

RESEARCH ARTICLE

# KrillDB: A *de novo* transcriptome database for the Antarctic krill (*Euphausia superba*)

Gabriele Sales<sup>1</sup>, Bruce E. Deagle<sup>2</sup>, Enrica Calura<sup>1</sup>, Paolo Martini<sup>1</sup>, Alberto Biscontin<sup>1</sup>, Cristiano De Pittà<sup>1</sup>, So Kawaguchi<sup>2</sup>, Chiara Romualdi<sup>1</sup>, Bettina Meyer<sup>3,4\*</sup>, Rodolfo Costa<sup>1\*</sup>, Simon Jarman<sup>2\*</sup>

**1** Department of Biology, University of Padova, Padova, Italy, **2** Australian Antarctic Division, Kingston, Tasmania, Australia, **3** Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany, **4** Institute for Chemistry and Biology of the Marine Environment, Carl von Ossietzky University Oldenburg, Oldenburg, Germany

\* [bettina.meyer@awi.de](mailto:bettina.meyer@awi.de) (BM); [rodolfo.costa@unipd.it](mailto:rodolfo.costa@unipd.it) (RC); [simon.jarman@gmail.com](mailto:simon.jarman@gmail.com) (SJ)



**OPEN ACCESS**

**Citation:** Sales G, Deagle BE, Calura E, Martini P, Biscontin A, De Pittà C, et al. (2017) KrillDB: A *de novo* transcriptome database for the Antarctic krill (*Euphausia superba*). PLoS ONE 12(2): e0171908. doi:10.1371/journal.pone.0171908

**Editor:** Cristiano Bertolucci, University of Ferrara, ITALY

**Received:** November 3, 2016

**Accepted:** January 27, 2017

**Published:** February 10, 2017

**Copyright:** © 2017 Sales et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Raw data are accessible on the National Center for Biotechnology Information Short Read Archive (SRA BioProject Accession PRJNA362526) at the URL: <http://www.ncbi.nlm.nih.gov/bioproject/362526>.

**Funding:** This work was supported by the following: \* the Helmholtz Virtual Institute “PolarTime” (VH-VI-500: Biological timing in a changing marine environment — clocks and rhythms in polar pelagic organisms, [www.polaritime.org](http://www.polaritime.org)), to BM; \* the Italian Programma Nazionale di Ricerche in Antartide — PNRA (grant

## Abstract

Antarctic krill (*Euphausia superba*) is a key species in the Southern Ocean with an estimated biomass between 100 and 500 million tonnes. Changes in krill population viability would have catastrophic effect on the Antarctic ecosystem. One looming threat due to elevated levels of anthropogenic atmospheric carbon dioxide (CO<sub>2</sub>) is ocean acidification (lowering of sea water pH by CO<sub>2</sub> dissolving into the oceans). The genetics of Antarctic krill has long been of scientific interest for both for the analysis of population structure and analysis of functional genetics. However, the genetic resources available for the species are relatively modest. We have developed the most advanced genetic database on *Euphausia superba*, KrillDB, which includes comprehensive data sets of former and present transcriptome projects. In particular, we have built a *de novo* transcriptome assembly using more than 360 million Illumina sequence reads generated from larval krill including individuals subjected to different CO<sub>2</sub> levels. The database gives access to: 1) the full list of assembled genes and transcripts; 2) their level of similarity to transcripts and proteins from other species; 3) the predicted protein domains contained within each transcript; 4) their predicted GO terms; 5) the level of expression of each transcript in the different larval stages and CO<sub>2</sub> treatments. All references to external entities (sequences, domains, GO terms) are equipped with a link to the appropriate source database. Moreover, the software implements a full-text search engine that makes it possible to submit free-form queries. KrillDB represents the first large-scale attempt at classifying and annotating the full krill transcriptome. For this reason, we believe it will constitute a cornerstone of future approaches devoted to physiological and molecular study of this key species in the Southern Ocean food web.

## Introduction

Antarctic krill (*Euphausia superba*) is a key species in the Southern Ocean with an estimated biomass between 100 and 500 million tons. The species provides a critical ecological link between primary producers and apex predators and supports commercial fisheries [1].

2013/C1.05, <http://www.pnra.it/>) to RC; \* the Australian Antarctic Science Program (AAS Project 4015, <http://www.antarctica.gov.au/science/information-for-scientists/research-guidelines/AASP-guidelines>) to SJ; \* the Alfred Wegener Institute through the research programme PACES II (topic 1, workpackage 5, <https://www.awi.de/en.html>) to BM.

**Competing interests:** The authors have declared that no competing interests exist.

Changes in krill population viability would have catastrophic effect on the Antarctic ecosystem [1,2]. One looming threat is ocean acidification (OA)—this is lowering of sea water pH due to elevated levels of anthropogenic atmospheric carbon dioxide (CO<sub>2</sub>) dissolving into the oceans [3]. This threat is recognized to be a global issue for the marine environment, but cold southern ocean waters are particularly susceptible due to high solubility of CO<sub>2</sub> [4]. Recent experimental work has documented the impact of increased CO<sub>2</sub> levels on krill development [2,5,6].

There has long been interest in the ecological genetics of Antarctic krill both for the analysis of population structure [7,8] and analysis of functional genetics [9–12]. However, the genetic resources available for this species are relatively modest. The genome size of the species is exceptionally large [13]; at 48.53 pg, it is more than 15 times larger than the human genome. Arthropods are the most diverse animal phylum, but their genomic resources have a relatively narrow taxonomic focus and little insight into the krill genome can be obtained from related species [14]. Even with rapid advance in DNA sequencing technology the current prospects for carrying out a complete genome sequencing and assembly project on krill is poor [15]. Sequencing of the krill transcriptome is much more feasible and could provide valuable information on coding sequences of many genes and will expedite the development of gene-linked markers [16].

Some transcriptome resources have been developed for this krill species including: sequencing of a few thousand ESTs [9,10], 454 assemblies from hundreds of thousands of sequence reads [11,17] and [18] (only raw reads available) and development of a 32,217 probe microarray [19]. However, all these resources are not yet available to the community as an organized resource as a database.

We have developed the most advanced genetic database on *Euphausia superba*, which include comprehensive data sets of former and present transcriptome projects. In particular, we have built a *de novo* transcriptome assembly using more than 360 million Illumina sequence reads generated from larval krill exposed to either normal or elevated CO<sub>2</sub> levels (short-term exposure employed to increase the representation of alternative transcripts, in particular with respect to an environmental stress). Using a meta-assembly strategy, we have produced and annotated a comprehensive Antarctic krill transcriptome that will contribute to the ongoing development of genomic resources for this species.

Here we present KrillDB, a web-based graphical interface to our annotation results. The database gives access to the full list of assembled genes and transcripts. This includes their reconstructed sequences, their level of similarity to transcripts and proteins from other species, the predicted protein domains contained within each transcript, the GO terms that can be linked to each transcript (based on previously described gene product characteristics) and the level of expression of each transcript in normal conditions and at the different CO<sub>2</sub> concentration levels. The KrillDB software implements a full-text search engine that makes it possible to quickly find all krill genes/transcripts linked to a query by submitting any free-form search text.

## Material and methods

### RNA-seq data

**Krill collection and larval krill CO<sub>2</sub> exposure.** Adult krill were collected in April 2011 from the Indian Ocean sector of the Southern Ocean (64°09' S, 100°460 E) and were maintained in the Australian Antarctic Division's marine research aquarium (see [20] for details). Krill were collected using permit AMLR 08-12-2337, issued under "Commonwealth of Australia, Antarctic Marine Living Resources Act 1981". These krill matured naturally in the laboratory, and gravid females spawned between December 2011 and February 2012. After hatching

batches of larvae were reared through the 12 recognizable stages of development over several months [21]. Larvae at two different stages of development were used in our experiment: (a) Stage 4 (*Calyptopsis I*) reached at approximately 2 weeks post-hatching and (b) Stage 11 (*Furcilia V*), approximately 16 weeks post-hatching. The procedure used for rearing these larvae is described in [20].

Larvae used in the experiment were reared under common ambient  $p\text{CO}_2$  conditions at  $0.5^\circ\text{C}$  and were randomly assigned to a  $p\text{CO}_2$  level treatment (control, 1000 or 2000  $\mu\text{atm } p\text{CO}_2$ ). A total of six experimental jars (2 larval stages  $\times$  3  $p\text{CO}_2$  levels) were used. For *Calyptopsis I*, 25 individuals were randomly selected and placed in each jar. Due to their larger size, only 5 *Furcilia V* larvae were used in each jar. The  $\text{CO}_2$  exposure experiment lasted 48 hours and was carried out concurrently for both larval stages in March 2012. Details of the experimental set-up are described in [5]. Briefly, sea water from a header tank (70 l at  $0.5^\circ\text{C}$ ) was equilibrated with air (control) or  $\text{CO}_2$ -enriched air before being delivered to experimental jars (250 ml clear polycarbonate) containing krill larvae. The  $p\text{CO}_2$  levels of  $\text{CO}_2$ -enriched air were monitored directly and also indirectly by pH measurement of the sea water. Experimental jars for each level of  $p\text{CO}_2$  were kept on separate shelves of a refrigerator maintained at  $0.5^\circ\text{C}$ . Overflow effluent from each jar was drained into a sump, and recirculated through a degassing unit before returning back to the header tank via a filtration system (see [5] for aquarium details).

**RNA extraction and high-throughput sequencing.** Larvae were transferred from the sea water treatment jars directly into RNeasy lysis solution using Pasteur pipettes. Total RNA was extracted using the ZR-Duet™ RNA miniPrep procedure (Zymo Research Corporation). Separate extractions were done for each *Furcilia* larva; these were minced with a scalpel blade and briefly homogenized with a microblender in lysis buffer during the first step of RNA extraction. Due to their small size *Calyptopsis* larvae were pooled and one RNA extraction performed for each  $\text{CO}_2$  treatment. These samples were homogenised by repeated pipetting in lysis buffer. Quantity and quality of extracted RNA was assessed using a Qubit® 2.0 Fluorometer (Life Technologies) and the Agilent 2100 bioanalyser (Agilent). For *Furcilia* larvae, RNA extracted from the five individuals in each treatment was pooled in equimolar amounts. Six samples corresponding to the experimental jars (2 larval stages  $\times$  3  $p\text{CO}_2$  levels) were sent to GeneWorks (Australia) for sequencing (Fig 1). GeneWorks used a TruSeq RNA sample prep kit (Illumina) and for each sample produced one lane of paired-end reads (2 $\times$ 100 bp) in a Genome Analyzer IIx sequencer (Illumina).

Raw data are accessible on the National Center for Biotechnology Information Short Read Archive (SRA BioProject ID: PRJNA362526, SRA sequences: . . . . .) at the URL: <http://www.ncbi.nlm.nih.gov/bioproject/362526>

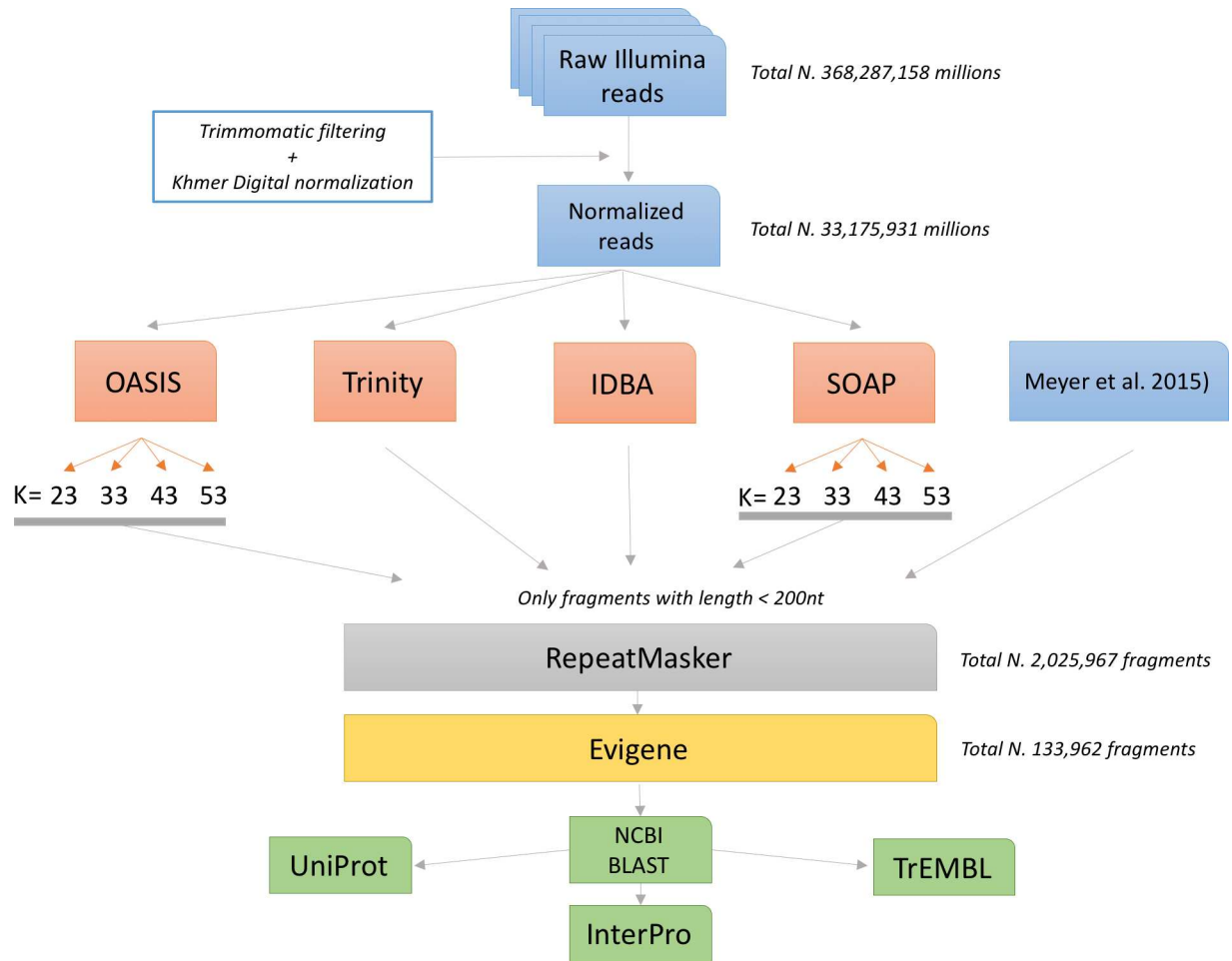
## Transcriptome assembly strategy

The general scheme of the transcriptome assembly and annotation is reported in Fig 1.

**Transcriptome assemblies.** We used Trimmomatic [22] to remove adapter sequences and other artifacts from the raw Illumina sequences. Reads were trimmed starting from the 3' end until the reported Q score was higher than 3. All resulting reads shorter than 75 nucleotides were discarded.

We then employed the khmer tool (<https://dx.doi.org/10.6084/m9.figshare.878460.v2>) to normalize the read coverage (“digital normalization”). As suggested by the software documentation, we also trimmed k-mers appearing less than two times in the entire dataset.

The resulting Illumina sequences were independently assembled using four different algorithms: Oases [23], Trinity [24], IDBA [25] and SOAPdenovo-Trans [26]. Both Oases and



**Fig 1. Outline of the assembly process.** Raw Illumina reads were first trimmed for adapters and for low quality bases at the 3' end. They were digitally normalized with the khmer software to reduce redundancy. These sequences were independently assembled using different software (OASIS, Trinity, IDBA and SOAP) and kmer sizes (23, 33, 43, 53). Information deriving from a previous assembly based on the 454 sequencing technology was added to further increase the transcriptome coverage. Repeated sequences were identified and removed using the RepeatMasker software in order to reduce the number of chimeric misassemblies. All surviving fragments were merged into a single transcriptome using the Evidential Gene pipeline. Results were annotated using sequence homology (BLAST) and protein domain searches (InterproScan).

doi:10.1371/journal.pone.0171908.g001

SOAPdenovo-Trans depend on the choice of the k-mer size used for the construction of the de-Bruijn graph. We thus performed multiple assemblies for different values of  $k$  (23, 33, 43 and 53) in order to compare the effects of this choice. In all cases, we discarded assembled fragments shorter than 200 nucleotides.

The presence of repeats in a species transcriptome has been linked to the erroneous generation of chimeric fragments. Those are formed by the transcriptome assembler mistakenly joining parts of different transcripts if they happen to share the same repeated sub-sequence. In order to reduce the number of these errors, we have filtered all the reconstructed fragments using the RepeatMasker software (RepeatMasker Open-4.0. 2013–2015 <http://www.repeatmasker.org>; we used the RepeatMasker library version 20150807). This procedure removed 29,645 (18%) of the fragments initially assembled.

**Merging of multiple assemblies.** We used the EvidentialGene pipeline (Gilbert D. 2013 <https://f1000research.com/posters/5-1695>) to identify redundant fragments and to select

sequences with the highest coding potential. Table 1 reports the quality measures for this approach. Both fragment lengths and protein matches against the NR database are significantly increased.

To further improve the quality of the transcriptome reconstruction, we used the same software pipeline to merge the results from a previous study which was based on the 454 sequencing technology (175,570 fragments total; see [11]). This approach provided further confirmation of the assembled sequences for 19% of the transcripts in KrillDB.

**Assembly comparison.** We computed a number of measures to assess the quality of the transcriptomes obtained from different assemblers: i) the total number of fragments, ii) the total number of assembled nucleotides, iii) the average fragment length, iv) the N50 index, v) the proportion of fragments linked to known sequences and vi) the ability to reconstruct (putative) full-length transcripts.

**Functional annotation.** We ran BLAST searches to annotate assembled fragments. We aligned our sequences against the NCBI NR (non-redundant) and UniProtKB/TrEMBL protein databases [27,28], and also against the NCBI NT nucleotide collection (data downloaded on 8/10/2015). Results with expectation values greater than 1e-6 for proteins (*blastx*) or 1e-9 for nucleotides (*blastn*) were discarded. Similarly, we ignored all aligned fragments having less than 50% of sequence identity with their target.

To further extend the annotation coverage, we employed InterproScan [29]. This software combines multiple tools to scan novel sequences in search of known functional domains. It is thus able to predict protein family membership.

## Database design and implementation

The pages of KrillDB are dynamically generated by a Python application based on the Flask framework and are served by the Nginx web server. Data is stored in a PostgreSQL 9.4 database (<http://www.postgresql.com>); full-text searches are implemented using the pg\_trgm module. The sequences of the assembled transcripts are available for download as FASTA files.

## Results

### Krill transcriptome assembly

Using Illumina sequencing, we generated a total of 368 million raw reads: 77, 69 and 67 million respectively for control, 1000 or 2000  $\mu\text{atm}$   $p\text{CO}_2$  in *Calyptopis I* samples and 47, 56 and 53 million for *Furcilia V* samples. After the filtering and cleaning process (see Methods for details), a total of 177 million (96,3%) high-quality reads were further used for the assembly.

Four de-novo assembly algorithms have been used on filtered and normalized (with digital normalization) reads: OASES [23], Trinity [24], IDBA [25] and SOAPdenovo-Trans [26]. Similar to what has been observed previously [30], we noticed that the choice of parameters for the assembly algorithms we used had a rather strong impact on the final results. Moreover, it is now increasingly clear that no best method exists but that different assemblers seemed to be

**Table 1. Quality measures of the three de-novo assembly algorithms.**

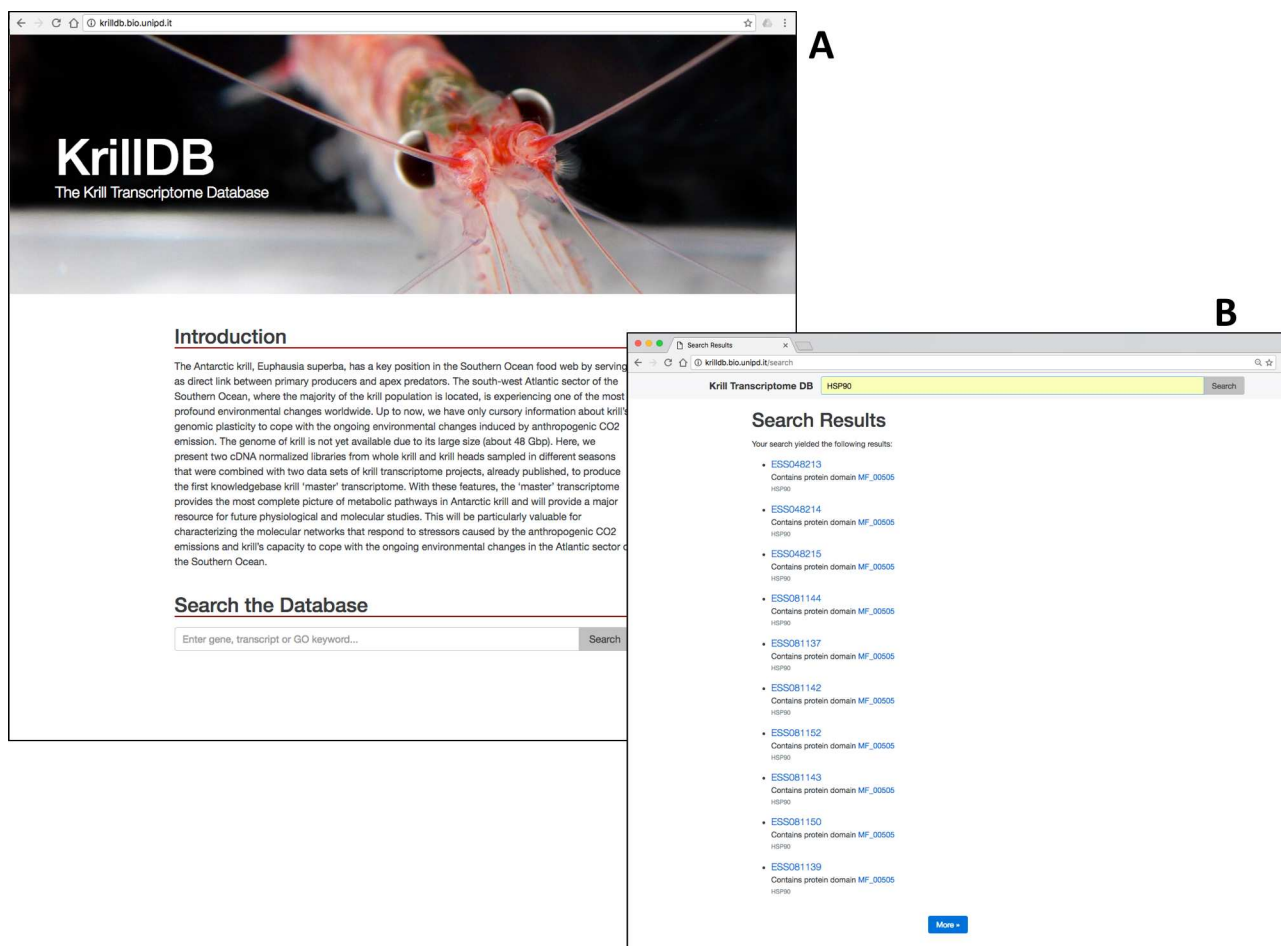
Software	# Frags	# tot Bases	Avg. Frag. Length	N50	% Matches	# Proteins
IDBA	183,688	78,682,848	428 ( $\pm$ 398)	445	13,06%	15,747
Butterfly	190,588	107,415,431	564 ( $\pm$ 596)	770	12,94%	14,927
Oases ( $k = 33$ )	237,717	159,270,607	670 ( $\pm$ 742)	1,009	13,79%	14,291
SOAP ( $k = 43$ )	123,855	54,662,513	441 ( $\pm$ 428)	484	16,88%	14,998
Evigene	133,962	129,183,922	964 ( $\pm$ 840)	1,294	67,27%	27,928

doi:10.1371/journal.pone.0171908.t001



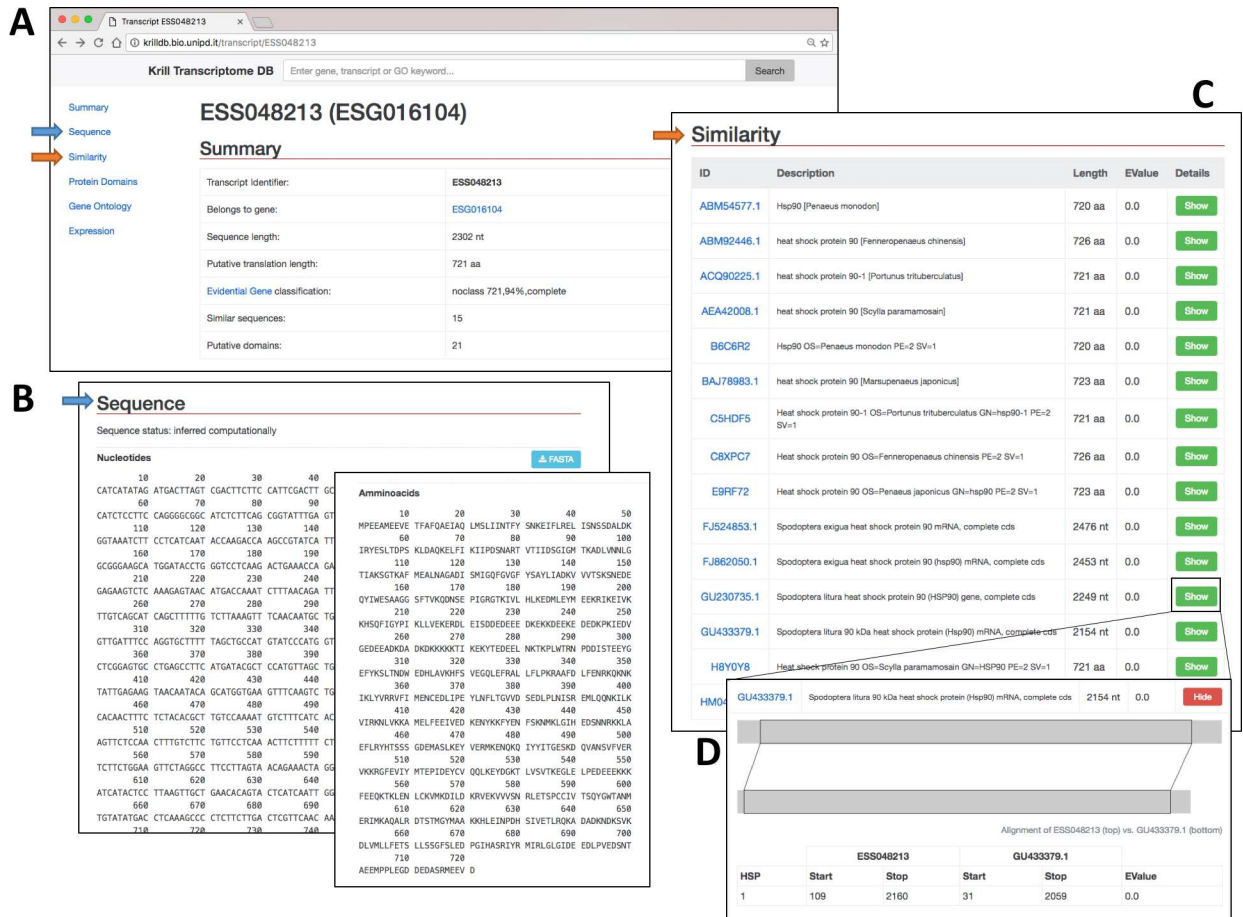
able to capture different sets of true transcripts. We thus decided to combine the results from all methods and choice of parameters in order to obtain a more thorough picture of the krill transcriptome. To this aim we use a recent approach (Gilbert D. 2013 <https://f1000research.com/posters/5-1695>) called EvidentialGene (hereafter evigene) that uses the results of different assembly algorithms extracting the best of their results. Specifically, evigene clusters fragments and then uses the presence and the length of CDS to select the best fragment for a specific transcript among all the available assemblies. Recent works [30,31] have shown that this approach is able to maximize the diversity of the assembled transcripts and their completeness, while limiting sequence redundancy.

For each assembly and for evigene final results we estimated several quality measures based on sequence features (the total number of fragments, their average length, the distribution of frags lengths, the total number of base pairs covered by the assembly and the N50 index) and functional annotation (proportion of fragments with a Blast hit, unique number of NR hits, and the distribution of the protein coverage). The results are reported in Table 1 and Fig 1. Looking at the result of each single assembler we found that OASES gives the highest number of fragments and then the highest number of total base pairs and N50, followed by Trinity and then IDBA and SOAP. The fragments identified by OASES are longer and match with a higher proportion of known proteins, followed by Trinity and then IDBA. On the other hand, OASES assembly seems



**Fig 2. (A)** The home page of KrillDB. **(B)** An example of search results. Here we queried the database for all hits containing annotations related to the 'HSP90' protein family.

doi:10.1371/journal.pone.0171908.g002



**Fig 3. Further exploration of results obtained with the search shown in Fig 2.** (A) The summary page for a single transcript. Links to the sequence and similarity sections are highlighted. (B) Sequence records. Nucleotide and amino acid sequences are displayed and can be downloaded as text file in the FASTA format. (C) Sequence similarity results obtained from BLAST are both summarized in a table and (D) depicted graphically to show the matches among different regions of the query and the target sequences.

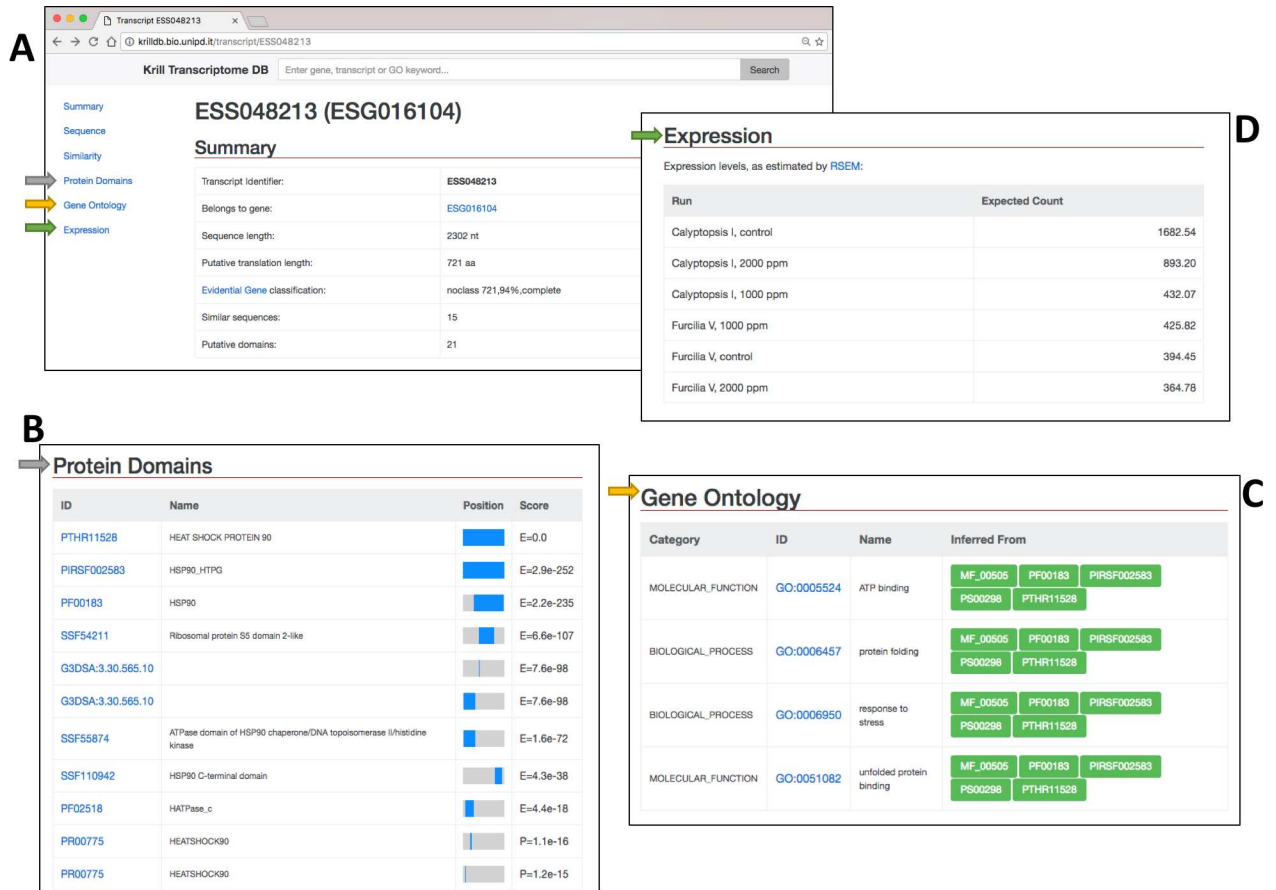
doi:10.1371/journal.pone.0171908.g003

to be characterized by a higher redundancy (showing a slightly lower number of recognized proteins) with respect to Trinity, SOAP and IDBA. The presence of a higher redundancy is not necessarily a negative aspect, since it may be due to the ability of identifying different isoforms of the same gene. The strategy adopted by evigene reduces the number of fragments, improving fragment length and annotation (Table 1 and Fig 1). It is interesting to note that evigene selects 133,962 fragments, of which 37% derived from OASES annotation, 27% from Trinity, 7% from IDBA, 10% from SOAP and 19% from the previously published transcriptome, confirming the above observation that OASES was on our data the best single assembly algorithm.

Of the 90,121 (67%) fragments of the evigene assembly that have a BLAST hit, 56,026 (62%) have associated at least one GO category. The InterProScan annotation gave us functional annotations for 1,300 other fragments which lacked any BLAST hit.

### Home page of KrillIDB

Users can access KrillIDB at <http://krilldb.bio.unipd.it/> (Fig 2A). The home page contains an introduction to KillIDB and PolarTime project and gives access to the full text search engine (Fig 2A). The user is free to search any keyword regarding for example the description of the transcript, the Gene Ontology category associated or the protein domain identified.



**Fig 4.** (A) Each transcript summary links to detailed sections about protein domains, gene ontology annotations and expression levels. (B) Protein domains detected within the transcript are visualized along with their ID, description, e-value and position on the sequence. (C) The list of Gene Ontology categories inferred by InterproScan. (D) Expression levels for each sequenced sample estimated by the RSEM software.

doi:10.1371/journal.pone.0171908.g004

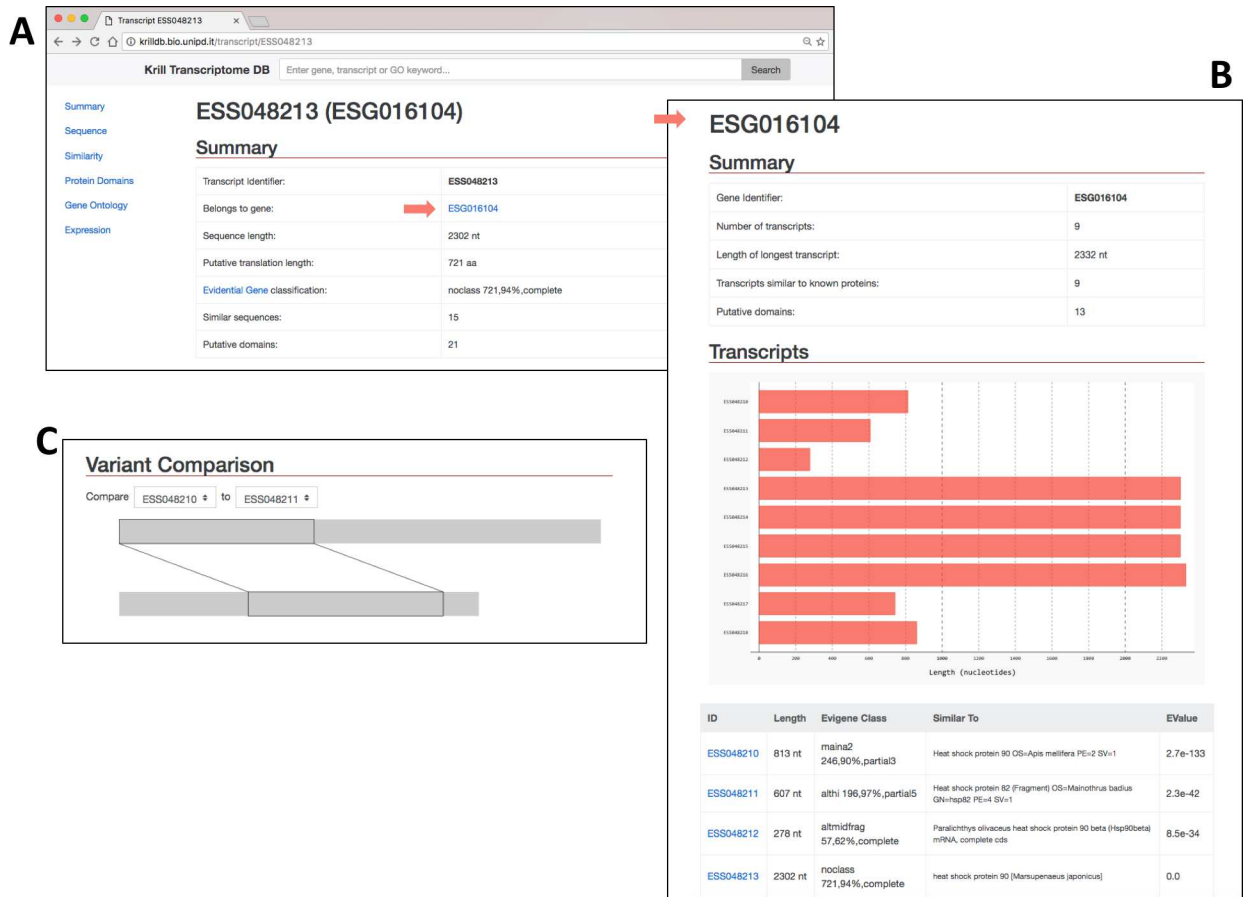
## Browsing the KrillDB

The search engine shows the resulted hits into a dynamic web page (Fig 2B). Every hit has a hyperlink that redirects the user to the detailed description the feature. An initial summary is reported with the name of the transcript, the reference gene, the length of the sequence and of the protein, the number of similar sequences (with BLAST) and the number of putative domain identified (Figs 3A, 4A and 5A).

Transcript and protein sequences can be visualized and downloaded as FASTA files (Fig 3B), while in the sequence similarity section (Fig 3C) the ID, the description, the length and the hit e-value are reported. Moreover, for each similar sequence the alignment is visualized and reported as start and end position of the conserved regions (Fig 3D).

In the protein domain section (Fig 4B) the ID (with the link to the reference database), the name, the e-value and the domain position within the protein sequence are reported. In case Gene Ontology categories have been associated to the transcript a dedicated section with the GO ID, the name of the term and the name of the protein domain from which the term definition derived is reported (Fig 4C). If the user selects the name of a protein domain, the GO terms associated with it will be highlighted the above table.





**Fig 5. (A)** Transcript fragments are clustered into groups putatively corresponding to genes. Each transcript page is thus linked to a group page. **(B)** Summary of a transcript group, showing a graphical comparison of the lengths of its members, the most significant BLAST hits and **(C)** pairwise alignment of all transcripts within the group.

doi:10.1371/journal.pone.0171908.g005

Finally, the expression quantification obtained with RSEM [32] is shown at the end of the page for each of the six libraries (Fig 4D).

Evigene assembly groups reconstructed transcripts according to their similarities. In the summary the ID of the gene (group of transcripts) is reported along with its hyperlink. The web page of the gene (Fig 5B) summarizes key information regarding number of transcripts, their lengths, their annotations along with a visual representation of their pairwise alignments (Fig 5C).

## Conclusions

Antarctic krill are incredibly abundant and represent key species of most food webs in the Southern Ocean. However, the lack of available genome sequences and the limited EST sequences stored in NCBI considerably restrict study progress on the molecular mechanisms of this species. In this study, we aim to produce large numbers of transcript sequences with corresponding annotation information and make these data freely accessible to users. As such, we developed KrillDB, which exhibits simplicity of use for researchers and contains 133,962 transcribed sequences assembled from the combined new Illumina sequences and previously published 454 assembled pyrosequencing data. This database provides information regarding the sequences and functional annotations of the transcripts. The database also includes the

developmental stages expression feature for each transcript, which was quantified on RNA-seq data. This database will represent a useful bioinformatics tool for studying molecular processes in krill. KrillDB represents the first large-scale attempt at classifying and annotating the full krill transcriptome. For this reason, we believe it will constitute a cornerstone of future approaches devoted to the study of this species.

## Acknowledgments

We thank Rob King and others at the Australian Antarctic Division krill aquarium for providing larval krill and setting up the CO<sub>2</sub> exposure experiment.

## Author Contributions

**Conceptualization:** GS BED BM RC SJ.

**Data curation:** GS BED SJ.

**Formal analysis:** GS EC PM AB CDP CR.

**Funding acquisition:** BM RC SJ.

**Investigation:** BED SK BM SJ.

**Methodology:** GS CR.

**Project administration:** CDP CR BM RC SJ.

**Resources:** CDP CR BM RC SJ.

**Software:** GS EC PM CR.

**Supervision:** CDP CR BM RC SJ.

**Validation:** BM RC SJ.

**Visualization:** GS EC PM.

**Writing – original draft:** GS CDP CR BM RC SJ.

**Writing – review & editing:** GS BED CDP CR BM RC SJ.

## References

1. Nicol S, Foster J, Kawaguchi S. The fishery for Antarctic krill—recent developments. *Fish Fish.* 2012; 13: 30–40.
2. Flores H, Atkinson A, Kawaguchi S, Krafft BA, Milinevsky G, Nicol S, et al. FEATURE ARTICLE: REVIEW Impact of climate change on Antarctic krill. *Mar Ecol Prog Ser.* 2012; 458: 1–19.
3. Raven J, Caldeira K, Elderfield H, Hoegh-Guldberg O, Liss PS, Riebesell U, et al. Ocean Acidification due to Increasing Atmospheric Carbon Dioxide. Royal Society Policy Document; 2005. Available: <https://ueaeprints.uea.ac.uk/32712/>
4. McNeil BI, Matear RJ. Southern Ocean acidification: a tipping point at 450-ppm atmospheric CO<sub>2</sub>. *Proc Natl Acad Sci U S A.* 2008; 105: 18860–18864. doi: [10.1073/pnas.0806318105](https://doi.org/10.1073/pnas.0806318105) PMID: [19022908](https://pubmed.ncbi.nlm.nih.gov/19022908/)
5. Kawaguchi S, Kurihara H, King R, Hale L, Berli T, Robinson JP, et al. Will krill fare well under Southern Ocean acidification? *Biol Lett.* 2011; 7: 288–291. doi: [10.1098/rsbl.2010.0777](https://doi.org/10.1098/rsbl.2010.0777) PMID: [20943680](https://pubmed.ncbi.nlm.nih.gov/20943680/)
6. Kawaguchi S, Ishida A, King R, Raymond B, Waller N, Constable A, et al. Risk maps for Antarctic krill under projected Southern Ocean acidification. *Nat Clim Change.* 2013; 3: 843–847.
7. Fevolden SE, Schneppenheim R. Genetic homogeneity of krill (*Euphausia superba* Dana) in the Southern Ocean. *Polar Biol.* 1989; 9: 533–539.

8. Bortolotto E, Bucklin A, Mezzavilla M, Zane L, Patarnello T. Gone with the currents: lack of genetic differentiation at the circum-continental scale in the Antarctic krill *Euphausia superba*. *BMC Genet.* 2011; 12: 32. doi: [10.1186/1471-2156-12-32](https://doi.org/10.1186/1471-2156-12-32) PMID: [21486439](https://pubmed.ncbi.nlm.nih.gov/21486439/)
9. De Pittà C, Bertolucci C, Mazzotta GM, Bernante F, Rizzo G, De Nardi B, et al. Systematic sequencing of mRNA from the Antarctic krill (*Euphausia superba*) and first tissue specific transcriptional signature. *BMC Genomics.* 2008; 9: 45. doi: [10.1186/1471-2164-9-45](https://doi.org/10.1186/1471-2164-9-45) PMID: [18226200](https://pubmed.ncbi.nlm.nih.gov/18226200/)
10. Seear PJ, Tarling GA, Burns G, Goodall-Copestake WP, Gaten E, Ozkaya O, et al. Differential gene expression during the moult cycle of Antarctic krill (*Euphausia superba*). *BMC Genomics.* 2010; 11: 582. doi: [10.1186/1471-2164-11-582](https://doi.org/10.1186/1471-2164-11-582) PMID: [20958982](https://pubmed.ncbi.nlm.nih.gov/20958982/)
11. Meyer B, Martini P, Biscontin A, De Pittà C, Romualdi C, Teschke M, et al. Pyrosequencing and *de novo* assembly of Antarctic krill (*Euphausia superba*) transcriptome to study the adaptability of krill to climate-induced environmental changes. *Mol Ecol Resour.* 2015; 15: 1460–1471. doi: [10.1111/1755-0998.12408](https://doi.org/10.1111/1755-0998.12408) PMID: [25818178](https://pubmed.ncbi.nlm.nih.gov/25818178/)
12. Biscontin A, Frigato E, Sales G, Mazzotta GM, Teschke M, De Pittà C, et al. The opsin repertoire of the Antarctic krill *Euphausia superba*. *Mar Genomics.* 2016; 29: 61–68. doi: [10.1016/j.margen.2016.04.010](https://doi.org/10.1016/j.margen.2016.04.010) PMID: [27157882](https://pubmed.ncbi.nlm.nih.gov/27157882/)
13. Jeffery NW. The first genome size estimates for six species of krill (Malacostraca, Euphausiidae): large genomes at the north and south poles. *Polar Biol.* 2011; 35: 959–962.
14. i5K Consortium. The i5K Initiative: Advancing Arthropod Genomics for Knowledge, Human Health, Agriculture, and the Environment. *J Hered.* 2013; 104: 595–600. doi: [10.1093/jhered/est050](https://doi.org/10.1093/jhered/est050) PMID: [23940263](https://pubmed.ncbi.nlm.nih.gov/23940263/)
15. Jarman S, Deagle BE. *Genetics of Antarctic Krill. Biology and Ecology of Antarctic Krill*, Springer 2016;
16. Cahais V, Gayral P, Tsagkogeorga G, Melo-Ferreira J, Ballenghien M, Weinert L, et al. Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Mol Ecol Resour.* 2012; 12: 834–845. doi: [10.1111/j.1755-0998.2012.03148.x](https://doi.org/10.1111/j.1755-0998.2012.03148.x) PMID: [22540679](https://pubmed.ncbi.nlm.nih.gov/22540679/)
17. Clark MS, Thorne MAS, Toullec J-Y, Meng Y, Guan LL, Peck LS, et al. Antarctic krill 454 pyrosequencing reveals chaperone and stress transcriptome. *PLoS One.* 2011; 6: e15919. doi: [10.1371/journal.pone.0015919](https://doi.org/10.1371/journal.pone.0015919) PMID: [21253607](https://pubmed.ncbi.nlm.nih.gov/21253607/)
18. Martins MJ, Lago-Leston A, Anjos A, Duarte CM, Agusti S, Serrão EA, Pearson GA. A transcriptome resource for Antarctic krill (*Euphausia superba* Dana) exposed to short-term stress. *Mar Genomics.* 2015; 23:45–47. doi: [10.1016/j.margen.2015.04.008](https://doi.org/10.1016/j.margen.2015.04.008) PMID: [25957695](https://pubmed.ncbi.nlm.nih.gov/25957695/)
19. De Pittà C, Biscontin A, Albiero A, Sales G, Millino C, Mazzotta GM, et al. The Antarctic krill *Euphausia superba* shows diurnal cycles of transcription under natural conditions. *PLoS One.* 2013; 8: e68652. doi: [10.1371/journal.pone.0068652](https://doi.org/10.1371/journal.pone.0068652) PMID: [23874706](https://pubmed.ncbi.nlm.nih.gov/23874706/)
20. Jia Z., Virtue P., Swadling K.M. A photographic documentation of the development of Antarctic krill (*Euphausia superba*) from egg to early juvenile *Polar Biol.* 2014; 37: 165.
21. Ikeda T. Development of the larvae of the Antarctic krill (*Euphausia superba* Dana) observed in the laboratory. *Journal of Experimental Marine Biology and Ecology.* 1984; 107–117.
22. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinforma Oxf Engl.* 2014; 30: 2114–2120.
23. Schulz MH, Zerbino DR, Vingron M, Birney E. Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinforma Oxf Engl.* 2012; 28: 1086–1092.
24. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011; 29: 644–652. doi: [10.1038/nbt.1883](https://doi.org/10.1038/nbt.1883) PMID: [21572440](https://pubmed.ncbi.nlm.nih.gov/21572440/)
25. Peng Y, Leung HCM, Yiu S-M, Lv M-J, Zhu X-G, Chin FYL. IDBA-tran: a more robust *de novo* de Bruijn graph assembler for transcriptomes with uneven expression levels. *Bioinforma Oxf Engl.* 2013; 29: i326–334.
26. Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, et al. SOAPdenovo-Trans: *de novo* transcriptome assembly with short RNA-Seq reads. *Bioinforma Oxf Engl.* 2014; 30: 1660–1666.
27. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 2016; 44: D7–19. doi: [10.1093/nar/gkv1290](https://doi.org/10.1093/nar/gkv1290) PMID: [26615191](https://pubmed.ncbi.nlm.nih.gov/26615191/)
28. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* 2015; 43: D204–212. doi: [10.1093/nar/gku989](https://doi.org/10.1093/nar/gku989) PMID: [25348405](https://pubmed.ncbi.nlm.nih.gov/25348405/)
29. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinforma Oxf Engl.* 2014; 30: 1236–1240.
30. Nakasugi K, Crowhurst R, Bally J, Waterhouse P. Combining transcriptome assemblies from multiple *de novo* assemblers in the allo-tetraploid plant *Nicotiana benthamiana*. *PLoS One.* 2014; 9: e91776. doi: [10.1371/journal.pone.0091776](https://doi.org/10.1371/journal.pone.0091776) PMID: [24614631](https://pubmed.ncbi.nlm.nih.gov/24614631/)

31. Chagné D, Crowhurst RN, Pindo M, Thrimawithana A, Deng C, et al. The draft genome sequence of European pear (*Pyrus communis* L. 'Bartlett'). PLoS One. 2014 Apr 3; 9: e92644. doi: [10.1371/journal.pone.0092644](https://doi.org/10.1371/journal.pone.0092644) PMID: [24699266](https://pubmed.ncbi.nlm.nih.gov/24699266/)
32. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011; 12: 323. doi: [10.1186/1471-2105-12-323](https://doi.org/10.1186/1471-2105-12-323) PMID: [21816040](https://pubmed.ncbi.nlm.nih.gov/21816040/)