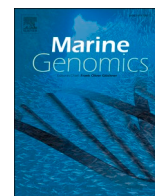


Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

## Marine Genomics

journal homepage: [www.elsevier.com/locate/margen](https://www.elsevier.com/locate/margen)

## De novo transcriptome assembly of the Southern Ocean copepod *Rhincalanus gigas* sheds light on developmental changes in gene expression

Cory A. Berger<sup>a,b</sup>, Deborah K. Steinberg<sup>c</sup>, Nancy J. Copley<sup>a</sup>, Ann M. Tarrant<sup>a,\*</sup><sup>a</sup> Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, United States<sup>b</sup> MIT-WHOI Joint Program in Oceanography/Applied Ocean Science & Engineering, Cambridge and Woods Hole, MA, USA<sup>c</sup> Virginia Institute of Marine Science, William & Mary, Gloucester Pt, VA 23062, United States

## ARTICLE INFO

## Keywords:

Southern Ocean  
Zooplankton  
Lipid metabolism  
Development  
Molting

## ABSTRACT

Copepods are small crustaceans that dominate most zooplankton communities in terms of both abundance and biomass. In the polar oceans, a subset of large lipid-storing copepods occupy central positions in the food web because of their important role in linking phytoplankton and microzooplankton with higher trophic levels. In this paper, we generated a high-quality de novo transcriptome for *Rhincalanus gigas*, the largest—and among the most abundant—of the Southern Ocean copepods. We then conducted transcriptional profiling to characterize the developmental transition between late-stage juveniles and adult females. We found that juvenile *R. gigas* substantially upregulate lipid synthesis and glycolysis pathways relative to females, as part of a developmental gene expression program that also implicates processes such as muscle growth, chitin formation, and ion transport. This study provides the first transcriptional profile of a developmental transition within *Rhincalanus gigas* or any endemic Southern Ocean copepod, thereby extending our understanding of copepod molecular physiology.

## 1. Introduction

The copepod *Rhincalanus gigas* (Fig. 1) is one of the dominant members of the mesozooplankton of the Southern Ocean, numerically and in terms of biomass (Ommanney, 1936; Mackey et al., 2012). It is also by far the largest of the Southern Ocean copepods, reaching more than 10 mm in length (Jansen et al., 2006). *R. gigas* is most abundant at the Polar Front where it is associated with the Antarctic Circumpolar Current, but its range extends south to the Antarctic continental shelf (Ward et al., 1997; Gleiber, 2014). Like the other dominant large-bodied copepods of the Southern Ocean—*Calanoides acutus* and *Calanus propinquus*—*R. gigas* has a circumpolar distribution and is abundant in both pelagic and coastal waters (Conover and Huntley, 1991). The diet of *R. gigas* is mostly herbivorous, but omnivory is an important aspect of its overwintering strategy when phytoplankton are scarce (Graeve et al., 1994; Kattner et al., 1994; Pasternak and Schnack-Schiel, 2002; Schmidt et al., 2003). *R. gigas* is a “slow and steady” species, characterized by low rates of growth, feeding, and metabolism (Conover and Huntley, 1991; Atkinson et al., 1992; Mayzaud et al., 2002; Maps et al., 2014), low fecundity (Ward and Shreeve, 1995), and relatively small seasonal fluctuations in abundance (Atkinson, 1991; Ward et al., 1997). Its large size and opportunistic feeding mode allow *R. gigas* to survive periods of

low food availability. Compared to *C. acutus*, *R. gigas* exhibits low winter mortality (Ward et al., 1997) and small seasonal changes in lipid content, losing only about half as much relative dry weight and lipids over the winter (Kattner et al., 1994; Hagen and Schnack-Schiel, 1996; Tarling et al., 2007).

Historically, there has been disagreement whether *R. gigas* has a one- (Ommanney, 1936) or two-year life cycle (Atkinson et al., 1992). In fact, like *Calanus finmarchicus* in the North Atlantic, *R. gigas* can almost certainly exhibit either a one- or two-year life cycle depending on the location and environmental conditions. Both field observations (Ommanney, 1936; Voronina, 1970; Marin, 1988; Chiba, 2002) and modelling studies (Tarling et al., 2007) indicate that populations are frequently composed of overlapping generations with both single- and multi-year lifespans. Towards the southern end of its distribution, the waters are probably too cold for *R. gigas* to complete its life cycle in one year, but its large size allows young juvenile (copepodite) stages to successfully overwinter and complete the life cycle during the following growing season (Ward et al., 1997; Maps et al., 2014).

Lipid storage is a critical strategy for zooplankton in highly seasonal polar environments. Several groups of copepods, particularly predominantly herbivorous copepods at high latitudes and in coastal upwelling zones, store large amounts of lipids that they use as fuel for reproduction

\* Corresponding author.

E-mail address: [atarrant@whoi.edu](mailto:atarrant@whoi.edu) (A.M. Tarrant).<https://doi.org/10.1016/j.margen.2021.100835>

Received 14 October 2020; Received in revised form 30 December 2020; Accepted 30 December 2020

1874-7787/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

and overwintering (Hagen and Schnack-Schiel, 1996; Lee et al., 2006). The overwintering strategy of these lipid-storing copepods frequently includes a period of seasonal dormancy called diapause, which takes place during a late juvenile or adult stage of development. During diapause, copepods descend to depth, arrest development, and become metabolically inactive; they ultimately emerge from diapause during the spring to molt (in the case of juveniles) and reproduce (reviewed in Baumgartner and Tarrant, 2017). It is unclear whether, or to what extent, *R. gigas* performs diapause. In contrast to the canonical diapause behavior of species heavily reliant on herbivory, such as the sympatric *Calanoides acutus*, omnivory allows *R. gigas* to remain active or semi-active throughout the winter. Although it accumulates large lipid stores and migrates seasonally to deeper waters (Schnack-Schiel and Hagen, 1994), it is capable of actively feeding and reproducing year-round (Rakusa-Suszczewski et al., 1976; Pasternak and Schnack-Schiel, 2002).

Like most of the large lipid-storing species in the families Calanidae, Neocalanidae, and Eucalanidae, the primary storage lipids of *R. gigas* are wax esters. However, these lipids contain mostly short saturated fatty acids and alcohols (Kattner et al., 1994; Ward et al., 1996), in contrast to the long-chain monounsaturated lipids that characterize predominantly herbivorous, diapause-associated copepods—such as *C. acutus*, and the Arctic species *Calanus glacialis* and *Calanus hyperboreus* (Kattner et al., 1994; Atkinson, 1998; Lee et al., 2006). Long-chain wax esters are more energy-efficient and are thought to be an adaptation for long periods of food deprivation (Lee et al., 1974), so the predominance of short-chain wax esters in *R. gigas* may be a biochemical signature of its omnivorous feeding mode and weak reliance on diapause (Kattner et al., 1994). These interspecific differences in lipid composition require differences in lipid synthesis and elongation pathways, but the genetic basis of these differences has yet to be closely examined in copepods.

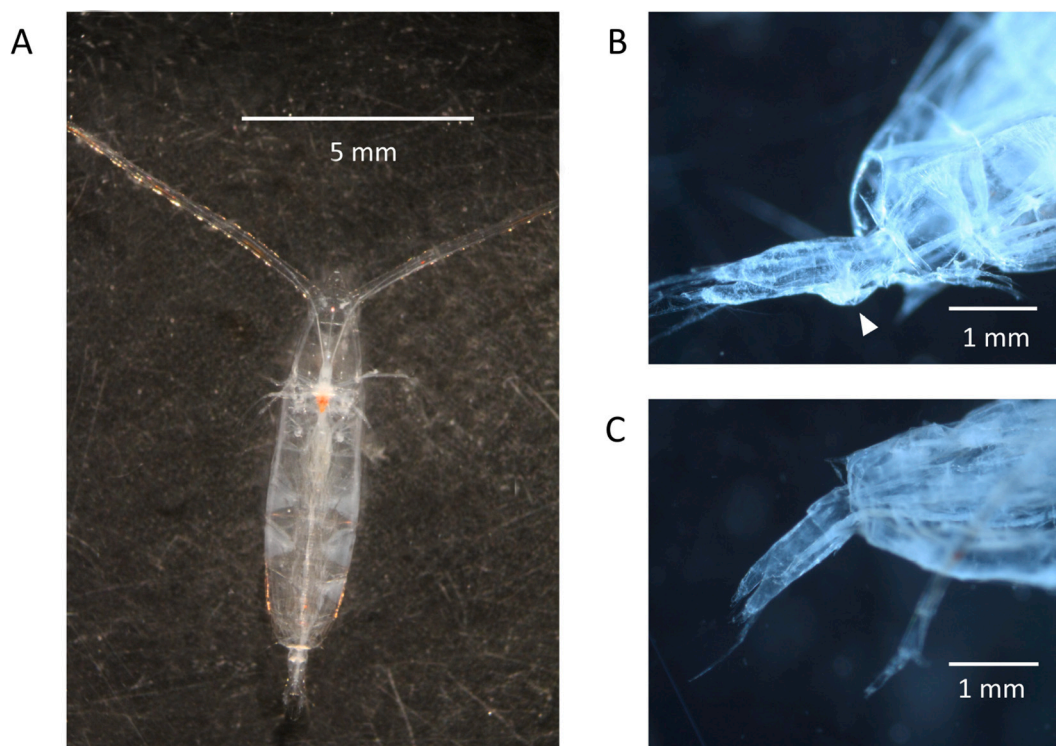
We report a highly-complete transcriptome for *R. gigas*, assembled from two life stages, to serve as a bioinformatic resource, and we investigated changes in gene expression between late stage juveniles (CV

copepodites) and adult females. Calanoid copepods progress through six naupliar (NI-NVI) and five copepodite (CI-CV) larval stages via a series of molts, which are accompanied by metabolic and physiological transitions and, in the case of the terminal molt, the formation of mature gonads (Mauchline, 1998). The developmental transition between late juveniles and adults has previously been profiled in a diapausing species (*C. finmarchicus*), where the terminal molt coincides with emergence from diapause. In *C. finmarchicus*, there is a clear connection between lipid metabolism and development, and transcriptomic studies identified large shifts in the expression of lipid synthesis and degradation enzymes between adults and juveniles (Lenz et al., 2014; Skottene et al., 2019). In contrast, the extent to which changes in lipid and carbohydrate metabolism are associated with development of *R. gigas* was unknown, so we specifically assessed changes in expression within these pathways during maturation to adulthood. Indeed, we found that *R. gigas* undergoes strong shifts in lipid synthesis and carbohydrate metabolism pathways during maturation, similar to previous observations in *C. finmarchicus*. This study opens the door for comparative analyses of copepod physiology and development.

## 2. Materials and methods

### 2.1. Animal collection

*Rhincalanus gigas* were collected during austral summer 2019 aboard the “ARSV Laurence M. Gould” in association with sampling conducted by the Palmer Antarctica Long-Term Ecological Research (PAL LTER) program. The collection site was on the shelf of the West Antarctic Peninsula (LTER grid Station 600.040; 64°56.04’S, 64°23.58’W). Copepods were collected at 18:25 on January 7, 2019 by conducting a double oblique net tow to 300 m with a 2 × 2 m rectangular frame, a 700- $\mu$ m mesh net, and a non-filtering cod end. The surface temperature was 1.5°C, and the bottom depth was 1422 m. *R. gigas* individuals were contained in an ice-chilled bucket during sorting. Copepods were stored



**Fig. 1.** a) Adult *Rhincalanus gigas* female (photo provided by Miram Gleiber). For clarity, the background was edited to remove large scratches. b) Close-up of urosome of adult female, with arrow showing enlarged genital segment (first urosomal segment) that contains a genital pore. c) Close-up of urosome of CV juvenile, without enlarged genital segment. a) Live image; b) and c) preserved in RNAlater.

in RNAlater at  $-20^{\circ}\text{C}$  or colder and shipped to Woods Hole Oceanographic Institution (WHOI) on dry ice. Staging was conducted in Woods Hole back at WHOI by examining the preserved individuals under a stereomicroscope, examining urosomes for signs of maturity by shape and presence of genital pores (Fig. 1), and measuring the prosome length (females  $>6$  mm) (Seret, 1979; Razouls, 1994).

## 2.2. RNA extraction, library preparation, and sequencing

Total RNA was extracted from individual copepods using the Aurum Fatty and Fibrous Tissue Kit (Bio-Rad). The optional DNase treatment was omitted because we have empirically found that omitting the step can improve RNA quality and yield without observable DNA contamination (for further discussion see Tarrant et al., 2019). RNA yield, purity, and quality were assessed using a Nanodrop Spectrophotometer and denaturing agarose gel. Because the 28S rRNA in arthropod samples frequently fragments during sample preparation (McCarthy et al., 2015), samples were deemed to be of suitable quality if they possessed a distinct 18S band with no evidence of genomic DNA contamination (large bands) or degradation (smear of smaller bands). RNA was submitted to Arraystar Inc. (Rockville, MD) for library construction and sequencing. Libraries were constructed from six individual copepods using the KAPA Stranded RNA-seq Library Preparation Kit. Library quality was assessed using an Agilent 2100 Bioanalyzer, and yield was assessed through quantitative PCR. All libraries showed the expected fragment size distribution (400–600 bp) without evidence of adapter dimer contamination. The barcoded libraries were sequenced with 150 base pair paired-end reads to a depth of 40 M reads (20 M pairs) on an Illumina HiSeq 4000. The MiXs (Minimum Information about any Sequence) descriptions, which provide a standardized checklist for sequencing metadata (Yilmaz et al., 2011), are given in Table 1. To further confirm species identity, mtCOI sequences from each individual library were BLASTed against the nr database; all sequences were  $>99\%$  identical to a deposited NCBI *R. gigas* mtCOI sequence (GenBank JN663361.1).

## 2.3. Sequence quality control, assembly and annotation

Sequence quality was checked using FastQC v0.11.7 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Raw sequences were error-corrected using RCorrector v1.0.4 (Song and Florea, 2015), and adapter and quality trimming were performed with TrimGalore v0.6.5 (<https://github.com/FelixKrueger/TrimGalore>) with a Phred score cutoff of 5 and a minimum length cutoff of 50. Ribosomal RNA (rRNA)

**Table 1**  
MiXs descriptors.

Item name	Description
Investigation type	Eukaryote
Organism	<i>Rhincalanus gigas</i>
Project name	De novo transcriptome assembly of the Southern Ocean copepod <i>Rhincalanus gigas</i>
Geographic location (latitude and longitude)	64.934 S 64.393 W
Geographic location (country and/or sea, region)	West Antarctic Peninsula continental shelf
Geographic location (depth)	0-300 m
Collection date	2019-01-07
Developmental stage	Adult and CV copepodite
Sample type	whole body
Environment (biome)	ENVO:01000032 (neritic pelagic zone)
Environment (feature)	ENVO:00000206 (marine neritic zone)
Environment (material)	ENVO:00002006 (water)
Sequencing method	Illumina HiSeq
Data accessibility	BioProject: PRJNA666170 BioSamples: SAMN16282538 - SAMN16282543 Raw read data: SRR12732903 - SRR12732908 TSA: GIVD00000000

was removed with SortMeRNA v2.1b (Kopylova et al., 2012). Six individual libraries were then assembled using Trinity v2.9.1 (Grabherr et al., 2011). The resulting assemblies were combined using TransFuse v0.5.0 (<https://github.com/cbournell/transfuse>). For comparison, an assembly was also constructed using the default Trinity workflow in which quality-filtered reads from all libraries are merged prior to assembly. At this point, the quality of both assemblies was assessed with TransRate v1.0.1 (Smith-Unna et al., 2016). TransRate calculates a cutoff for contig quality that maximizes the overall assembly score, so we retained only the “good” contigs that pass this cutoff. Finally, assemblies were clustered using CD-HIT-EST v4.7 (Li and Godzik, 2006) at 98% sequence similarity and an alignment bandwidth of three ( $-c$  0.98 -b 3). Assembly quality was again assessed with TransRate and these are the statistics reported in Table 2. Completeness was assessed with BUSCO v4.0.5 (Simão et al., 2015) using the arthropoda\_odb9 database.

Sequences were annotated with the Trinotate pipeline (Bryant et al., 2017). Peptide sequences were predicted with TransDecoder v5.5.0 (<https://github.com/TransDecoder/TransDecoder>). BLASTX and BLASTP v2.7.1 (Altschul et al., 1990) were used to search transcripts and peptides against the Swiss-Prot database (Boutet et al., 2007, downloaded March 18, 2020), and hmmscan (<http://hmmer.org/>) was used to search the PFAM database (Finn et al., 2014, downloaded March 16, 2020). Any peptides with a hit (e-value =  $1e-10$ ) to either SwissProt or PFAM were retained in addition to computationally predicted peptides. Sequence whose top five BLASTP hits by e-value contained a non-metazoan sequence and no metazoan sequences were considered contamination and removed. We also used DIAMOND v0.9.30 (Buchfink et al., 2015) with the ‘more-sensitive’ setting to search proteins and nucleotides against NCBI’s non-redundant protein database (nr, downloaded March 23, 2019).

We used the “shuffle” mode of FEELnc v0.1.1 (Wucher et al., 2017) to annotate long non-coding RNA’s (lncRNA’s). This is a machine learning approach that can be used to predict lncRNA’s in de novo assemblies. The training set for coding transcripts was constructed from the set of complete universal single copy orthologs found from the BUSCO analysis (as in Huerlimann et al., 2018).

## 2.4. Mapping and differential gene expression analysis

Transcript quantification was performed with Salmon v1.1.0 (Patro et al., 2017) using the ‘gcBias’, ‘seqBias’, and ‘validateMappings’ flags. Transcripts were clustered using Corset v1.09 (Davidson and Oshlack, 2014) with default settings, and counts were summarized to the gene level using the R package ‘tximport’ v1.10.1 (Soneson et al., 2016). Differential expression analysis was performed with DESeq2 v1.22.2 (Love et al., 2014). Gene ontology enrichment was performed with the Fisher’s exact test function of the GO MWU R package (Wright et al., 2015). We then used the REVIGO web server (<http://revigo.irb.hr/>) (Supek et al., 2011) to visualize and summarize significant GO terms (adj.  $p$ -value  $< 0.05$ ). All statistical analyses were performed within RStudio (RStudio Team, 2020). For visualization, we used the ‘plotPCA’ function of DESeq2 as well as the R packages ‘pheatmap’ (<https://cran.r-project.org/web/packages/pheatmap/index.html>) and ‘Rtsne’ (<https://github.com/jkrijthe/Rtsne>). Since we only had six samples, we used a perplexity parameter of 1.6 with Rtsne.

## 2.5. Annotation and gene set analysis

Because automatic annotation of non-model organisms is not always reliable, we manually annotated certain KEGG (Kanehisa and Goto, 2000) metabolic pathways of particular interest (as described in Section 3.5) using BLAST and a gene tree-based approach. For each KEGG gene ID, we aligned sequences with MAFFT (Katoh et al., 2009), constructed gene trees using iQTREE (Nguyen et al., 2015), and visualized trees with FigTree (<http://tree.bio.ed.ac.uk/software/figtree>). Sequences were annotated based on their top BLAST hits and their placement in the gene



**Table 2**  
*R. gigas* assembly statistics.

	Trans rate score	Mapping rate	Assembly contigs	N50	BUSCO
Full assembly	0.44049	89.1%	265,173	1344	C:93.7% [D:80.6%], F:2.0%, M:4.3%, n:1013
Default assembly	0.29849	80.1%	152,092	1138	C:91.3% [D:49.8%], F:2.8%, M:5.9%, n:1013

tree. Gene set enrichment analysis (GSEA) was performed on manually annotated KEGG pathways in the R package ‘clusterProfiler’ (Yu et al., 2012). As input, we ordered genes by shrunken log-fold change and used all genes annotated for a particular pathway as a custom gene set. A gene set was considered enriched at an FDR of 0.05.

### 3. Results

#### 3.1. Sequence read data

Three field-collected individuals from each of two developmental stages (CV copepodite and adult female, six individuals total) were sequenced to an average depth of  $18.5 \pm 2.2$  M (mean  $\pm$  SD) read pairs per sample, generating 111.1 M total reads. K-mer correction, quality trimming, and rRNA removal removed 17.6% of the raw reads for a final set of 91.5 M reads in total and  $15.2 \pm 2.4$  M (mean  $\pm$  SD) read pairs per sample. GC content of the final assembly was 39.7%.

#### 3.2. Transcriptome assembly and quality control

We used a non-standard transcriptome assembly workflow in which the transcriptomes for each of the six individuals were assembled individually and subsequently merged. This follows a similar approach used by Roncalli et al. (2018) to successfully assemble another recent copepod transcriptome; in that study, this strategy was necessary to obtain a high-quality assembly. To validate this approach in our case, we compared our assembly to an alternative transcriptome generated from the same libraries using the default Trinity pipeline, in which all reads are initially merged (“Default assembly”, Table 2). The final non-standard assembly contained 265,173 contigs and had a TransRate score of 0.44, with 89.1% of all reads mapping back to the transcriptome. BUSCO results indicate a relatively complete assembly with 93.7% of single-copy arthropod orthologs recovered, although many of these were duplicated (C:93.7%[S:13.1%, D:80.6%], F:2.0%, M:4.3%, n:1013). The default Trinity assembly contained a smaller number of sequences (152,092 versus 265,173) but had a lower TransRate score (0.30 versus 0.44). The mapping rate of the default assembly was lower at 80%, and BUSCO scores for the default assembly were also slightly lower than for our final assembly, albeit with lower levels of duplication. For these reasons, we continued to use the non-standard assembly for downstream analysis.

#### 3.3. Annotation

TransDecoder predicted peptides for 155,948 (58.8%) transcripts in the full assembly. We removed 2860 non-metazoan sequences, retaining a filtered set of 262,313 contigs for downstream analysis. Of the filtered contigs, 188,349 had at least ten total counts and were retained by Corset, which grouped them into 89,528 clusters; 62,148 (69.4%) Corset clusters had at least one predicted peptide and 39,361 (44%) had at least one BLASTP hit to the NCBI non-redundant database (nr) with a threshold e-value of  $1e-10$ . Based on these BLAST results, 34,624 (38.7%) clusters received at least one gene ontology annotation, and 27,298 (30.5%) were annotated with at least one KEGG term.

#### 3.4. Differential gene expression analysis

We sequenced six *R. gigas* individuals, three of which were identified as CV copepodites and three as adult females based on microscopic

examination. Images of *R. gigas* are shown in Fig. 1. However, principal component and distance-based clustering analyses identified one juvenile (“RJ2”) that strongly grouped with the adult animals (Fig. 2). This grouping was also apparent using an alternative dimensionality-reduction approach, t-SNE (van der Maaten and Hinton, 2008). It is possible that the RJ2 sample was misidentified as a juvenile and is actually another adult female, or that it is a late-stage CV close to its terminal molt, which could explain its greater transcriptional similarity to adults. This individual was also the largest of the juvenile samples. To avoid any ambiguity, we excluded this sample from subsequent analyses, thus comparing three adults and two juveniles. DE analysis without RJ2 identified 5198 (5.9%) genes with higher expression in juveniles and 1234 (1.4%) with higher expression in adults (FDR  $\leq$  0.05, LFC  $\geq$  1).

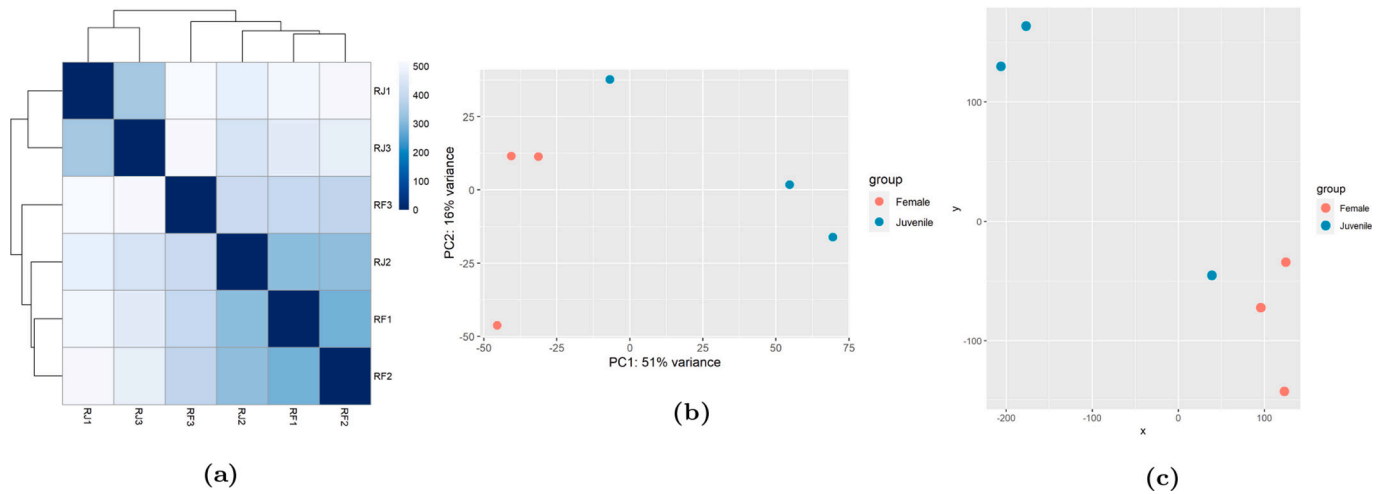
In order to see which of our results were robust to including the outlying sample, we also performed DE and GO enrichment analysis including the outlier. DE analysis using all samples identified 1344 (1.5%) genes with higher expression in juveniles and 353 (0.39%) with higher expression in adults. Complete differential expression results and gene annotations are given in Supplementary File 1.

#### 3.5. GO enrichment and KEGG pathway analysis

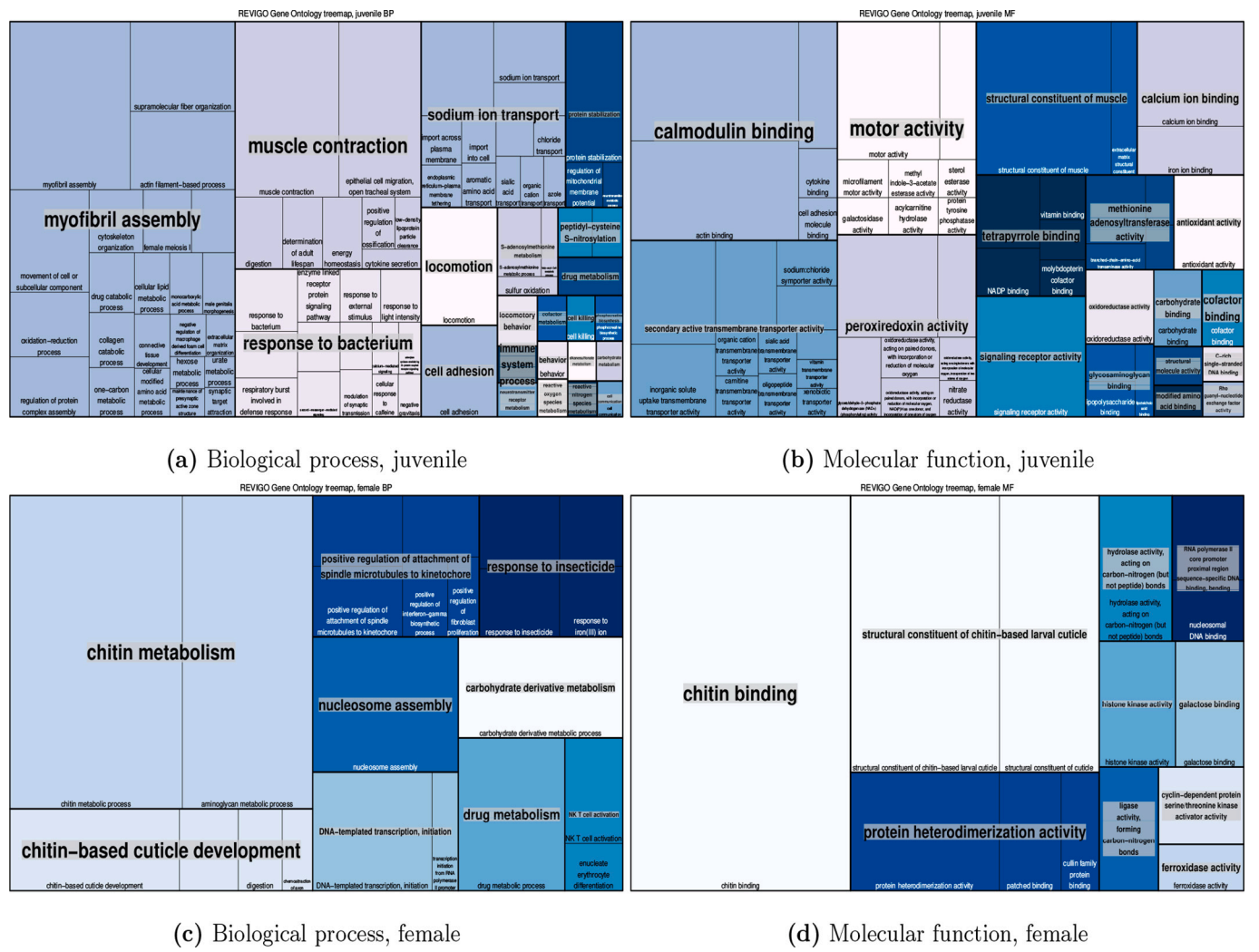
After excluding the outlier, we identified 449 gene ontology (GO) categories enriched among genes with higher expression in juveniles (263 BP, 67 CC, 119 MF), and 70 GO categories enriched among genes with higher expression in females (41 BP, 15 CC, 14 MF). Including the outlier, we found a total of 110 GO categories enriched among genes with higher expression in juveniles (70 BP, 10 CC, 30 MF), and 11 GO categories enriched among genes with higher expression in females (4 BP, 4 CC, and 3 MF). The GO terms with strongest statistical significance (lowest adjusted *p*-value) were identical or similar in both cases (Supplementary File 2), so we only discuss the results excluding the outlier in the rest of the paper. Complete GO enrichment results are given in Supplementary File 2.

We used REVIGO to interpret and visualize the large number of GO categories, and the results are summarized in Fig. 3. After clustering with REVIGO, the top five biological process (BP) terms by *p*-value with higher expression in juveniles were “myofibril assembly”, “muscle contraction”, “epithelial cell migration, open tracheal system”, “supramolecular fiber organization”, and “locomotion”; the top five molecular function (MF) terms were “structural constituent of muscle”, “calmodulin binding”, “actin binding”, “motor activity”, and “signaling receptor activity”. The top five BP terms with higher expression in females were “chitin metabolic process”, “aminoglycan metabolic process”, “nucleosome assembly”, “carbohydrate derivative metabolic process”, and “drug metabolic process”; the top five MF terms were “chitin binding”, “structural constituent of chitin-based larval cuticle”, “structural constituent of cuticle”, “protein heterodimerization activity”, and “hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds”.

Because automated BLAST-based pipelines typically result in incomplete annotation (e.g., Tarrant et al., 2016; Skottene et al., 2019), we manually annotated genes in KEGG pathways related to lipid and carbohydrate metabolism (glycolysis, fatty acid synthesis, fatty acid elongation, fatty acid degradation, the pentose phosphate pathway, glycerophospholipid metabolism, and the TCA cycle) as well as steroid biosynthesis. We then performed gene set enrichment analysis on the manually annotated pathways. Three of the eight pathways were enriched among genes upregulated in juveniles: glycolysis, fatty acid



**Fig. 2.** Clustering analyses identify an additional putative female sample. Group labels (“Female” and “Juvenile”) were based on microscopy. a) Heatmap of distance matrix based on variance stabilizing transformed counts, implemented in the R package ‘pheatmap’. Samples labelled “RF” were identified as female and “RJ” samples were identified as juvenile. b) Principal component analysis based on the 500 genes exhibiting the highest variance across all samples, implemented in ‘DESeq2’. c) t-SNE of normalized counts for all genes, implemented in the R package ‘Rtsne’.



**Fig. 3.** Summary of gene ontology enrichment analysis. Top: GO categories upregulated in juveniles. Bottom: GO categories upregulated in females. Generated using REVIGO (<http://revigo.irb.hr/>) “TreeMap” script. Each colored rectangle is a cluster of one or more semantically-related categories. Rectangle size is inversely proportional to the p-value for that category.

synthesis, and glycerophospholipid metabolism (Fig. 4). These results broadly agree with the GO enrichment analysis, as GO terms such as “lipid metabolic process”, “lipid transport”, “calcium-dependent phospholipid binding”, and “carbohydrate catabolic process” were upregulated in juveniles. For the list of genes annotated with these KEGG pathways, see Supplementary File 3.

Lenz et al. (2014) identified three lipid metabolism genes with differential expression between CV copepodite and adult female *Calanus finmarchicus*. Two of these genes, an elongation of very long-chain fatty acids protein (*Elov*) and a delta-9 desaturase (*Scd*), had much higher expression in *C. finmarchicus* CV juveniles compared to adult females. The *C. finmarchicus Elov* gene was the best BLASTN hit for two *R. gigas* transcripts, and neither was differentially expressed in our dataset. However, three other *R. gigas Elov* genes were upregulated in juveniles, suggesting that upregulation of enzymes in the *Elov* family may be a general feature of late-stage copepodite development, even if the specific genes involved differ between species. The *C. finmarchicus Scd* gene had no significant similarity to any transcripts in the *R. gigas* assembly (BLASTN search with an e-value cutoff of  $1e-5$ ), but five other *Scd*-like genes were upregulated in juvenile *R. gigas*, again suggesting that developmental similarities may be found within gene families but not necessarily with the same orthologs. The third gene identified by Lenz et al. (2014), a diacylglycerol o-acyltransferase (*Dgat*), also had no significant similarity to any *R. gigas* transcripts, and we did not identify any *Dgat* enzymes with differential expression in *R. gigas*.

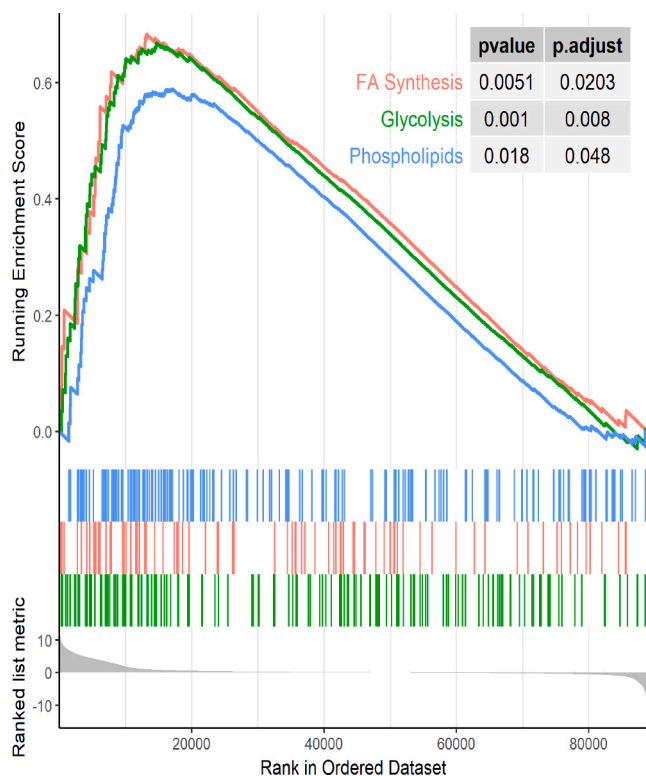


Fig. 4. Gene set enrichment analysis of selected KEGG pathways. Curves show running enrichment scores for each pathway (top), as calculated in the R package ‘clusterProfiler’. Enrichment scores quantify enrichment of genes in a pathway relative to the ranked list of all genes ordered by log-fold change (bottom). Vertical lines (middle) mark the positions of pathway genes in the ranked list. Genes on the left of this figure were upregulated in juveniles. ‘pvalue’ indicates unadjusted p-value, ‘p.adjust’ indicates p-values after multiple testing correction. Plotted with the R package ‘enrichplot’. The list of genes in each gene set is given in Table S3. FA: fatty acid; Phospholipids: glycerophospholipid metabolism.

### 3.6. Identification of long non-coding RNA’s

We used FEELnc to annotate potential long non-coding RNA’s (lncRNA’s) in our transcriptome. We used the assembly’s complete BUSCO orthologs as the training set for classification of coding and non-coding transcripts. FEELnc annotated 99,375 out of 262,313 transcripts as lncRNA’s (37.9%). Of these, 14,645 transcripts had at least one hit to the nr protein database (e-value cutoff  $1e-5$ ) and were considered false positives, leaving 84,730 (32.3%) putative lncRNA’s. This 85% accuracy of lncRNA annotation is similar to the error rate reported by the authors of FEELnc (Wucher et al., 2017) and in another recent crustacean transcriptome (Huerlimann et al., 2018). Of the 188,349 contigs included in Corset clusters, 50,293 (26.7%) were annotated as lncRNA’s, and 22,361 out of 89,528 clusters (25.0%) consisted entirely of predicted lncRNA transcripts. Consistent with the fact that lncRNA’s generally have low expression (Ilott and Ponting, 2013), 41% of lncRNA’s fell below Corset’s expression filter compared to 22% of mRNA transcripts. Indeed, the mean normalized counts value for lncRNA transcripts was 24.6, compared to 160.7 for mRNA transcripts. In total, 634 lncRNA clusters were differentially expressed between adults and juveniles (~9% of all DE genes), with 272 more highly expressed in adults and 362 more highly expressed in juveniles. Overall, a similar proportion of our assembly was annotated as lncRNA compared to the recent crustacean transcriptome assembled by Huerlimann et al. (2018).

## 4. Discussion

Despite the global abundance and ecological importance of marine copepods, the development of transcriptomic resources for these taxa has been somewhat limited (reviewed in Tarrant et al. (2019); see Asai et al. (2020) and Russo et al. (2020) for other recent copepod transcriptomes). The current study provides a high-quality transcriptome for the important Southern Ocean endemic *Rhincalanus gigas*. Another assembly for *R. gigas* was recently published by Lauritano et al. (2020). Their assembly is derived only from adult individuals, whereas the transcriptome reported here was assembled from more than one life stage and is thus considerably more complete (94% versus 70% BUSCO). Lauritano et al. (2020) focused on the identification of antifreeze proteins and glutathione metabolism enzymes in adult copepods. On the other hand, our study profiles differences in gene expression between two developmental stages (copepods and adult females), highlighting that juveniles upregulate lipid synthesis and glycolysis pathways compared to adults.

### 4.1. Transition between juvenile and adult is accompanied by a large-scale transcriptional shift

In this study, we observed extensive changes in gene expression across the terminal molt between juvenile (CV copepodite) and adult *Rhincalanus gigas* copepods. Relative to adult females, we found that juvenile *R. gigas* substantially upregulate muscle-related genes, including functions and processes such as actin binding, myofibril assembly, and muscle contraction (Fig. 3a, b). This likely reflects preparation for molt-associated muscle remodelling. Crustacean molting is associated with cycles of muscle atrophy and growth, and pre-molt periods are often characterized by enhanced muscle growth and muscle-related gene expression (Mykles and Skinner, 1982; Kuballa et al., 2011). A previous study of copepod development observed large changes in myosin expression within the CV copepodite stage of *Calanus finmarchicus* (Tarrant et al., 2014).

This developmental gene expression program involves substantial changes to cell signalling, as juveniles upregulate ion transport, transmembrane transporter activity, and signalling receptor activity. Regulation of ion balance is a critical aspect of crustacean molting because molting requires major changes in ion concentrations and water balance as the exoskeleton is shed and replaced (Ahearn and Zhuang, 1996).



Another interesting category of genes with higher expression in juveniles are those related to oxidative stress, including peroxiredoxin and antioxidant activity. Oxidative stress proteins are known to cycle over the course of the molt cycle in crabs, perhaps because oxidative species are associated with cell signalling and/or increased mitochondrial activity and steroid synthesis (Head et al., 2019). Female *R. gigas* displayed higher expression of genes related to chitin binding, chitin metabolism, and cuticle development (Fig. 3c, d). Copepod exoskeletons are made mostly of chitin—in fact, copepods are probably the largest source of chitin in the ocean (Souza et al., 2011). There is evidence that chitin plays a role in copepod reproduction, providing a possible explanation for the upregulation of chitin metabolism in females. The application of a chitin synthesis inhibitor to adult copepods (*Acartia tonsa*) substantially reduced egg and larval viability of exposed females, although it did not cause mortality for the females themselves (Tester and Costlow, 1981). Chitin also lines the reproductive tissues of adult copepods of both sexes (Sugier et al., 2018). Interestingly, female banana shrimp (a decapod crustacean) express more transcripts related to chitin binding and catabolism compared to males (Powell et al., 2018).

We found that long non-coding RNA's (lncRNA's) make up roughly 30% of the *R. gigas* transcriptome, in line with other recent crustacean assemblies (Huerlimann et al., 2018). The functions of lncRNA's remain poorly understood, particularly outside of model systems, although they generally have roles as regulators of gene expression (Engreitz et al., 2016). In *Drosophila*, lncRNA's are implicated in embryogenesis and development, and are often expressed specifically in the nervous system and gonads (Li et al., 2019). Interestingly, lncRNA's play roles in sex determination and X-chromosome dosage compensation in both humans (Pontier and Gribnau, 2011) and flies (Li et al., 2019), and have also been implicated in sex determination in crustaceans (Yang et al., 2018; Kato et al., 2018). The only other existing study of lncRNA's in a copepod showed that some lncRNA transcripts were perturbed by drug exposure (Valenzuela-Miranda et al., 2017). In the current study, lncRNA's accounted for just ~9% of differentially expressed genes despite representing ~25% of all genes, a pattern that can be explained (at least in part) by low lncRNA abundance. Nonetheless, some lncRNA transcripts are apparently involved in the developmental transition between CV juvenile and adult. The annotation of lncRNA's in non-model organisms such as copepods may help clarify the functions of non-coding transcripts in these groups, and perhaps identify homologous lncRNA's between closely-related species.

#### 4.2. Juvenile *R. gigas* upregulate lipid and carbohydrate metabolism

Aspects of lipid and carbohydrate metabolism, including glycolysis, fatty acid synthesis, and glycerophospholipid metabolism, were upregulated in juveniles compared to adult females. These broad changes in energy metabolism may reflect preparation for an energetically expensive molt. For example, molt-associated increases in muscle growth would require large amounts energy. Other studies of crustacean development have similarly suggested that actively growing juveniles may have higher energetic requirements than adults (Kuballa et al., 2011), consistent with the upregulation of glycolysis. Juvenile copepods must also accumulate large amounts of lipids (Baumgartner and Tarrant, 2017), providing a straightforward explanation for the upregulation of fatty acid synthesis observed in *R. gigas*.

Among large lipid-storing copepods, the molecular basis of development is by far the best-studied in *Calanus finmarchicus*, which has a North Atlantic and subarctic distribution. In this species, large stores of wax esters are accumulated during the CV stage, which serve as fuel for molting, reproduction and diapause (Lee et al., 2006). Studies of *C. finmarchicus* development identified substantial shifts in gene expression related to lipid synthesis and degradation; these shifts occur both between CV and adult individuals (Lenz et al., 2014; Skottene et al., 2019), as well as within the CV stage (Tarrant et al., 2014, 2016). Tarrant et al. (2016) profiled gene expression in early- and late-stage CV

copepodites of *C. finmarchicus*, and found that many genes related to glycolysis and lipid metabolism were upregulated in early-stage CV's (Table S2 in Tarrant et al., 2016). Skottene et al. (2019) found that fatty acid degradation genes were downregulated in adult *C. finmarchicus* and suggested that protein catabolism is relatively more important than lipid catabolism in adults under starved conditions. Pathway analysis did not identify a consistent difference in the expression of fatty acid degradation genes in *R. gigas* (Fig. 4), although several individual genes in this pathway were expressed more highly in juveniles (Supplementary File 2).

Lipid metabolism is clearly linked to the development of polar *Calanus* spp. and *Calanoides* spp. because they undergo their terminal molt after emerging from a long period of lipid-fueled dormancy as juveniles (Lee et al., 2006). In contrast, *R. gigas* does not appear to rely on diapause as its primary overwintering strategy and is active or semi-active year-round in the Southern Ocean (Pasternak and Schnack-Schiel, 2002). Although *R. gigas* does store wax esters in oil sacs, it accumulates smaller lipid stores relative to species with a well-defined winter diapause (Tarling et al., 2007) and loses less lipid weight during the winter, probably because it actively feeds during that time (Kattner et al., 1994). Nonetheless, we observed upregulation of lipid synthesis in juvenile *R. gigas* relative to adult females. Regulation of lipid metabolism may be a general aspect of development for large lipid-storing copepods, rather than being strictly connected to diapause. Lipid metabolism is also important in the development of decapod crustaceans (Kuballa et al., 2011) and insects (Gilbert and O'Connor, 2012), and may be a conserved feature in pancrustacean development.

In a recent study of the lipid-storing copepod *Neocalanus flemingerii*, high expression of glycolysis and glycerophospholipid metabolism were characteristic of females in the early-to-mid stage of oogenesis (Roncalli et al., 2020). Phospholipids can be important components of copepod eggs (Ohman and Runge, 1994; Lee et al., 2006)—although this does not seem to be true in all copepod species, and some studies of the same species differ in the reported abundance of egg phospholipids (Lee et al., 1972; Leiknes et al., 2016). In *R. gigas*, the main lipid components of eggs are triglycerides and polar lipids, which include phospholipids (Ward et al., 1996). The upregulation of glycolysis and glycerophospholipid metabolism observed in *R. gigas* juveniles could reflect early oogenesis in CV copepodites. Although the reproductive biology of *R. gigas* has not been well-studied, this would be consistent with other copepods, where gonadogenesis can begin during the CV stage (Hirche, 1996) or in some species even earlier (Kosobokova, 1999). Regardless of when oogenesis begins, the lipids accumulated by juveniles serve as fuel for adult reproduction as well as for the molt itself in many copepods (Lee et al., 2006). There is evidence that this is true in *R. gigas* because it is able to reproduce in apparently food-poor conditions (Rakusa-Suszczewski et al., 1976; Ward and Shreeve, 1995; Shreeve et al., 2002).

Some of the other changes in gene expression observed in this study are likely associated specifically with female reproductive development, as we did not sample adult males. Many calanoid copepods have highly female-skewed sex ratios, due to male-specific mortality (Hirst et al., 2010) and/or skewed sex ratios at birth (Burriss and Dam, 2015). Indeed, low male:female sex ratios have frequently been observed for *R. gigas* (Marin, 1988; Spiridonov and Kosobokova, 1997). Future studies of males could clarify which aspects of the developmental program observed here are generally related to molting and development and which are specific to female reproductive development; lncRNA transcripts may be particularly interesting in this regard because of their known roles in sex determination and gonadal function (Yang et al., 2018; Kato et al., 2018; Li et al., 2019).

## 5. Conclusions and future directions

We present a high-quality transcriptome assembly for *Rhincalanus gigas*, a large and abundant lipid-storing copepod endemic to the Southern Ocean. Using this assembly, we found that the developmental

transition between late-stage (CV) juveniles and adults is marked by substantial changes in gene expression. Compared to adult females, juveniles upregulated lipid and carbohydrate metabolic pathways, which likely reflects ongoing lipid accumulation and preparation for molting and reproduction. A wide variety of other molecular processes are also implicated in molting and maturation in *R. gigas*, including muscle growth, chitin metabolism, and regulation of ion transport. The new transcriptome presented here extends the taxonomic breadth of molecular studies in copepods and contributes to our understanding of developmental transitions and metabolic adaptations of pelagic copepods. In conjunction with the ongoing development of molecular resources for other species, this assembly will enable future comparative studies for this ecologically important group of zooplankton. This will help determine whether copepods share a common developmental gene expression program, or if differences in lifestyle (herbivory versus omnivory) and lipid composition (wax esters versus triglycerides) are associated with differences in developmental transitions at the gene expression level.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.margen.2021.100835>.

### Data availability

Sequenced reads and assemblies are available on the Short Read Archive and the Transcriptome Shotgun Assembly (TSA) database, respectively (BioProject PRJNA666170). The TSA project has been deposited at GenBank under the accession GIVD00000000.

### Funding

Funding for this project was provided by the National Science Foundation (Grants OPP-1746087 to AMT and OPP-1440435 to DKS).

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

We thank the Captain, crew and technical staff of the ARSV *Lawrence M. Gould*. We also thank the Steinberg Lab, Mark Baumgartner and Phil Alatalo for technical assistance.

### References

- Ahearn, G.A., Zhuang, Z., 1996. Cellular mechanisms of calcium transport in crustaceans. *Physiol. Zool.* 69, 383–402.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Asai, S., Sanges, R., Lauritano, C., Lindeque, P.K., Esposito, F., Ianora, A., Carotenuto, Y., 2020. De novo Transcriptome assembly and gene expression profiling of the copepod *Calanus helgolandicus* feeding on the PUA-producing diatom *Skeletonema marinoi*. *Marine Drugs* 18, 392. <https://doi.org/10.3390/md18080392>.
- Atkinson, A., 1991. Life cycles of *Calanoides acutus*, *Calanus simillimus* and *Rhincalanus gigas* (Copepoda: Calanoida) within the Scotia Sea. *Mar. Biol.* 109, 79–91. <https://doi.org/10.1007/BF01320234>.
- Atkinson, A., 1998. Life cycle strategies of epipelagic copepods in the Southern Ocean. *J. Mar. Syst.* 15, 289–311. [https://doi.org/10.1016/S0924-7963\(97\)00081-X](https://doi.org/10.1016/S0924-7963(97)00081-X).
- Atkinson, A., Ward, P., Williams, R., Poulet, S.A., 1992. Feeding rates and diel vertical migration of copepods near South Georgia: comparison of shelf and oceanic sites. *Mar. Biol.* 114, 49–56. <https://doi.org/10.1007/BF00350855>.
- Baumgartner, M.F., Tarrant, A.M., 2017. The physiology and ecology of diapause in marine copepods. *Annu. Rev. Mar. Sci.* 9, 387–411. <https://doi.org/10.1146/annurev-marine-010816-060505>.
- Boutet, E., Lieberherr, D., Tognolli, M., Schneider, M., Bairoch, A., 2007. UniProtKB/Swiss-Prot. *Methods in Molecular Biology* (Clifton, N.J.) 406, 89–112. [https://doi.org/10.1007/978-1-59745-535-0\\_4](https://doi.org/10.1007/978-1-59745-535-0_4).
- Bryant, D.M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M.B., Payzin-Dogru, D., Lee, T.J., Leigh, N.D., Kuo, T.H., Davis, F.G., Bateman, J., Bryant, S., Guzikowski, A., Tsai, S.L., Coyne, S., Ye, W.W., Freeman, R.M., Peshkin, L., Tabin, C.J., Regev, A., Haas, B.J., Whited, J.L., 2017. A tissue-mapped axolotl De novo Transcriptome enables identification of limb regeneration factors. *Cell Rep.* 18, 762–776. <https://doi.org/10.1016/j.celrep.2016.12.063>.
- Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. <https://doi.org/10.1038/nmeth.3176>.
- Burris, Z.P., Dam, H.G., 2015. First evidence of biased sex ratio at birth in a calanoid copepod. *Limnol. Oceanogr.* 60, 722–731. <https://doi.org/10.1002/lno.10056>.
- Chiba, S., 2002. Spatio-temporal variability in life cycle strategy of four pelagic Antarctic copepods: *Rhincalanus gigas*, *Calanoides acutus*, *Calanus propinquus* and *Metridia gerlachei*. *Polar Bioscience* 15, 27–44.
- Conover, R., Huntley, M., 1991. Copepods in ice-covered seas—distribution, adaptations to seasonally limited food, metabolism, growth patterns and life cycle strategies in polar seas. *J. Mar. Syst.* 2, 1–41. [https://doi.org/10.1016/0924-7963\(91\)90011-1](https://doi.org/10.1016/0924-7963(91)90011-1).
- Davidson, N.M., Oshlack, A., 2014. Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biol.* 15, 410. <https://doi.org/10.1186/s13059-014-0410-6>.
- Engreitz, J.M., Ollikainen, N., Guttman, M., 2016. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat. Rev. Mol. Cell Biol.* 17, 756–770. <https://doi.org/10.1038/nrm.2016.126>.
- Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., Punta, M., 2014. Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. <https://doi.org/10.1093/nar/gkt1223>.
- Gilbert, L.I., O'Connor, J.D., 2012. Lipid metabolism and transport in arthropods. *Chemical Zoology* 5, 229–253.
- Gleiber, M., 2014. Long-Term Change in Copepod Community Structure in the Western Antarctic Peninsula: Linkage to Climate and Implications for Carbon Cycling. Ph.D. thesis. College of William and Mary - Virginia Institute of Marine Science.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mucelli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. <https://doi.org/10.1038/nbt.1883>.
- Graeve, M., Hagen, W., Kattner, G., 1994. Herbivorous or omnivorous? On the significance of lipid compositions as trophic markers in Antarctic copepods. *Deep-Sea Res. I Oceanogr. Res. Pap.* 41, 915–924. [https://doi.org/10.1016/0967-0637\(94\)90083-3](https://doi.org/10.1016/0967-0637(94)90083-3).
- Hagen, W., Schnack-Schiel, S.B., 1996. Seasonal lipid dynamics in dominant Antarctic copepods: energy for overwintering or reproduction? *Deep-Sea Res. I Oceanogr. Res. Pap.* 43, 139–158. [https://doi.org/10.1016/0967-0637\(96\)00001-5](https://doi.org/10.1016/0967-0637(96)00001-5).
- Head, T.B., Mykles, D.L., Tomanek, L., 2019. Proteomic analysis of the crustacean molting gland (Y-organ) over the course of the molt cycle. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* 29, 193–210. <https://doi.org/10.1016/j.cbd.2018.11.011>.
- Hirche, H.J., 1996. The reproductive biology of the marine copepod, *Calanus finmarchicus* — a review. *Ophelia* 44, 111–128. <https://doi.org/10.1080/00785326.1995.10429842>.
- Hirst, A.G., Bonnet, D., Conway, D.V.P., Kjørboe, T., 2010. Does predation controls adult sex ratios and longevity in marine pelagic copepods? *Limnol. Oceanogr.* 55, 2193–2206. <https://doi.org/10.4319/lno.2010.55.5.2193>.
- Huerlimann, R., Wade, N.M., Gordon, L., Montenegro, J.D., Goodall, J., McWilliam, S., Tinning, M., Siemerling, K., Giardina, E., Donovan, D., Sellars, M.J., Cowley, J.A., Condon, K., Coman, G.J., Khatkar, M.S., Raadsma, H.W., Maes, G.E., Zenger, K.R., Jerry, D.R., 2018. De novo assembly, characterization, functional annotation and expression patterns of the black tiger shrimp (*Penaeus monodon*) transcriptome. *Sci. Rep.* 8, 1–14. <https://doi.org/10.1038/s41598-018-31148-4>.
- Ilott, N.E., Ponting, C.P., 2013. Predicting long non-coding RNAs using RNA sequencing. *Methods* 63, 50–59. <https://doi.org/10.1016/j.ymeth.2013.03.019>.
- Jansen, S., Klaas, C., Krägfesky, S., von Harbou, L., Bathmann, U., 2006. Reproductive response of the copepod *Rhincalanus gigas* to an iron-induced phytoplankton bloom in the Southern Ocean. *Polar Biol.* 29, 1039–1044. <https://doi.org/10.1007/s00300-006-0147-0>.
- Kanehisa, M., Goto, S., 2000. KEGG: Kyoto Encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. <https://doi.org/10.1093/nar/28.1.27>.
- Kato, Y., Perez, C.A.G., Mohamad Ishak, N.S., Nong, Q.D., Sudo, Y., Matsuura, T., Wada, T., Watanabe, H., 2018. A 5' UTR-overlapping lncRNA activates the male-determining gene doublesex1 in the crustacean *Daphnia magna*. *Current Biology* 28, 1811–1817 e4. <https://doi.org/10.1016/j.cub.2018.04.029>.
- Katoh, K., Asimeno, G., Toh, H., 2009. Multiple alignment of DNA sequences with MAFFT. In: Posada, D. (Ed.), *Bioinformatics for DNA Sequence Analysis*. Humana Press, Totowa, NJ, pp. 39–64. [https://doi.org/10.1007/978-1-59745-251-9\\_3](https://doi.org/10.1007/978-1-59745-251-9_3).
- Methods in Molecular Biology.
- Kattner, G., Graeve, M., Hagen, W., 1994. Ontogenetic and seasonal changes in lipid and fatty acid/alcohol compositions of the dominant Antarctic copepods *Calanus propinquus*, *Calanoides acutus* and *Rhincalanus gigas*. *Mar. Biol.* 118, 637–644. <https://doi.org/10.1007/BF00347511>.
- Kopylova, E., Noé, L., Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28, 3211–3217. <https://doi.org/10.1093/bioinformatics/bts611>.
- Kosobokova, K.N., 1999. The reproductive cycle and life history of the Arctic copepod *Calanus glacialis* in the White Sea. *Polar Biol.* 22, 254–263. <https://doi.org/10.1007/s003000050418>.



- Kuballa, A.V., Holton, T.A., Paterson, B., Elizur, A., 2011. Moulting cycle specific differential gene expression profiling of the crab *Portunus pelagicus*. *BMC Genomics* 12, 147. <https://doi.org/10.1186/1471-2164-12-147>.
- Lauritano, C., Roncalli, V., Ambrosino, L., Cieslak, M.C., Ianora, A., 2020. First De novo Transcriptome of the copepod *Rhincalanus gigas* from Antarctic waters. *Biology* 9, 410. <https://doi.org/10.3390/biology9110410>.
- Lee, R., Hagen, W., Kattner, G., 2006. Lipid storage in marine zooplankton. *Marine Ecology-Progress Series* 307, 273–306. <https://doi.org/10.3354/meps307273>.
- Lee, R.F., Nevenzel, J.C., Paffenhöfer, G.A., 1972. The presence of waxes in marine planktonic copepods. *Die Naturwissenschaften* 59, 406–411. <https://doi.org/10.1007/BF00623130>.
- Lee, R.F., Nevenzel, J.C., Lewis, A.G., 1974. Lipid changes during life cycle of marine copepod, *Euchaeta japonica* marukawa. *Lipids* 9, 891–898. <https://doi.org/10.1007/BF02532615>.
- Leiknes, Ø., Etter, S.A., Tokle, N.E., Bergvik, M., Vadstein, O., Olsen, Y., 2016. The effect of essential fatty acids for the somatic growth in nauplii of *Calanus finmarchicus*. *Front. Mar. Sci.* 3, 33. <https://doi.org/10.3389/fmars.2016.00033>.
- Lenz, P.H., Roncalli, V., Hassett, R.P., Wu, L.S., Cieslak, M.C., Hartline, D.K., Christie, A.E., 2014. De novo assembly of a Transcriptome for *Calanus finmarchicus* (Crustacea, Copepoda) – the dominant zooplankton of the North Atlantic Ocean. *PLoS One* 9, e88589. <https://doi.org/10.1371/journal.pone.0088589>.
- Li, K., Tian, Y., Yuan, Y., Fan, X., Yang, M., He, Z., Yang, D., 2019. Insights into the functions of lncRNAs in *Drosophila*. *Int. J. Mol. Sci.* 20, 4646. <https://doi.org/10.3390/ijms20184646>.
- Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>.
- Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol.* 15, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
- Mackey, A.P., Atkinson, A., Hill, S.L., Ward, P., Cunningham, N.J., Johnston, N.M., Murphy, E.J., 2012. Antarctic macrozooplankton of the Southwest Atlantic sector and Bellingshausen Sea: baseline historical distributions (discovery investigations, 1928–1935) related to temperature and food, with projections for subsequent ocean warming. *Deep-Sea Res. II Top. Stud. Oceanogr.* 59–60, 130–146. <https://doi.org/10.1016/j.dsr2.2011.08.011>.
- Maps, F., Record, N.R., Pershing, A.J., 2014. A metabolic approach to dormancy in pelagic copepods helps explaining inter- and intra-specific variability in life-history strategies. *J. Plankton Res.* 36, 18–30. <https://doi.org/10.1093/plankt/fbt100>.
- Marin, V., 1988. Qualitative models of the life cycles of *Calanoides acutus*, *Calanus propinquus*, and *Rhincalanus gigas*. *Polar Biol.* 8, 439–446. <https://doi.org/10.1007/BF00264720>.
- Mauchline, J., 1998. The biology of calanoid copepods. *Adv. Mar. Biol.* 33.
- Mayzaud, P., Razouls, S., Errhif, A., Tirelli, V., Labat, J.P., 2002. Feeding, respiration and egg production rates of copepods during austral spring in the Indian sector of the Antarctic Ocean: role of the zooplankton community in carbon transformation. *Deep-Sea Res. I Oceanogr. Res. Pap.* 49, 1027–1048. [https://doi.org/10.1016/S0967-0637\(02\)00012-2](https://doi.org/10.1016/S0967-0637(02)00012-2).
- McCarthy, S.D., Dugon, M.M., Power, A.M., 2015. ‘Degraded’ RNA profiles in Arthropoda and beyond. *PeerJ* 3, e1436. <https://doi.org/10.7717/peerj.1436>.
- Mykles, D.L., Skinner, D.M., 1982. Molt cycle-associated changes in calcium-dependent proteinase activity that degrades actin and myosin in crustacean muscle. *Dev. Biol.* 92, 386–397. [https://doi.org/10.1016/0012-1606\(82\)90184-1](https://doi.org/10.1016/0012-1606(82)90184-1).
- Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. <https://doi.org/10.1093/molbev/msu300>.
- Ohman, M.D., Runge, J.A., 1994. Sustained fecundity when phytoplankton resources are in short supply: Omnivory by *Calanus finmarchicus* in the Gulf of St. Lawrence. *Limnol. Oceanogr.* 39, 21–36. <https://doi.org/10.4319/lo.1994.39.1.0021>.
- Ommanney, F.D., 1936. *Rhincalanus gigas* (Brady): a copepod of the southern macropelagial. *Discovery Report* 13, 277–384.
- Pasternak, A.F., Schnack-Schiel, S.B., 2002. Feeding patterns of dominant Antarctic copepods: an interplay of diapause, selectivity, and availability of food. In: Lopes, R.M., Reid, J.W., Rocha, C.E.F. (Eds.), *Copepoda: Developments in Ecology, Biology and Systematics*, 156. Kluwer Academic Publishers, Dordrecht, pp. 25–36. [https://doi.org/10.1007/0-306-47537-5\\_3](https://doi.org/10.1007/0-306-47537-5_3).
- Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. *Nat. Methods* 14, 417–419. <https://doi.org/10.1038/nmeth.4197>.
- Pontier, D.B., Gribnau, J., 2011. Xist regulation and function eXplored. *Hum. Genet.* 130, 223–236. <https://doi.org/10.1007/s00439-011-1008-7>.
- Powell, D., Elizur, A., Knibb, W., 2018. Sex-specific transcript expression in the hepatopancreas of the banana shrimp (*Fenneropenaeus merguensis*). *Hydrobiologia* 825, 81–90. <https://doi.org/10.1007/s10750-017-3438-2>.
- Rakusa-Suszczewski, S., McWhinnie, M.A., Cahoon, M.O., 1976. Respiration of the antarctic copepod, *Rhincalanus gigas* 1. *Limnol. Oceanogr.* 21, 763–765. <https://doi.org/10.4319/lo.1976.21.5.0763>.
- Razouls, C., 1994. Manuel d'identification des principales espèces de copépodes pélagiques antarctiques et subantarctiques. *Annales de L'Institut Océanographique* 70, 3–204.
- Roncalli, V., Cieslak, M.C., Sommer, S.A., Hopcroft, R.R., Lenz, P.H., 2018. De novo transcriptome assembly of the calanoid copepod *Neocalanus flemingeri*: a new resource for emergence from diapause. *Mar. Genomics* 37, 114–119. <https://doi.org/10.1016/j.margen.2017.09.002>.
- Roncalli, V., Cieslak, M.C., Hopcroft, R.R., Lenz, P.H., 2020. Capital breeding in a Diapausing copepod: a Transcriptomics analysis. *Front. Mar. Sci.* 7 <https://doi.org/10.3389/fmars.2020.00056>.
- RStudio Team, 2020. RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA.
- Russo, E., Lauritano, C., d'Ippolito, G., Fontana, A., Sarno, D., von Elert, E., Ianora, A., Carotenuto, Y., 2020. RNA-Seq and differential gene expression analysis in *Temora stylifera* copepod females with contrasting non-feeding nauplii survival rates: an environmental transcriptomics study. *BMC Genomics* 21, 693. <https://doi.org/10.1186/s12864-020-07112-w>.
- Schmidt, K., Atkinson, A., Stübing, D., McClelland, J.W., Montoya, J.P., Voss, M., 2003. Trophic relationships among Southern Ocean copepods and krill: some uses and limitations of a stable isotope approach. *Limnol. Oceanogr.* 48, 277–289. <https://doi.org/10.4319/lo.2003.48.1.0277>.
- Schnack-Schiel, S., Hagen, W., 1994. Life cycle strategies and seasonal variations in distribution and population structure of four dominant calanoid copepod species in the eastern Weddell Sea, Antarctica. *J. Plankton Res.* 16, 1543–1566. <https://doi.org/10.1093/plankt/16.11.1543>.
- Seret, C., 1979. *Taxonomie, Biologie et Biogéographie Des Copépodes Pélagiques Récoltés Au Cours de La Campagne MD 03 Du "Marin Dufresne" (Iles Crozet, Kerguelen, et Heard)*. Ph.D. thesis. Univ. Pierre et Marie Curie.
- Shreeve, R., Ward, P., Whitehouse, M., 2002. Copepod growth and development around South Georgia: relationships with temperature, food and krill. *Mar. Ecol. Prog. Ser.* 233, 169–183. <https://doi.org/10.3354/meps233169>.
- 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, 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Skottene, E., Tarrant, A.M., Olsen, A.J., Altin, D., Østensen, M.A., Hansen, B.H., Choquet, M., Jenssen, B.M., Olsen, R.E., 2019. The  $\beta$ -oxidation pathway is downregulated during diapause termination in *Calanus* copepods. *Sci. Rep.* 9, 1–13. <https://doi.org/10.1038/s41598-019-53032-5>.
- Smith-Unna, R., Bournsnel, C., Patro, R., Hibberd, J.M., Kelly, S., 2016. TransRate: reference-free quality assessment of de novo transcriptome assemblies. *Genome Res.* 26, 1134–1144. <https://doi.org/10.1101/gr.196469.115>.
- Soneson, C., Love, M.I., Robinson, M.D., 2016. Differential analyses for RNA-Seq: transcript-level estimates improve gene-level inferences. *F1000Research* 4, 1521. <https://doi.org/10.12688/f1000research.7563.2>.
- Song, L., Florea, L., 2015. Rcorrector: efficient and accurate error correction for Illumina RNA-Seq reads. *GigaScience* 4, 48. <https://doi.org/10.1186/s13742-015-0089-y>.
- Souza, C.P., Almeida, B.C., Colwell, R.R., Rivera, I.N.G., 2011. The importance of chitin in the marine environment. *Mar. Biotechnol.* 13, 823. <https://doi.org/10.1007/s10126-011-9388-1>.
- Spiridonov, V.A., Kosobokova, K.N., 1997. Winter ontogenetic migrations and the onset of gonad development in large dominant calanoid copepods in the Weddell gyre (Antarctica). *Mar. Ecol. Prog. Ser.* 157, 233–246. <https://doi.org/10.3354/meps157233>.
- Sugier, K., Vacherie, B., Cornils, A., Wincker, P., Jamet, J.L., Madoui, M.A., 2018. Chitin distribution in the Oithona digestive and reproductive systems revealed by fluorescence microscopy. *PeerJ* 6, e4685. <https://doi.org/10.7717/peerj.4685>.
- Supek, F., Bošnjak, M., Skunca, N., Šmuc, T., 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLoS ONE* 6 (7). <https://doi.org/10.1371/journal.pone.0021800>.
- Tarling, G., Shreeve, R., Ward, P., 2007. Life-cycle and population dynamics of *Rhincalanus gigas* (Copepoda: Calanoida) in the Scotia Sea. *Mar. Ecol. Prog. Ser.* 338, 145–158. <https://doi.org/10.3354/meps338145>.
- Tarrant, A.M., Baumgartner, M.F., Hansen, B.H., Altin, D., Nordtug, T., Olsen, A.J., 2014. Transcriptional profiling of reproductive development, lipid storage and molting throughout the last juvenile stage of the marine copepod *Calanus finmarchicus*. *Front. Zool.* 11, 91. <https://doi.org/10.1186/s12983-014-0091-8>.
- Tarrant, A.M., Baumgartner, M.F., Lysiak, N.S.J., Altin, D., Størseth, T.R., Hansen, B.H., 2016. Transcriptional profiling of metabolic transitions during development and diapause preparation in the copepod *Calanus finmarchicus*. *Integr. Comp. Biol.* 56, 1157–1169. <https://doi.org/10.1093/icb/icw060>.
- Tarrant, A.M., Nilsson, B., Hansen, B.W., 2019. Molecular physiology of copepods: from biomarker to transcriptomes and back again. *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics* 2019, 230–247. <https://doi.org/10.1016/j.cbd.2019.03.005>.
- Tester, P.A., Costlow, J.D., 1981. Effect of insect growth regulator Dimilin (TH 6040) on fecundity and egg viability of the marine copepod *Acartia tonsa*. *Mar. Ecol. Prog. Ser.* 5, 297–302.
- Valenzuela-Miranda, D., Etebari, K., Asgari, S., Gallardo-Escárate, C., 2017. Long noncoding RNAs: unexplored players in the drug response of the sea louse *Caligus rogercresseyi*. *Agri Gene* 4, 1–7. <https://doi.org/10.1016/j.aggene.2017.03.001>.
- van der Maaten, L., Hinton, G., 2008. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605.
- Voronina, N., 1970. Seasonal cycles of some common Antarctic copepod species. *Antarctic Ecology* 1, 162–172.
- Ward, P., Shreeve, R., 1995. Egg production in three species of Antarctic Calanoid copepods during an austral summer. *Deep-Sea Res. I Oceanogr. Res. Pap.* 42, 721–735. [https://doi.org/10.1016/0967-0637\(95\)00018-2](https://doi.org/10.1016/0967-0637(95)00018-2).
- Ward, P., Shreeve, R., Cripps, G., 1996. *Rhincalanus Gigas* and *Calanus simillimus*: lipid storage patterns of two species of copepod in the seasonally ice-free zone of the Southern Ocean. *J. Plankton Res.* 18, 1439–1454. <https://doi.org/10.1093/plankt/18.8.1439>.

- Ward, P., Atkinson, A., Schnack-Schiel, S., Murray, A., 1997. Regional variation in the life cycle of *Rhincalanus gigas* (Copepoda: Calanoida) in the Atlantic sector of the Southern Ocean—re-examination of existing data (1928 to 1993). *Mar. Ecol. Prog. Ser.* 157, 261–275. <https://doi.org/10.3354/meps157261>.
- Wright, R.M., Aglyamova, G.V., Meyer, E., Matz, M.V., 2015. Gene expression associated with white syndromes in a reef building coral, *Acropora hyacinthus*. *BMC Genomics* 16, 371. <https://doi.org/10.1186/s12864-015-1540-2>.
- Wucher, V., Legeai, F., Hédan, B., Rizk, G., Lagoutte, L., Leeb, T., Jagannathan, V., Cadieu, E., David, A., Lohi, H., Cirera, S., Fredholm, M., Botherel, N., Leegwater, P.A. J., Le Béguec, C., Fieten, H., Johnson, J., Alföldi, J., André, C., Lindblad-Toh, K., Hitte, C., Derrien, T., 2017. FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res.* 45, e57. <https://doi.org/10.1093/nar/gkw1306>.
- Yang, X., Ikhwanuddin, M., Li, X., Lin, F., Wu, Q., Zhang, Y., You, C., Liu, W., Cheng, Y., Shi, X., Wang, S., Ma, H., 2018. Comparative Transcriptome analysis provides insights into differentially expressed genes and long non-coding RNAs between ovary and testis of the mud crab (*Scylla paramamosain*). *Mar. Biotechnol.* 20, 20–34. <https://doi.org/10.1007/s10126-017-9784-2>.
- Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J.R., Amaral-Zettler, L., Gilbert, J.A., Karsch-Mizrachi, I., Johnston, A., Cochrane, G., Vaughan, R., Hunter, C., Park, J., Morrison, N., Rocca-Serra, P., Sterk, P., Arumugam, M., Bailey, M., Baumgartner, L., Birren, B.W., Blaser, M.J., Bonazzi, V., Booth, T., Bork, P., Bushman, F.D., Buttigieg, P.L., Chain, P.S.G., Charlson, E., Costello, E.K., Huot-Creasy, H., Dawyndt, P., DeSantis, T., Fierer, N., Fuhrman, J.A., Gallery, R.E., Gevers, D., Gibbs, R.A., Gil, I.S., Gonzalez, A., Gordon, J.I., Guralnick, R., Hankeln, W., Highlander, S., Hugenholtz, P., Jansson, J., Kau, A.L., Kelley, S.T., Kennedy, J., Knights, D., Koren, O., Kuczynski, J., Kyrpides, N., Larsen, R., Lauber, C.L., Legg, T., Ley, R.E., Lozupone, C.A., Ludwig, W., Lyons, D., Maguire, E., Methé, B.A., Meyer, F., Muegge, B., Nakielnny, S., Nelson, K.E., Nemergut, D., Neufeld, J.D., Newbold, L.K., Oliver, A.E., Pace, N.R., Palanisamy, G., Peplies, J., Petrosino, J., Proctor, L., Pruesse, E., Quast, C., Raes, J., Ratnasingham, S., Ravel, J., Relman, D.A., Assunta-Sansone, S., Schloss, P.D., Schriml, L., Sinha, R., Smith, M.I., Sodergren, E., Spor, A., Stombaugh, J., Tiedje, J.M., Ward, D.V., Weinstock, G.M., Wendel, D., White, O., Whiteley, A., Wilke, A., Wortman, J.R., Yatsunenko, T., Glöckner, F.O., 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. *Nat. Biotechnol.* 29, 415–420. <https://doi.org/10.1038/nbt.1823>.
- Yu, G., Wang, L.G., Han, Y., He, Q.Y., 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS: A Journal of Integrative Biology* 16, 284–287. <https://doi.org/10.1089/omi.2011.0118>.