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**: gabriele.sales@unipd.it, bettina.meyer@awi.de

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 already altered 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 it 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. Our study covered the entire developmental process, from larval stages to adult individuals, providing information of central relevance for ecological studies. Here we describe the KrillDB^2 database, a resource combining 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 initial indications about non-coding RNAs, a class of molecules whose contribute to krill physiology has 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 [1]. 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 [2, 3, 4], population dynamics and structuring [5, 6] and above all to understand its complex genetics [5, 7, 8, 9]. A sizable fraction of these studies focused on the DNA, specifically on mtDNA variation; however, the information available about krill genetics remains relatively modest. The difficulty in the progression of this kind of study mainly depends on the extraordinarily large krill genome size [10], which is more than 15 times larger than the human genome. This aspect largely complicates DNA sequencing, which is the reason why in recent years - together with the advances in high-throughput RNA-sequencing techniques - different krill transcriptome resources have been developed [11-16]. However, it was with the KrillDB project [17] that a detailed and advanced genetic resource was produced and made available to the community as an organized database. KrillDB is a web-based graphical interface with annotation results coming from the *de novo* reconstruction of krill transcriptome, assembling more than 360 million Illumina sequence reads.

Here we describe an update to KrillDB, now renamed KrillDB^2 (available at the address https://krilldb2.bio.unipd.it/), specifically focusing on two aspects: the improvement of the quality and breadth of the krill transcriptome sequences previously reconstructed, thanks to the addition of an unprecedented amount of RNA-sequencing data; and,

correspondingly, an increase in the amount of annotation information associated to each transcript. This improvement is made available through interactive graphs, images and downloadable files.

Material and methods

Krill collection

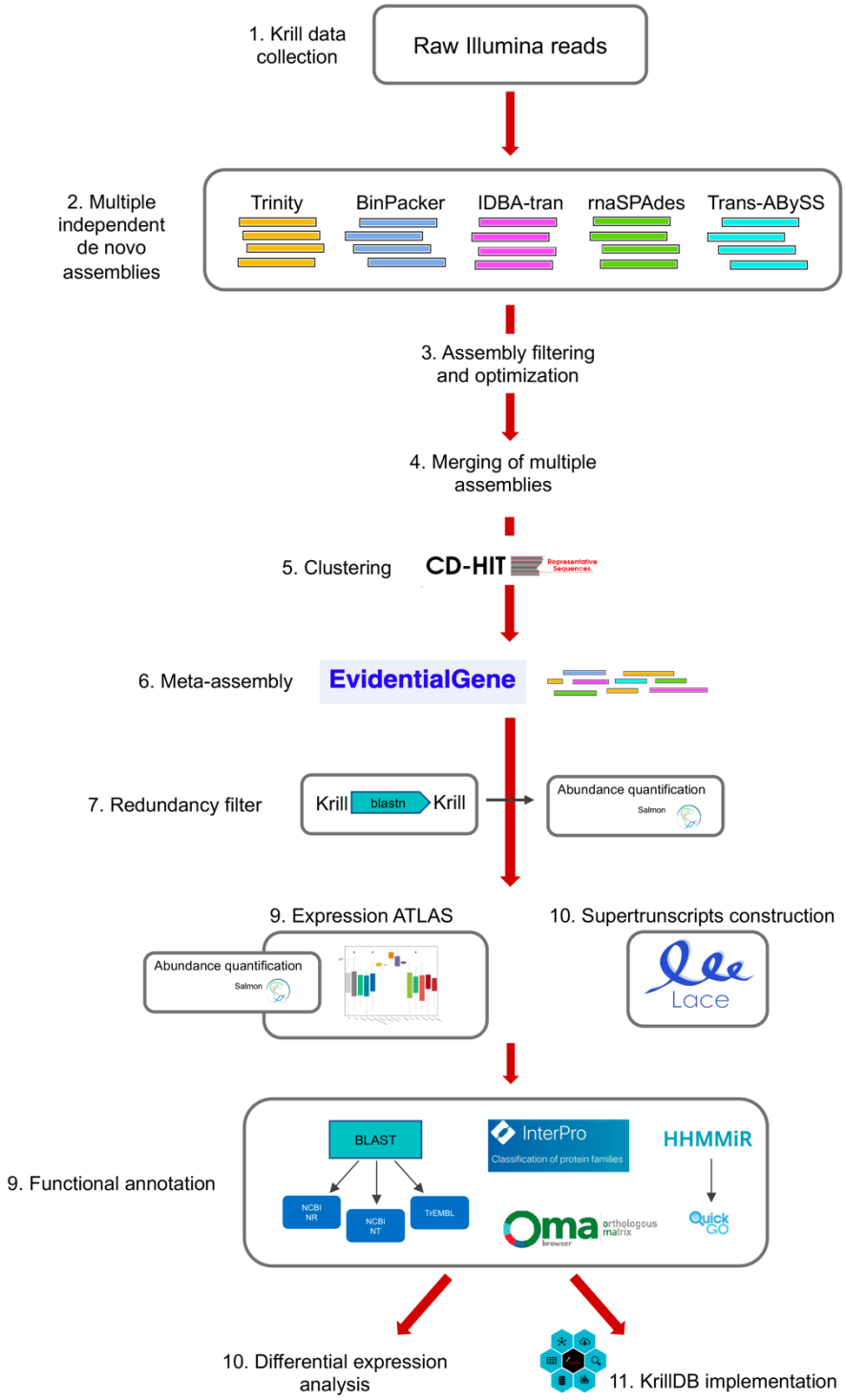
This study aims at covering the entire developmental process of krill. Therefore, we used samples coming from different developmental stages to cover the entire *E. superba* transcriptome, from larval to adult specimens. Specifically, adults included both male and female specimens, as well as summer and winter individuals and they also came from 3 different geographical regions: Lazarev Sea, South Georgia, and Bransfield Strait/South Orkney. The entire samples collection used to produce the new transcriptomic reference and carry out all downstream analysis is listed in **Table S1** (Supplementary Material).

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 selected the five which are arguably the most popular in the field: Trinity [18], BinPacker [19], rnaSPAdes [20], TransABySS [21] and IDBA-tran [22]. We summarized all the steps of the assembly reconstruction strategy, annotation process and downstream analyses in **Fig 1**.

88 **Fig 1. Workflow of the assembly process, annotation, database re-design and**
89 **downstream analyses.**



90 At first, we performed a separate transcriptome reconstruction with each of the tools
91 listed above. We evaluated their respective advantages through a series of
92 independent measures, such as: the total number of transcripts; %GC content; the

93 average fragment length; the total number of bases; the N50 value; and finally, the
94 results of the BUSCO analysis, which provides a measure of transcriptome
95 completeness based on evolutionarily informed expectations of gene content from
96 near-universal single-copy orthologs [23].

97 Assembly filtering and optimization

98 The raw sequencing data we used for the assemblies was obtained from different
99 experiments and included both stranded (**Table S1** – Group 2) and unstranded
100 libraries (**Table S1** – Group 1). As mixing these two types of libraries in a single
101 assembly is not well supported, we decided to run each software twice: we thus
102 generated a total of ten different *de novo* assemblies.

103 We used Trimmomatic [24] to remove adapter sequences and other artifacts from raw
104 Illumina sequences. The, the quality of trimmed reads was checked with the program
105 FastQC [25] (version 0.11.9). *De novo* transcriptome assembly was performed using
106 specific parameters depending on the library type (the actual commands used are
107 listed in **Table S2**, Supplementary Material).

108 Once assembled, a combination of two filtering steps was then applied to the newly
109 reconstructed transcriptomes to discard artifacts and improve the assembly quality.

110 First, we estimated the abundances of all the transcripts reconstructed by each
111 assembler using the Salmon software [26] (v. 1.4.0). Specifically, we used the
112 following parameters were used: samples coming from unstranded library (**Table S1**
113 – Group 1) were aligned using the options "-l ISR -1 --validateMappings"; samples
114 coming from stranded library (**Table S1** – Group 2) were aligned using the options "-l
115 IU --validateMappings". Samples were grouped according to the main experimental
116 conditions: (1) sex, with female and male levels; (2) geographical area, covering

117 Bransfield Strait, South Georgia, South Orkney and Lazarev Sea; and (3) season, with
118 summer and winter levels. Abundance estimates were imported in the R statistical
119 environment using the *tximport* package [27] and we implemented a filter to keep only
120 those transcripts showing an expression level of at least 1 transcript per million (TPM)
121 within each of the three experimental conditions.

122 In a second step, we considered the results of all assemblers jointly, and we ran the
123 "cd-hit-est" program [28] in order to cluster similar sequences and to produce a set of
124 non-redundant representative transcripts. Specifically, we collapsed all sequences
125 sharing 95% or more of their content, thus reducing the number of transcripts from
126 1,650,404 to 551,110.

127 Meta-assembly

128 The procedure described above was designed to identify near-duplicate sequences
129 deriving from different software, but likely corresponding to the same biological
130 transcript. As a further refinement, we were also interested in grouping resulting
131 transcripts into units corresponding to genes. To this end, we relied on the
132 EvidentialGene pipeline [29, 30]. We applied the "tr2aacds" tool which clusters
133 transcripts and classifies them to identify the most likely coding sequence representing
134 each gene. The software subdivides sequences into different categories, including
135 primary transcript with alternates (main), primary without alternates (noclass),
136 alternates with high and medium alignment to primary (althi1, althi, altmid) and partial
137 (part) incomplete transcripts. A "coding potential" flag is also added, separating coding
138 from non-coding sequences (see section "KrillDB^2 Web Interface"). The meta-
139 assembly thus obtained consisted in 274,840 putative transcripts, subdivided into
140 173,549 genes.

141 As these figures remained unrealistically high, we performed another round of
142 analyses to identify redundant or mis-assembled sequences still appearing in our
143 transcriptome. Here we used a combination of BLAST searches against known protein
144 and nucleotide databases (NR, NT, TREMBL) and information deriving from full-
145 length, experimentally validated transcripts from a previous study [31]. Results
146 confirmed that the newly reconstructed transcriptome fully represented krill RNAs, but
147 the large amount of input reads, together with the number of independent *de novo*
148 assemblers, likely led to an inflation in the number of alternative splicing variants being
149 reconstructed. Moreover, transcript alignments against BUSCO genes [23] and the
150 *doubletime*, *cry1, shaggy* and *vrille* full-length transcripts from [31] highlighted the fact
151 that multiple fragments of the same gene were incorrectly assembled as separate
152 transfrags. To remove these artifacts, first we aligned all transcript sequences in our
153 meta-assembly against each other using the *blastn* tool. We discarded all sequences
154 already included in a longer transcript for more than the 90% of their length. This filter
155 helped us remove 78,731 redundant sequences (29% of transcripts, overall). Then,
156 we ran a new abundance quantification using Salmon and we discarded all transcripts
157 with an average abundance below 0.1 TPM.

158 The combination of all the filters discussed above allowed us to reduce the number of
159 transcripts to 151,585 and, correspondingly, that of genes to 85,905. Our approach
160 discarded redundant genes, while retaining alternative transcripts with a sufficient level
161 of uniqueness in their sequence. This was confirmed by the fact that although we
162 removed almost 45% of the initially assembled transcripts, this filtering barely affected
163 the average read mapping rate, which went from 89% (initial EvidentialGene output)
164 to 88% (full filtering).

In order to enhance the interpretability of the transcriptome reconstruction, we also employed a SuperTranscripts analysis, on the basis of the workflow proposed by [32]. Specifically, we ran the Lace software (https://github.com/Oshlack/Lace) to reconstruct the block structure of each gene (see section “KrillDB^2 Web Interface”).

170 ## Functional Annotation

171 Assembled fragments were aligned against the NCBI NR (non-redundant)
172 UniProtKB/TrEMBL protein databases and against the NCBI NT nucleotide collection
173 (data downloaded on 22/04/2021). We also ran InterproScan (version 5.51-85.0) in
174 order to search for known functional domains and to predict protein family
175 membership. Results with an *e-value* greater than 1e-6 for proteins (*blastx*) or 1e-9
176 for nucleotides (*blastn*) were discarded.

177 Orthology inference was performed using the Orthologus MAtrix (OMA) standalone
178 package [33] (https://omabrowser.org/standalone/) which relies on a complete catalog
179 of orthologous genes among more than 2,300 genomes covering the entire tree of life.
180 This analysis helped us identify, based on protein sequences, those krill transcripts
181 showing an orthology relationship with genes from other species and which sets of
182 genes derived from a single common ancestral gene at a given taxonomic range [34].

183 Finally, all krill transcripts were compared against the RNAcentral database
184 (https://rnacentral.org/; https://doi.org/10.1093/nar/gkw1008) in order to identify any
185 homology with the mature sequences of known microRNAs from other species.

186 ## Expression Atlas

187 We used the final assembly described above to re-estimate transcript abundances
188 over a wide range of RNAseq dataset (see **Table S1**) including:

189 - Larval krill at two different stages of development exposed to different CO_2
190 conditions, coming from [17] (**Table S1** – Group 1)
191 - Adult krill (48 samples) coming from different geographical areas (Bransfield
192 Strait, Lazarev Sea, South Georgia, South Orkney) and different seasons

193 (summer and winter), divided into male and female specimens [35] (**Table S1**
194 – Group 2)

195 - Adult krill exposed to three different temperatures – Low Temperature, Mid
196 temperature, High Temperature (**Table S1** – Group 3)

197 - Adult krill divided into male and female specimens [36] (**Table S1** – Group 4)

198 Overall, these datasets include six experimental factors: geographical area, season,
199 developmental stage, pCO_2 exposure condition, sex and temperature. Newly
200 computed transcript abundances and raw counts were imported using R (version
201 4.0.5) and the package *tximport* (version 1.18.0). Batch effect removal was performed
202 using the *removeBatchEffect* function implemented in the *limma* package (version
203 3.46.0). The resulting count matrix of transcripts (rows) across samples (columns) was
204 then converted to the transcripts per million (TPM) scale. Finally, results were
205 summarized to the gene level using the *isoformToGeneExp* function
206 (IsoformSwitchAnalyzeR version 1.12.0). The expression levels for each experimental
207 condition are displayed in KrillDB^2 as a barplot, as part of the webpage for each gene
208 or transcript (see section "KrillDB^2 Web Interface").

209 Differential Expression Analysis

210 Transcript-level abundances and estimated counts were summarized at the gene-level
211 using the package *tximport*. Resulting counts were normalized to remove unwanted
212 variation by means of the RUVg method [37]. Specifically, we performed a preliminary
213 between-sample normalization (EDASeq, version 2.24.0) to adjust for sequencing
214 depth. Following the workflow outlined in the RUVseq vignette, we identified a set of
215 negative control genes with an FDR level larger than 0.8. We applied the RUVg
216 method to estimate k=2 factors of unwanted variation and we included those in the

217 design matrix for the final differential expression analysis, performed using the GLM
218 method implemented by the edgeR software (version 3.32.1). All p-values were
219 corrected using the Benjamini-Hochberg method.

220 MicroRNAs

221 We also investigated the possibility that the new transcriptome included sequences
222 corresponding to the precursors of krill microRNAs.

223 To this aim, we ran the HHMMiR software [38], which combines structural and
224 sequence information to train a Hierarchical Hidden Markov Model for the identification
225 of microRNA genes. We also performed a *blastn* search of all our assembled
226 transcripts against the collection of miRBase (http://www.mirbase.org/) mature
227 sequences. Results from these two analyses were combined: we collected all
228 transcripts with a HHMMiR score below or equal to 0.71 and an alignment to a known
229 mature microRNA with at most two mismatches. We then used the QuickGO tool
230 (https://www.ebi.ac.uk/QuickGO/) to identify any potential association among our
231 putatively identified microRNA precursors and GO categories.

232 Opsin phylogeny

233 To identify novel opsin genes in krill, we manually examined manually the list of
234 transcripts that were annotated as “opsin” by our automated pipeline. Furthermore, the
235 entire krill transcriptome was aligned against a curated opsin dataset (including 996
236 visual and non-visual opsins [39]) using Blast+ (version 2.11.0). For genes with
237 multiple alternative variants, we selected the longest transcript as a representative
238 sequence. Secondary structure was assessed by the NCBI Conserved Domain
239 Search (CDD database, May 2021). A phylogenetic tree was generated using the

MUSCLE alignment tool and the Maximum Likelihood method (Dayhof substitution matrix and Nearest-Neighbor-Interchange method) as implemented in MEGA X (version 10.2.6, https://www.megasoftware.net/). New opsins were aligned against a curated invertebrate-only opsin data set [40], the previously cloned krill opsins [41], and the full-length onychopsin and arthropsin sequences available on the NCBI Protein database (May 2021, ncbi.nlm.nih.gov/protein). The tree was rooted using the human G protein-coupled receptor VIPR1 as an outgroup.

Web Interface Implementation

The website was developed as a Python application based on the Flask framework. Data is stored in a PostgreSQL 12.8 database (http://www.postgresql.com). The sequences of the assembled transcripts and corresponding proteins are available for download as FASTA files. Gene and transcript pages have been updated with barplots implemented using the Seaborn Python library (version 0.11.1).

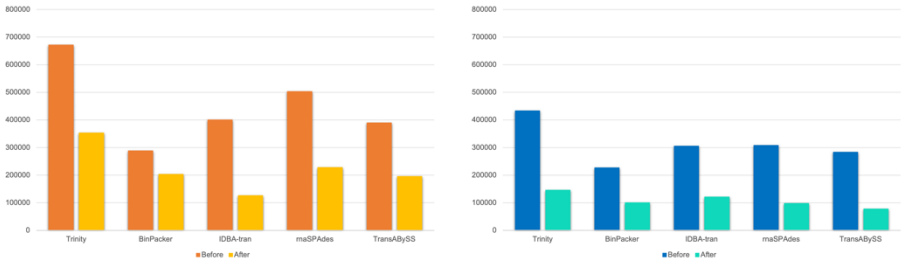
Results

Transcriptome Quality

We checked the quality of our transcriptome reconstruction quality step by step, starting from the 10 independent *de novo* assemblies, then evaluating the potential of merging all assemblies into a unique meta-assembly through EvidentialGene, and finally filtering the transcriptome for redundancy. All these results are summarized in **Fig 2**. As previously mentioned, the result of our reconstruction strategy was evaluated using different measures: the N50 statistics highlighted an increase in transfrags lengths at each step. Recent benchmarks, such as [42], have shown that, while reconstructing the transcriptome of a species, no single approach is uniformly superior at reconstructing the transcriptome of a species: the quality of each result is influenced by a number of factors, both technical (*k*-mer size, strategy for duplicate resolution) and biological (genome size, presence of contaminants). In our study we observed that, although a consistent number of sequences was removed through each step of the assembly, merging and filtering procedure, we didn't encounter any decline in the quality described by the basic statistics of the reconstructed transcripts.

Fig 2. **Transcriptome quality assessment results.** Panel **(a)** 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 **(c)** 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.

(a)



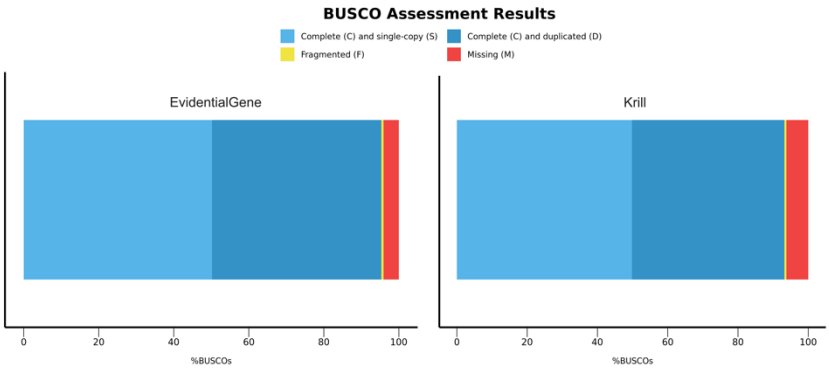
(b)

	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	151585
%GC	36.18	36.34
Median Contig Length	756	1156
N50	2164	2761
# Bases	360989701	264149525

(c)



We then explored the completeness of the krill transcriptome according to conserved ortholog content using BUSCO searching our sequences among all expected orthologs from Arthropoda phylum. This analysis confirms that our strategy for reducing redundancy did not affect transcriptome completeness: indeed, the fraction of complete single-copy essential genes drops by 2.1% while our strategy discards more than 44% redundant transfrags.

We finally compared our quality assessment results with those from previously released krill transcriptomes (**Table 1**). Our latest assembly significantly improves all

290 the metrics we have discussed above and highlights the potential of the filtering

291 strategy we have devised.

292 **Table 1.** Quality statistics of the previously released krill transcriptomes compared to
293 the newly assembled KrillDB[2]. GenBank accession GFCS00000000.1 refers to the
294 SuperbaSe krill transcriptome reference [43].

	GFCS00000000.1	KrillDB	KrillDB[2]
#Total Transcript	484080	133965	151585
Median Contig Length	439	683	1156
N50	1071	1294	2761
BUSCO - Complete	827 (81.6%)	536 (52.9%)	944 (93.2%)

295 ## Functional Classification

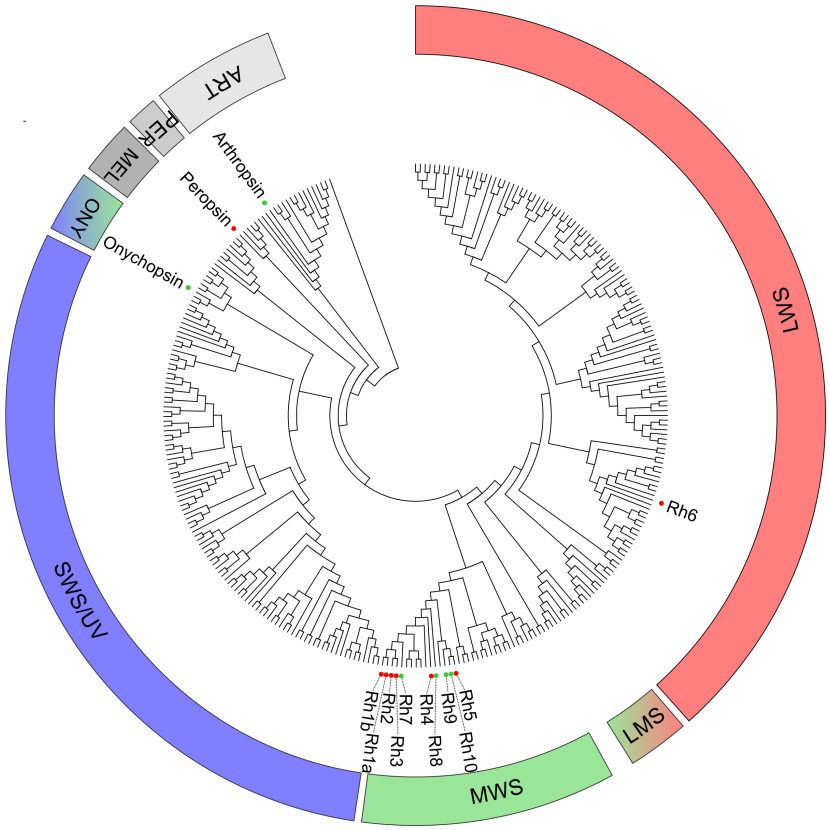
296 Results from the functional annotation analyses showed that 63,633 contigs matched
297 at least one protein from the NCBI NR (non-redundant) collection, corresponding to
298 about 42% of the total krill transcriptome, while 62,249 transfrags found a match
299 among UniProtKB/TREMBL protein sequences (41% of the total). Furthermore,
300 22,071 krill transcripts (15% of the total) had significant matches with sequences in
301 the NCBI NT nucleotide database. To classify transcripts by putative function, we
302 performed a GO assignment. Specifically, 2,612 GO terms (corresponding to 13068
303 genes) were assigned: 1,128 of those (corresponding to 1178 genes) represented
304 molecular functions; 1,099 terms (corresponding to 6991 genes) were linked to
305 biological processes; 385 terms (corresponding to 4303 genes) represented cellular
306 components.

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

308 To evaluate the gene discovery potential of the new assembly, we searched the
309 transcriptome for novel members of the opsin family. Opsins are a group of light
310 sensitive G protein-coupled receptors with seven transmembrane domains. 14 genes
311 were annotated as putative opsins and the conserved domains analysis revealed that
312 all of them possess the distinctive 7 α-helix transmembrane domain structure. The 8
313 previously cloned opsins [41] were all represented in KrillDB[2] (sequence identity
314 >90%; **Table S3**, Supplementary Material). The other 6 genes we identified can
315 therefore be considered new putative opsins. Among those, we found 4 putative
316 rhabdomeric opsins: *Es*Rh7 and *Es*Rh8, with 70% and 59 % of amino acid identity to
317 *Es*Rh1a and *Es*Rh4, respectively; *Es*Rh9 and *Es*Rh10 showing high sequence identity
318 (87% and 74%, respectively) to *Es*Rh5. Further, we identified 2 putative ancestral

opsins: a non-visual arthropsin (*Es*Arthropsin), and an onychopsin (*Es*Onychopsin) with 70% and 49% of sequence identity with crustacean and onychophoran orthologous, respectively. Phylogenetic analysis (**Fig 3**) suggested that *Es*Rh7-10 are middle-wavelength-sensitive (MWS) rhabdomeric opsins, and further confirmed *Es*Arthropsin and *Es*Onychopsin annotation.

Fig 3. **Phylogenetic relationships of *Euphausia superba* opsins shown as circular cladogram.** Colored dots indicate krill opsins: red, previously cloned opsins; green, novel identified opsins. The spectral sensitivities of rhabdomeric opsin clades were inferred from the curated invertebrate-only opsin dataset proposed by DeLeo & Bracken-Grissom, 2020. Represented opsin classes: LWS, long-wavelenght-sensitive; LSM, long/middle-wavelenght-sensitive; MWS, middle-wavelenght-sensitive; SWS/UV, short/UV-wavelenght-sensitive; ONY, onychopsins; MEL, melanopsins; PER, peropsin; ART, arthropsin. Rectangular phylogram is reported in **Fig S1** (Supplementary Material).



Differential Expression

The availability of a new assembly of the krill transcriptome, reconstructed collecting the largest amount of experimental data available thus far, suggested the possibility of performing a more detailed investigation of differential expression patterns. We decided to reanalyze the dataset from [35] to assess the possibility of identifying differentially expressed genes which were not detected in the original study due to the use of an older reference transcriptome [15].

Our design matrix for the model included all the independent factors (*season*, *area* and *sex*) and, in addition, the interaction between *area* and *season*, *sex* and *area*, *sex* and *season.*

In total 1,741 genes were found to be differentially expressed (DEG) among experimental conditions. They correspond to around 2% of the total reconstructed genes. In the previous work by [35] the same samples were quantified against a total of 58,581 contigs [15] producing a total of 1,654 DEGs. **Table 2** summarizes the list of contrasts that were performed, each one with the number of differentially expressed up and down regulated genes.

Table 2. List of contrast computed with total number of differentially expressed genes and numbers of up- and downregulated genes.

Reference condition	Alternative condition	Sample group	# Total	# Upregulated	# Downregulated
Summer	Winter	Group 2	1195	1078	117
Male	Femae	Group 2	14	7	7
Male/Summer	Female/Winter	Group 2	12	6	6
South Georgia	Lazarev Sea	Group 2	79	26	53
South Georgia	Bransfield Strait-South Orkney	Group 2	28	6	22
Lazarev Sea	Bransfield Strait-South Orkney	Group 2	17	13	4
South Georgia/Male	Bransfield Strait-South Orkney/Female	Group 2	10	6	4
South Georgia/Male	Lazarev Sea/Male	Group 2	19	8	11
South Georgia/Summer	Bransfield Strait-South Orkney/Winter	Group 2	75	66	9
Lazarev Sea/Summer	Bransfield Strait-South Orkney/Winter	Group 2	359	173	186
South Georgia/Summer	Lazarev Sea/Summer	Group 2	188	150	38
Lazarev Sea/Male	Bransfield Strait-South Orkney/Female	Group 2	20	10	10

353 1,195 DEGs were identified in the comparison between summer and winter
354 specimens: 1,078 were up-regulated and 117 down-regulated. 396 of such DEGs had
355 some form of functional annotation. In general, these results are in accordance with
356 the discussion by Höring [35], which found that seasonal differences are predominant
357 in comparison to regional ones. A summary of the DEGs is listed in **Table 3**. Complete
358 tables of differentially expressed genes are downloadable on KrillDB[2] (**Fig 4c**;
359 https://krilldb2.bio.unipd.it/, Section "Differentially Expressed Genes (DEGs)").

360 **Table 3. List of biologically relevant DEGs identified, starting from those already**
361 **described by Höring et al** [35]. Genes that were already found to be differentially
362 expressed in the work by Höring are reported in black, while newly DEGs identified by
363 our analysis are reported in red.

Process	Gene	KrillDB[2] Gene
Development of cuticle (moult cycle)	Peritrophin	ESG063925
	Chitooligosaccharidolytic beta-N- acetylglucosaminidase	ESG040750
	Carbohydrate sulfotransferase 11	ESG043538
	Trypsin like	ESG046724
	Chitinase 1	ESG041912
	Chitinase 3	ESG043598
	Chitinase 4	ESG040248
	Endochitinase-like	ESG041048
	Glycosyltransferase 8 domain-containing protein 1-like	ESG047683
	Collagen aplha-1	ESG039607
	Glutamine-fructose-6-phosphate aminotransferase	ESG040051
	Pupal cuticle protein 20-like	ESG045660
	Early cuticle protein 3	ESG054542
	Endocuticle	ESG037580
	Crustin 1	ESG059398
Immune response	Laccase	ESG048485
	Leucine rich repeat only protein 2	ESG048485
Embryogenesis	Blastula protease 10	ESG045350
Development and reproduction	Aldehyde dehydrogenase family 8	ESG043319
	Retinoid-inducible serine carboxypeptidase	ESG040940
	Dehydrogenase/reductase SDR family member 11	ESG048936
Reproduction	Vitellogenin	ESG035720
	Hematopoietic prostaglandin D synthase	ESG056241
	Carboxylic ester hydrolase	ESG040590
	Adiponectin receptor protein	ESG049090
	Type I iodothyronine deiodinase	ESG061750
	Ovochymase 1	ESG044749
	Ovochymase 2	ESG052923
	Serine/threonine-protein phosphatase PP1-gamma catalytic subunit	ESG045461
	Doublesex and mab-3 related transcription factor 1	ESG045173
Metalloendopeptidase activity	Neprilysin 1	ESG037511
Steroid metabolism	Inactive hydroxysteroid dehydrogenase-like protein 1	ESG050201
	Short-chain dehydrogenase/ reductase family 42E member 1	ESG041089
Lipid metabolism	Epoxide hydrolase	ESG048309
	Enoyl-CoA isomerase	ESG051749
	Long-chain-fatty-acid–CoA ligase	ESG040433
Glucose metabolic process	Furin-1 precursor	ESG037914
Cell cycle	Histone-lysine M-methyltransferase MLL5	ESG035391
Circadian clock	Euphausia superba cry gene for cryptochrome, exons 1-7	ESG035391
	Vrille	ESG040113
Photoreception	Opsin 5	ESG047639

364 ## Summer vs Winter

365 We selected a series of genes among seasonal DEGs according to what has been
366 already described in the literature. Höring et al. [35] previously identified and described
367 35 relevant DEGs involved in seasonal physiology and behavior: we recovered the
368 same gene signature in our own analysis by comparing summer to winter samples.
369 The majority of these DEGs appear to be involved in the development of cuticles
370 (*chitine synthase*, *carbohydrate sulfotransferase 11*), lipid metabolism (*fatty acid*
371 *synthase 2*, *enoyl-CoA ligase*), reproduction (*vitellogenin*, *hematopoietic prostaglandin*
372 *D synthase*), metabolism of different hormones (*type 1 iodothyronine deiodinase*) and
373 in the circadian clock (*cryptochrome*). Our results also include DEGs that were found
374 to be involved in the moult cycle of krill in other studies [16]. Specifically, we identified
375 a larger group of genes involved in the different stages of cuticle developmental
376 process (*peritrophin-A domain*, *calcified cuticle protein*, *glycosyltransferase 8-domain*
377 *containing protein 1*, *collagen alpha 1*, *glutamine-fructose 6 phosphate*), including
378 proteins such as *cuticle protein-3,6,19.8, early cuticle protein, pupal cuticle protein,*
379 *endocuticle structural glycoprotein, chitinase-3* and *chitinase-4*, the latter representing
380 a group of chitinase which have been shown to be expressed predominantly in gut
381 tissue during larval and/or adult stages in other arthropods and are proposed to be
382 involved in digestion of chitin-containing substrates [44]. Finally, in addition to *trypsin*
383 and *crustin 4* (immune-related gene, essential in early pre-moult stage when krill still
384 have a soft cuticle to protect them from pathogen attack, as seen by Seear et al. [16])
385 we also identified *crustin-1,2,3,5* and *7*. All the reported genes were up-regulated in
386 summer, the period in which growth take place and krill moult regularly.

387 Cuticle development genes were also identified as differentially expressed in the
388 analysis of the interaction of multiple factors, in particular between male samples

389 coming from South Georgia and female specimens coming from the area of Bransfield
390 Strait-South Orkney (considered as a unique area since they are placed at similar
391 latitudes). Strikingly, we also identified a pro-resilin gene, whose role in many insects
392 consists in providing efficient energy storage, being upregulated in South Georgia
393 male specimens.

394 Interaction Effects

395 A number of relevant DEGs were found among specific interactions of regional and
396 seasonal factors. In the comparison between krill samples in South Georgia in summer
397 and individuals sampled in Bransfield Strait-South Orkney in winter we found genes,
398 up-regulated in summer in South Georgia, that are related to reproductive activities,
399 such as *doublesex* and *mab-3 related transcription factor*. The latter is a transcription
400 factor crucial for sex determination and sexual differentiation which was already
401 described in other arthropods [45]. Since no differentially expressed gene related to
402 reproduction was found by Höring et al. [35] in the same comparisons, this suggests
403 that the new krill transcriptome improves the interpretability of expression studies and
404 the characterizations of krill samples.
405 Finally, the comparison between male individuals from the Lazarev Sea and female
406 specimens from the Bransfield Strait-South Orkney showed additional DEGs involved
407 in reproduction, such as *ovochymase 2*, usually highly expressed in female adults or
408 eggs, *serine protease* and a *trypsin-like gene*. In particular, *trypsin-like genes* are
409 usually thought to be digestive serine proteases, but previous works suggested that
410 they can play other roles [46]; many trypsins show female or male-specific expression
411 patterns and have been found exclusively expressed in males, as in our analysis,
412 suggesting that they play a role in the reproductive processes.

413 The simultaneous presence of differentially expressed genes involved in different
414 steps of the krill moulting cycle, in the reproductive process and in sexual maturation
415 that appear to be differentially expressed in same comparisons is in accordance with
416 what was already observed in krill [47] and other krill species [48]. In particular, there
417 is evidence of a strong relation between the krill moulting process and its growth and
418 sexual maturation during the year, which supports and confirms the reliability of our
419 results in terms of genes involved in such krill life cycle steps.

420 Identification of microRNA Precursors

421 In total we identified 261 krill transcripts with sequence homology to 644 known
422 microRNAs from other species. 306 sequences were linked to at least one GO term,
423 matching 54 krill transcripts (**Table S4**, Supplementary Material). Among them, we
424 identified 5 putative microRNAs involved with changes in cellular metabolism (age-
425 dependent general metabolic decline - GO:0001321, GO:0001323), as well as
426 changes in the state or activity of cells (age-dependent response to oxidative stress -
427 GO:0001306, GO:0001322, GO:0001324), 35 microRNAs involved in interleukin
428 activity and production. We found 26 putative microRNAs likely involved in
429 *ecdysteroidogenesis* (specifically GO:0042768), a process resulting in the production
430 of ecdysteroids, moulting and sex hormones found in many arthropods. In addition,
431 we found a microRNA involved in fused antrum stage (GO:0048165) which appears
432 to be related in other species to oogenesis. We also identified 27 microRNAs related
433 to *rhombomere* morphogenesis, formation and development (GO:0021661,
434 GO:0021663, GO:0021570). These functions have been linked to the development of
435 portions of the central nervous system in vertebrates, which share the same structure
436 of those found in arthropod brains. Lastly, 26 krill sequences showed high similarity
437 with 2 mature microRNA related to the formation of tectum (GO:0043676), which

438 represents in arthropods and, specifically, crustaceans, the part of the brain acting as
439 visual center.

440 KrillDB^2 Web Interface

441 The KrillDB website has been re-designed to include the new version of the
442 transcriptome assembly. **Figs 4**, **5** and **6** collect images taken from the new main
443 sections of the database. The integrated full-text search engine allows the user to
444 search for a transcript ID, gene ID, GO term, a microRNA ID or any other free-form
445 query. Results of full-text searches are now organized into several separate tables,
446 each representing a different data source or biological aspect (**Fig 4b**). Results of GO
447 term searches are summarized in a table reporting the related genes with
448 corresponding domain (**Fig 5a**) or microRNA (**Fig 5b**) match and associated
449 description. Both gene and transcript-centric pages have been extended with two new
450 sections: “Orthology" and “Expression” (**Fig 6a**). The Orthology section summarizes
451 the list of orthologous sequences coming from the OMA analysis, each one with the
452 species it belongs to and the identity score.

453 The “Expression” section shows a barplot representing abundances estimates
454 obtained from Salmon. An additional section, called “Gene Structure” (**Fig 6a**), was
455 added to the gene page on the basis of the results coming from the SuperTranscript
456 analysis. Specifically, we modified the STViewer.py Python script (from Lace),
457 optimizing and adapting it to our own data and database structure, in order to produce
458 a visualization of each gene with its transcripts. Since Lace relies on the construction
459 of a single directed splice graph and it is not able to compute it for complex clusters
460 with more than 30 splicing variants, this section is available for a selection of genes
461 only.

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

465 The KrillDB^2 homepage now includes two additional sections (**Fig 4a**): one is
466 represented by the possibility to perform a BLAST search. Any nucleotide or protein
467 sequence (*query*) can be aligned against krill sequences stored in the database.
468 Results are summarized in a table containing information about the krill transcripts
469 (*target*) that matched with the user's query, and the e-value corresponding to the
470 alignment. The other new section, called "Differentially Expressed Genes", allows the
471 user to browse all the tables listing the genes that were found to be differentially
472 expressed among the conditions we have described above. A drop-down menu gives
473 access to the different comparisons; DEG tables (**Fig 4c**) list for each gene its log fold-
474 change, p- and FDR values as estimated by edgeR. Moreover, each gene is linked to
475 a functional description (if available) inferred from sequence homology searches.

476 Information about krill transcripts that showed homology with an annotated microRNA
477 is available in the section "Predicted Hairpin" (**Fig 6b**). It contains a summary table
478 with details about the hairpin length and the similarity score (as estimated by
479 HHMMiR), followed by full listing of all the corresponding mature microRNAs (including
480 links to their miRBase page). In addition, an image displaying the predicted secondary
481 structure of the hairpin is included (computed by the "fornac" visualization software
482 from the ViennaRNA suite).
483

Fig 4. 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)**.

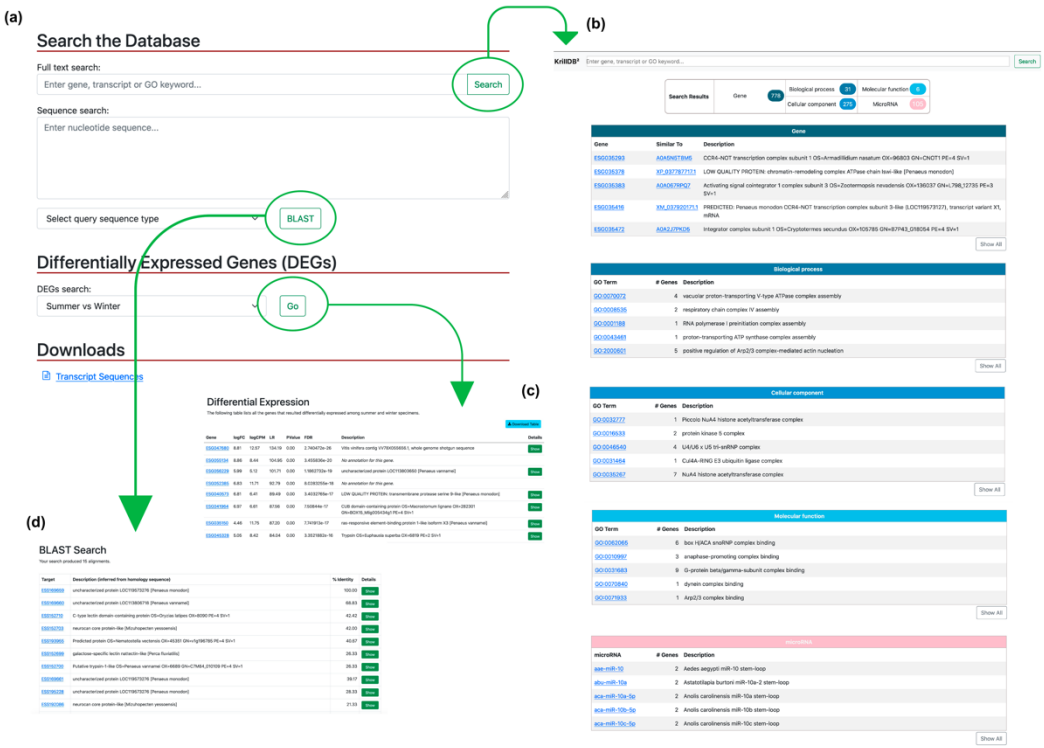


Fig 5. Table summarizing results of a GO term searc. An example of a GO term associated to genes matching different protein domains **(a)** and a GO term associated to genes matching a known mature microRNA **(b)**.

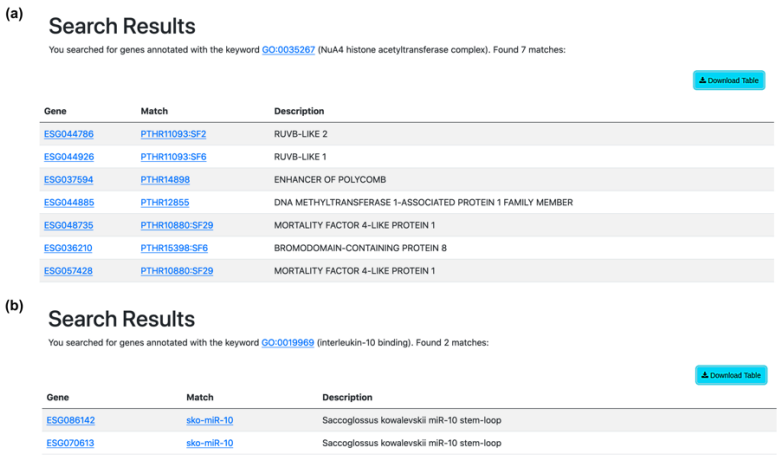
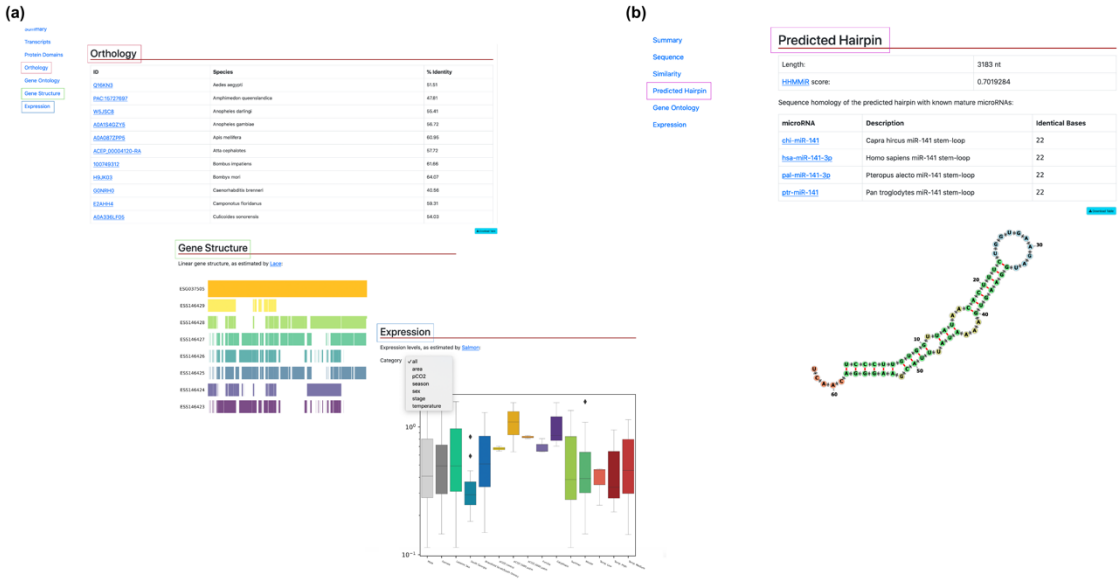


Fig 6. Additional sections in gene and transcript pages. The new sections in the gene-centric page show a table listing the orthologous sequences with their belonging species and the identity score, a visualization of the gene structure as estimated by Lace software and a boxplot coming from Expression Atlas analyses **(a)**. Both Orthology and Expression section are integrated also in the transcript-centric page. When a transcript is 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 microRNAs matching **(b)**.



Discussion

The availability of a large amount of public RNA-seq data capturing krill transcripts has given us the possibility to re-assemble its transcriptome and to significantly extend its annotation. We have now covered the entire developmental process of this species and included in our analysis individuals belonging to different seasons and affected by different environmental conditions. KrillDB[2] provides the most complete source of information about the krill transcriptome and will offer a reliable starting point development of novel ecological studies. As shown in **Table 1**, the analysis of the quality of previously released krill transcriptome in comparison to the newly assembled KrillDB[2] confirmed how the strategy applied did not produce any loss in terms of

509 quality, although a consistent number of transcripts was removed. The quality metrics,
510 in contrast, were improved both in terms of N50 statistics and transcriptome
511 completeness: the fraction of complete single-copy essential genes reached the
512 93.2%.

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

516 The identification of 6 novel putative opsin sequences almost double the eight that
517 were previously cloned, demonstrating a significant improvement in the gene
518 discovery potential of this new version of krill transcriptome. The finding of four novel
519 MWS rhabdomeric opsins, an onychopsin, and a non-visual arthropsin further enrich
520 the opsin repertoire of *E. superba* shedding light on a complex photoreception system
521 able to coordinate the physiological and behavioral responses to the extreme daily
522 (diel vertical migration) and seasonal changes in photoperiod and spectral
523 composition. Arthropsins are rhabdomeric non-visual opsins and its clade is the sister
524 group of the bilaterian rhabdomeric opsins [49, 50]. It was first discovered in the
525 crustacean *Daphnia pulex* and subsequently in other arthropods, onychophoran,
526 molluscs, anallids and flatworms [49-53]. Of relevance is the identification of an
527 onychopsin which has been suggested to be the common ancestor of *Panarthropoda*
528 visual opsins [49], and possibly sensitive to wavelength from UV to green light [54].
529 *Es*Onychopsin could represent the short-wavelength sensitive opsin (SWS/UV) which
530 we have long been searching for. Indeed, the absence of a SWS/UV opsin was truly
531 unexpected in an organism that shows daily vertical migration reaching depth beyond
532 the 30 m, where only short wavelength light can penetrate.

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

538 Acknowledgments

539 The position of Ilenia Urso was supported by the Helmholtz Virtual Institute
540 "PolarTime": Biological timing in a changing marine environment - clocks and rhythms
541 in polar pelagic organisms (VH-VI-500), headed by Bettina Meyer. Alberto Biscontin
542 was funded by the “Programma Nazionale di Ricerche in Antartide – PNRA” (grant
543 2016_00225) and by the Promega Corporation 2019 Real-Time PCR Grant Program.
544 We would also like to acknowledge the CAPRI initiative (Calcolo ad Alte Prestazioni
545 per la Ricerca e l’Innovazione”, University of Padova Strategic Research Infrastructure
546 Grant 2017) for the technical support and the HPC resources we have used for the
547 analyses. Cristiano Bertolucci was supported by the “Programma Nazionale di
548 Ricerche in Antartide – PNRA” (grant 2016_00225) and by the University of Ferrara
549 research grant (FIR2020 and FAR2021).

550

References

1. Nicol, S., & Endo, Y. (1997). Krill fisheries of the world (No. 367). Food & Agriculture Org..
2. 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
3. Hofmann, E. E., & Murphy, E. J. (2004). Advection, krill, and Antarctic marine ecosystems. Antarctic Science, 16(4). doi: 10.1017/s0954102004002275
4. 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
5. 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
6. 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.
7. 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
8. Goodall-Copestake, W. P., Perez-Espona, S., Clark, M. S., Murphy, E. J., Seear, P. J., & Tarling, G. A. (2010). Swarms of diversity at the gene cox1 in Antarctic krill. Heredity, 104(5), 513-518. doi: 10.1038/hdy.2009.188
9. 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
10. Jeffery, N. W. (2012). The first genome size estimates for six species of krill (Malacostraca, Euphausiidae): large genomes at the north and south poles. Polar Biology, 35(6), 959-962. doi: 10.1007/s00300-011-1137-4
11. 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
12. 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
13. 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
14. 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

15. 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
16. 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
17. 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
18. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., ... & Regev, A. (2011). Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature biotechnology, 29(7), 644. doi: 10.1038%2Fnbt.1883
19. 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
20. 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
21. 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
22. 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
23. 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
24. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. doi: 10.1093/bioinformatics/btu170
25. Andrews, S. (2017). FastQC: a quality control tool for high throughput sequence data. 2010.
26. 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
27. Love, M. I., Soneson, C., & Robinson, M. D. (2017). Importing transcript abundance datasets with tximport. Dim Txi. Inf. Rep. Sample1, 1, 5.
28. 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
29. Gilbert, D. G. (2019). Genes of the pig, Sus scrofa, reconstructed with EvidentialGene. *PeerJ*, *7*, e6374. doi: 10.7717/peerj.6374

30. Gilbert, D. G. (2019). Longest protein, longest transcript or most expression, for accurate gene reconstruction of transcriptomes? bioRxiv, 829184. doi: 10.1101/829184
31. Biscontin, A., Wallach, T., Sales, G., Grudziecki, A., Janke, L., Sartori, E., ... & Costa, R. (2017). Functional characterization of the circadian clock in the Antarctic krill, Euphausia superba. *Scientific reports*, *7*(1), 1-13. doi: 10.1038/s41598-017-18009-2
32. 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
33. 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
34. 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
35. Höring, F., Biscontin, A., Harms, L., Sales, G., Reiss, C. S., De Pittà, C., & Meyer, B. (2021). Seasonal gene expression profiling of Antarctic krill in three different latitudinal regions. Marine Genomics, 56, 100806. doi: 10.1016/j.margen.2020.100806
36. 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
37. Risso, D., & Course, I. B. S. (2015). RNA-seq Normalization and Batch Effect Removal.
38. Kadri, S., Hinman, V., & Benos, P. V. (2009). HHMMiR: efficient de novo prediction of microRNAs using hierarchical hidden Markov models. BMC bioinformatics, 10(1), 1-12. doi: 10.1186/1471-2105-10-S1-S35
39. Henze, M. J., & Oakley, T. H. (2015). The dynamic evolutionary history of pancrustacean eyes and opsins. Integrative and comparative biology, 55(5), 830-842. doi: 10.1093/icb/icv100
40. 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
41. 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
42. Hölzer, M., & Marz, M. (2019). De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. GigaScience, 8(5), giz039. doi: 10.1093/gigascience/giz039
43. Hunt, B. J., Özkaya, Ö., Davies, N. J., Gaten, E., Seear, P., Kyriacou, C. P., Tarling G., Rosato, E. (2017). The Euphausia superba transcriptome database, Superba SE: An online, open resource for researchers. *Ecology and evolution*, *7*(16), 6060-6077. doi: 10.1002/ece3.3168
44. Khajuria, C., Buschman, L. L., Chen, M. S., Muthukrishnan, S., & Zhu, K. Y. (2010). A 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

45. Jia, L. Y., Chen, L., Keller, L., Wang, J., Xiao, J. H., & Huang, D. W. (2018). Doublesex evolution is correlated with social complexity in ants. Genome biology and evolution, 10(12), 3230-3242. doi: 10.1093/gbe/evy250
46. 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
47. 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
48. 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
49. Hering, L, Henze, M.J., Kohler, M., Kelber, A., Bleidorn, C., Leschke, M., Nickel, B., Meyer, M., Kircher, M., Sunnucks, P., Mayer, G. (2012) Opsins in onychophora (velvet worms) suggest a single origin and subsequent diversification of visual pigments in arthropods. Molecular Biology & Evolution, 29(11), 3451-3458. doi: 10.1093/molbev/mss148
50. Hering, L., and Mayer, G. (2014). Analysis of the opsin repertoire in the tardigrade Hypsibius dujardini provides insights into the evolution of opsin genes in Panarthropoda. Genome Biol. Evol. 6, 2380–2391. doi: 10.1093/gbe/evu193
51. 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
52. 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
53. Futahashi, R., Kawahara-Miki, R., Kinoshita, M., Yoshitake, K., Yajima, S., Arikawa, K., et al. (2015). Extraordinary diversity of visual opsin genes in dragonflies. Proc. Natl. Acad. Sci. U.S.A. 112, E1247–E1256. doi: 10.1073/pnas.1424670112
54. 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

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.