561. doi: 10.1126/science.1197761

Davidson, N. M., Hawkins, A. D., & Oshlack, A. (2017). SuperTranscripts: a data driven reference for analysis and visualization of transcriptomes. Genome biology, 18(1), 1-10. doi: 10.5281/zenodo.830594

De Pittà, C., Bertolucci, C., Mazzotta, G. M., Bernante, F., Rizzo, G., De Nardi, B., ... & Costa, R. (2008). Systematic sequencing of mRNA from the Antarctic krill (Euphausia superba) and first tissue specific transcriptional signature. BMC genomics, 9(1), 1-14. doi: 10.1186/1471-2164-9-45

De Pittà, C., Biscontin, A., Albiero, A., Sales, G., Millino, C., Mazzotta, G. M., ... & Costa, R. (2013). The Antarctic krill Euphausia superba shows diurnal cycles of transcription under natural conditions. PLoS One, 8(7), e68652. doi: 10.1371/journal.pone.0068652

DeLeo, D. M., & Bracken-Grissom, H. D. (2020). Illuminating the impact of diel vertical migration on visual gene expression in deep-sea shrimp. Molecular Ecology, 29(18), 3494-3510. doi: 10.1111/mec.15570

Eriksson, B. J., Fredman, D., Steiner, G., and Schmid, A. (2013). Characterisation and localisation of the opsin protein repertoire in the brain and retinas of a spider and an onychophoran. BMC Evol. Biol. 13:186. doi: 10.1186/1471-2148-13-186

550 Futahashi, R., Kawahara-Miki, R., Kinoshita, M., Yoshitake, K., Yajima, S., Arikawa,
551 K., et al. (2015). Extraordinary diversity of visual opsin genes in dragonflies. Proc. Natl.
552 Acad. Sci. U.S.A. 112, E1247–E1256. doi: 10.1073/pnas.1424670112

553 Gilbert, D. G. (2019). Longest protein, longest transcript or most expression, for
554 accurate gene reconstruction of transcriptomes? bioRxiv, 829184. doi:
555 10.1101/829184

556 Goodall-Copestake, W. P., Perez-Espona, S., Clark, M. S., Murphy, E. J., Seear, P.
557 J., & Tarling, G. A. (2010). Swarms of diversity at the gene cox1 in Antarctic krill.
558 Heredity, 104(5), 513-518. doi: 10.1038/hdy.2009.188

559 Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... &
560 Regev, A. (2011). Trinity: reconstructing a full-length transcriptome without a genome
561 from RNA-Seq data. Nature biotechnology, 29(7), 644. doi: 10.1038%2Fnbt.1883

562 Henze, M. J., & Oakley, T. H. (2015). The dynamic evolutionary history of
563 pancrustacean eyes and opsins. Integrative and comparative biology, 55(5), 830-842.
564 doi: 10.1093/icb/icv100

565 Hering, L, Henze, M.J., Kohler, M., Kelber, A., Bleidorn, C., Leschke, M., Nickel, B.,
566 Meyer, M., Kircher, M., Sunnucks, P., Mayer, G. (2012) Opsins in onychophora (velvet
567 worms) suggest a single origin and subsequent diversification of visual pigments in
568 arthropods. Molecular Biology & Evolution, 29(11), 3451-3458. doi:
569 10.1093/molbev/mss148

570 Hering, L., and Mayer, G. (2014). Analysis of the opsin repertoire in the tardigrade
571 Hypsibius dujardini provides insights into the evolution of opsin genes in
572 Panarthropoda. Genome Biol. Evol. 6, 2380–2391. doi: 10.1093/gbe/evu193

573 Hofmann, E. E., & Murphy, E. J. (2004). Advection, krill, and Antarctic marine
574 ecosystems. Antarctic Science, 16(4). doi: 10.1017/s0954102004002275

575 Hölzer, M., & Marz, M. (2019). De novo transcriptome assembly: A comprehensive
576 cross-species comparison of short-read RNA-Seq assemblers. GigaScience, 8(5),
577 giz039. doi: 10.1093/gigascience/giz039

578 Höring, F., Biscontin, A., Harms, L., Sales, G., Reiss, C. S., De Pittà, C., & Meyer, B.
579 (2021). Seasonal gene expression profiling of Antarctic krill in three different latitudinal
580 regions. Marine Genomics, 56, 100806. doi: 10.1016/j.margen.2020.100806

581 Jeffery, N. W. (2012). The first genome size estimates for six species of krill
582 (Malacostraca, Euphausiidae): large genomes at the north and south poles. Polar
583 Biology, 35(6), 959-962. doi: 10.1007/s00300-011-1137-4

584 Jia, L. Y., Chen, L., Keller, L., Wang, J., Xiao, J. H., & Huang, D. W. (2018). Doublesex
585 evolution is correlated with social complexity in ants. Genome biology and evolution,
586 10(12), 3230-3242. doi: 10.1093/gbe/evy250

587 Kadri, S., Hinman, V., & Benos, P. V. (2009). HHMMiR: efficient de novo prediction of
588 microRNAs using hierarchical hidden Markov models. BMC bioinformatics, 10(1), 1-
589 12. doi: 10.1186/1471-2105-10-S1-S35

590 Khajuria, C., Buschman, L. L., Chen, M. S., Muthukrishnan, S., & Zhu, K. Y. (2010). A
591 gut-specific chitinase gene essential for regulation of chitin content of peritrophic

matrix and growth of Ostrinia nubilalis larvae. Insect biochemistry and molecular biology, 40(8), 621-629. doi: 10.1016/j.ibmb.2010.06.003

Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics, 22(13), 1658-1659. doi: 10.1093/bioinformatics/btl158

Liu, J., Li, G., Chang, Z., Yu, T., Liu, B., McMullen, R., ... & Huang, X. (2016). BinPacker: packing-based de novo transcriptome assembly from RNA-seq data. PLoS computational biology, 12(2), e1004772. doi: 10.1371/journal.pcbi.1004772

Martins, M. J. F., Lago-Leston, A., Anjos, A., Duarte, C. M., Agusti, S., Serrão, E. A., & Pearson, G. A. (2015). A transcriptome resource for Antarctic krill (Euphausia superba Dana) exposed to short-term stress. Marine genomics, 23, 45-47. doi: 10.1016/j.margen.2015.04.008

Meyer, B., Martini, P., Biscontin, A., De Pittà, C., Romualdi, C., Teschke, M., ... & Kawaguchi, S. (2015). Pyrosequencing and de novo assembly of Antarctic krill (E uphausia superba) transcriptome to study the adaptability of krill to climate-induced environmental changes. Molecular ecology resources, 15(6), 1460-1471. doi: 10.1111/1755-0998.12408

Nicol, S., & Endo, Y. (1997). Krill fisheries of the world (No. 367). Food & Agriculture Org..

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nature methods, 14(4), 417-419. doi: 10.1038/nmeth.4197

Peng, Y., Leung, H. C., Yiu, S. M., Lv, M. J., Zhu, X. G., & Chin, F. Y. (2013). IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels. Bioinformatics, 29(13), i326-i334. doi: 10.1093/bioinformatics/btt219

Risso, D., & Course, I. B. S. (2015). RNA-seq Normalization and Batch Effect Removal.

Sales, G., Deagle, B. E., Calura, E., Martini, P., Biscontin, A., De Pittà, C., ... & Jarman, S. (2017). KrillDB: A de novo transcriptome database for the Antarctic krill (Euphausia superba). PLoS One, 12(2), e0171908. doi: 10.1371/journal.pone.0171908

Seear, P. J., Goodall-Copestake, W. P., Fleming, A. H., Rosato, E., & Tarling, G. A. (2012). Seasonal and spatial influences on gene expression in Antarctic krill Euphausia superba. Marine Ecology Progress Series, 467, 61-75. doi: 10.3354/meps09947

Seear, P. J., Tarling, G. A., Burns, G., Goodall-Copestake, W. P., Gaten, E., Özkaya, Ö., & Rosato, E. (2010). Differential gene expression during the moult cycle of Antarctic krill (Euphausia superba). BMC genomics, 11(1), 1-13. doi: 10.1186/1471-2164-11-582

Siegel, V. (2005). Distribution and population dynamics of Euphausia superba: summary of recent findings. Polar Biology, 29(1), 1-22. doi: 10.1007/s00300-005-0058-5

Siegel, V. (Ed.). (2016). Biology and ecology of Antarctic krill. Cham, Switzerland: Springer.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, 31(19), 3210-3212. doi: 10.1093/bioinformatics/btv351

Suter, L., Polanowski, A. M., King, R., Romualdi, C., Sales, G., Kawaguchi, S., ... & Deagle, B. E. (2019). Sex identification from distinctive gene expression patterns in Antarctic krill (Euphausia superba). Polar Biology, 42(12), 2205-2217. doi: 10.1007/s00300-019-02592-3

Tarling, G. A., & Cuzin-Roudy, J. (2003). Synchronization in the molting and spawning activity of northern krill (Meganyctiphanes norvegica) and its effect on recruitment. Limnology and Oceanography, 48(5), 2020-2033. doi: 10.4319/lo.2003.48.5.2020

Valentine, J. W., & Ayala, F. J. (1976). Genetic variability in krill. Proceedings of the National Academy of Sciences, 73(2), 658-660. doi: 10.1073/pnas.73.2.658.

Zane, L., Ostellari, L., Maccatrozzo, L., Bargelloni, L., Battaglia, B., & Patarnello, T. (1998). Molecular evidence for genetic subdivision of Antarctic krill (Euphausia superba Dana) populations. Proceedings of the Royal Society of London. Series B: Biological Sciences, 265(1413), 2387-2391. doi: 10.1098/rspb.1998.0588

Zhao, Q. Y., Wang, Y., Kong, Y. M., Luo, D., Li, X., & Hao, P. (2011, December). Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. In BMC bioinformatics (Vol. 12, No. 14, pp. 1-12). BioMed Central. doi: 10.1186/1471-2105-12-S14-S2

# Data Accessibility

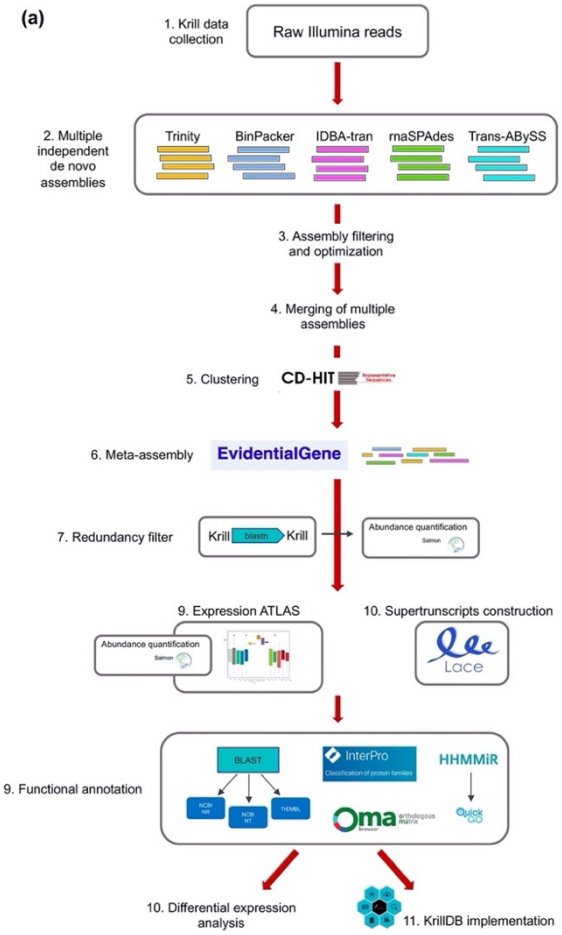
Data used for the krill transcriptome reconstruction and for the generation of the Expression Atlas was downloaded from the NCBI Short Read Archive, under accessions: PRJEB30084, PRJNA362526, PRJEB30084, PRJNA362526 and PRJNA640244.

# Author Contributions

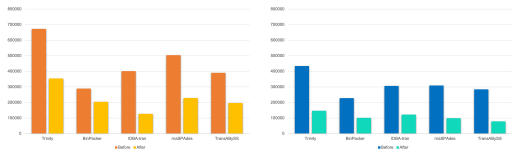
IU, CDP, BM and GS conceived the study. IU and DC performed the analyses. IU, AB, BM and GS wrote the manuscript. CB, CR and CDP advised on data analysis and reviewed the text.

# Tables and Figures

**FIGURE 1.** Workflow of the assembly process, annotation, database re-design and downstream analyses (**a**) with transcriptome quality assessment results (**b**, **c**, **d**). Panel **b** shows the results of the first assembly filtering in terms of total number of transcripts. Quality measures computed at each assembly step are reported in panel **b**, from the five de novo assembly algorithms (top), after the first filtering process (bottom) and finally comparing the quality of the EvidentialGene meta-assembly and the final krill transcriptome after the redundancy filter (right). BUSCO assessment results (**d**) on EvidentialGene transcriptome (left) and krill transcriptome after last filter (right): the EvidentialGene transcriptome was characterized by 95.3% Complete sequences (50.2% single-copy, 45.1% duplicated), 0.6% Fragmented and 4.1% Missing sequences. The same analysis on the final krill transcriptome reconstruction produced 93.2% Complete transcripts (49.8% Single-copy, 43.4% Duplicated), 0.6% Fragmented and 6.2% Missing sequences.




(b)



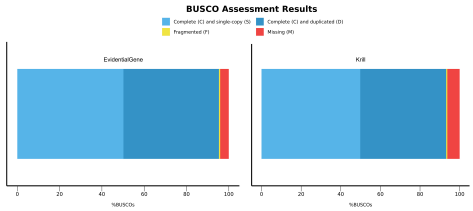
(c)

| | Trinity | BinPacker | maSPAdes | IDBA-tran | TransABySS |
|---|---|---|---|---|---|
| # Transcripts | 671837 | 288476 | 503293 | 400750 | 389351 |
| %GC | 35.56 | 34.66 | 35.39 | 34.77 | 35.07 |
| Median Contig Length | 368 | 938 | 347 | 336 | 332 |
| N50 | 1140 | 2213 | 1533 | 553 | 730 |
| # Bases | 470615830 | 413787317 | 392082747 | 198618480 | 218288366 |

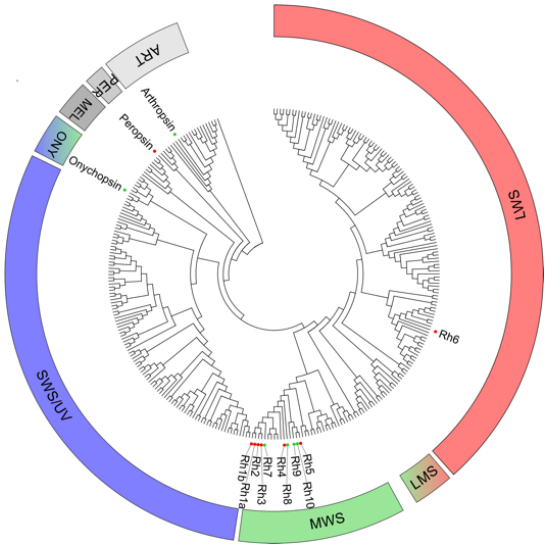
| | Trinity | BinPacker | maSPAdes | IDBA-tran | TransABySS |
|---|---|---|---|---|---|
| # Transcripts | 353340 | 203274 | 228038 | 125886 | 195764 |
| %GC | 35.77 | 34.80 | 35.61 | 35.18 | 35.17 |
| Median Contig Length | 452 | 1074 | 762 | 352 | 426 |
| N50 | 1455 | 2317 | 1958 | 670 | 1100 |
| # Bases | 301600820 | 321478538 | 280807245 | 69975046 | 144053299 |

| | EvidentialGene | Krill |
|---|---|---|
| # Transcripts | 274840 | 151595 |
| %GC | 36.18 | 36.34 |
| Median Contig Length | 756 | 1156 |
| N50 | 2164 | 2761 |
| # Bases | 360989701 | 264149525 |

(d)



686 **FIGURE 2.** Phylogenetic relationships of *Euphausia superba* opsins shown as circular
687 cladogram. Colored dots indicate krill opsins: red, previously cloned opsins; green,
688 novel identified opsins. The spectral sensitivities of rhabdomeric opsin clades were
689 inferred from the curated invertebrate-only opsin dataset proposed by DeLeo &
690 Bracken-Grissom, 2020. Represented opsin classes: LWS, long-wavelenght-
691 sensitive; LSM, long/middle-wavelenght-sensitive; MWS, middle-wavelenght-
692 sensitive; SWS/UV, short/UV-wavelenght-sensitive; ONY, onychopsins; MEL,
693 melanopsins; PER, peropsin; ART, arthropsin. Rectangular phylogram is reported in
694 **Figure S1**.




695 **TABLE 1.** List of contrast computed with total number of differentially expressed genes
696 and numbers of up- and downregulated genes.

| Contrast | # Total | # Upregulated | # Downregulated |
|---|---|---|---|
| Summer vs. Winter | 1195 | 1078 | 117 |
| Male vs. Female | 14 | 7 | 7 |
| Male/Summer vs. Female/Winter | 12 | 6 | 6 |
| South Georgia vs. Lazarev Sea | 79 | 26 | 53 |
| South Georgia vs. Bransfield Strait-South Orkney | 28 | 6 | 22 |
| Lazarev Sea vs. Bransfield Strait-South Orkney | 17 | 13 | 4 |
| South Georgia/Male vs. Bransfield Strait-South Orkney/Female | 10 | 6 | 4 |
| South Georgia/Male vs. Lazarev Sea/Male | 19 | 8 | 11 |
| South Georgia/Summer vs. Bransfield Strait-South Orkney/Winter | 75 | 66 | 9 |
| Lazarev Sea/Summer vs. Bransfield Strait-South Orkney/Winter | 359 | 173 | 186 |
| South Georgia/Summer vs. Lazarev Sea/Summer | 188 | 150 | 38 |
| Lazarev Sea/Male vs. Bransfield Strait-South Orkney/Female | 20 | 10 | 10 |

**FIGURE 3.** New search engine of KrillDB[2]: the homepage sections **(a)** with an example of the results of a fulltext-search **(b)**, of a blast-search **(d)** and the tables listing differentially expressed genes related to the contrast selected **(c)**.

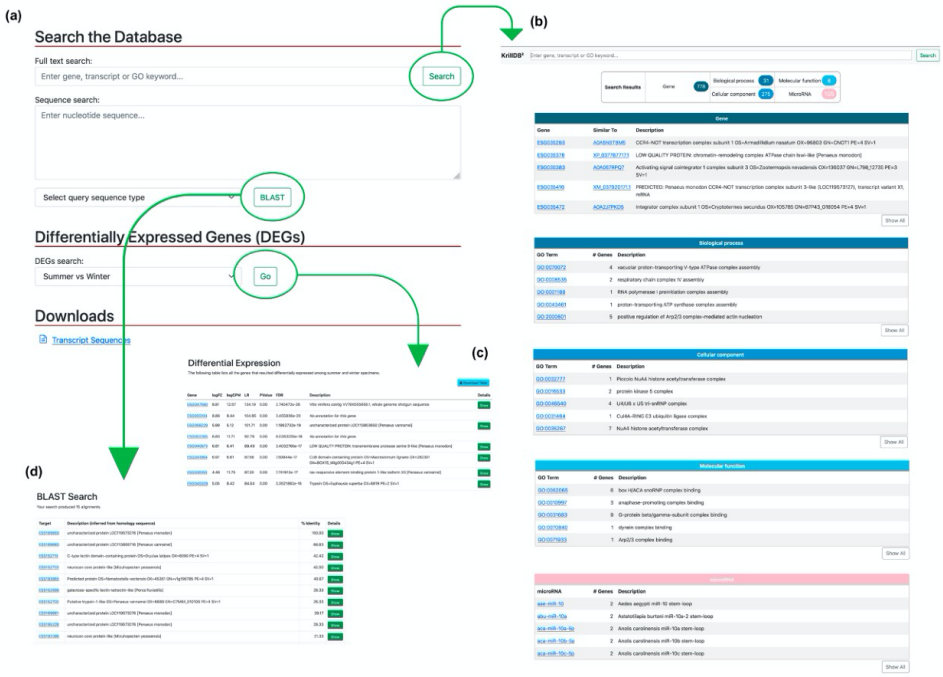



**FIGURE 4.** Table summarizing results of a GO term search: an example of a GO term associated to genes matching different protein domains **(a)** and a GO term associated to genes matching a known mature microRNA **(b)**.

(a)

## Search Results

You searched for genes annotated with the keyword GO:0035267 (NuA4 histone acetyltransferase complex). Found 7 matches:

Download Table

| Gene | Match | Description |
|---|---|---|
| ESG044786 | PTHR11093:SF2 | RUVB-LIKE 2 |
| ESG044926 | PTHR11093:SF6 | RUVB-LIKE 1 |
| ESG037594 | PTHR14898 | ENHANCER OF POLYCOMB |
| ESG044885 | PTHR12855 | DNA METHYLTRANSFERASE 1-ASSOCIATED PROTEIN 1 FAMILY MEMBER |
| ESG048735 | PTHR10880:SF29 | MORTALITY FACTOR 4-LIKE PROTEIN 1 |
| ESG036210 | PTHR15398:SF6 | BROMODOMAIN-CONTAINING PROTEIN 8 |
| ESG057428 | PTHR10880:SF29 | MORTALITY FACTOR 4-LIKE PROTEIN 1 |

(b)

## Search Results

You searched for genes annotated with the keyword GO:0019969 (interleukin-10 binding). Found 2 matches:

Download Table

| Gene | Match | Description |
|---|---|---|
| ESG086142 | sko-miR-10 | Saccoglossus kowalevskii miR-10 stem-loop |
| ESG070613 | sko-miR-10 | Saccoglossus kowalevskii miR-10 stem-loop |

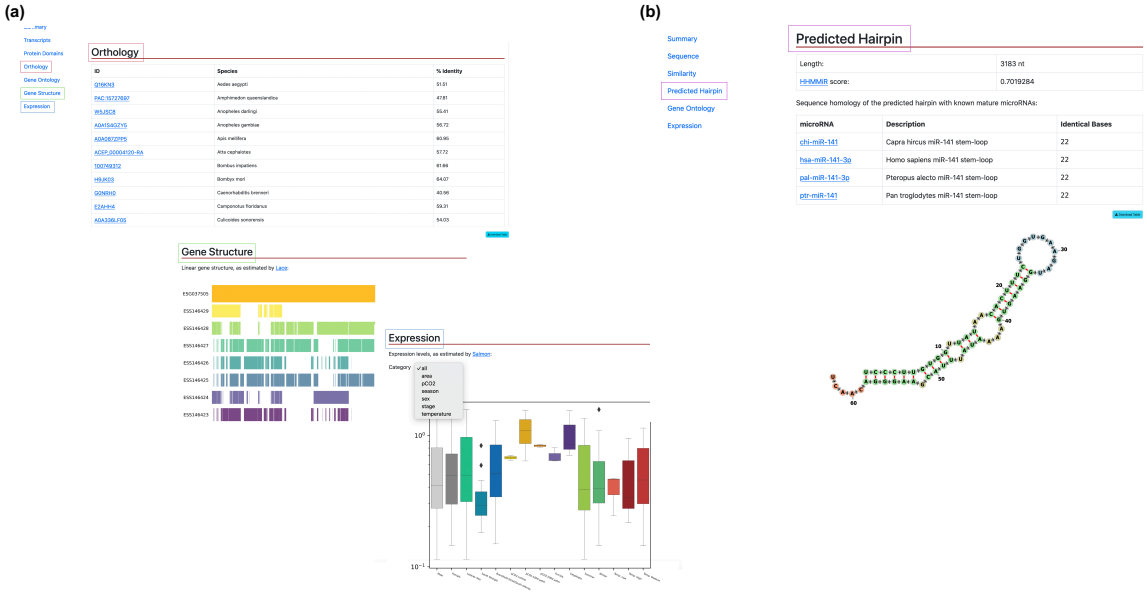



**FIGURE 5.** Additional sections in gene and transcript pages. The new sections in the gene-centric page show a table listing the orthologous sequences with their belonging species and the identity score, a visualization of the gene structure as estimated by Lace software and a boxplot coming from Expression Atlas analyses **(a)**. Both Orthology and Expression section are integrated also in the transcript-centric page. When a transcript was annotated as a putative microRNA, a "Predicted Hairpin" section displays a visualization of the hairpin predicted secondary structure and tables showing the alignment length, the HHMMiR score and the list of mature microRNA matching **(b)**.