Embryogenesis of a calanoid copepod analyzed by transcriptomics

Cifuentes Acebal, Miguel; Dalgaard, Louise Torp; Jørgensen, Tue Sparholt; Hansen, Benni Winding

Published in:
Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics

DOI:
10.1016/j.cbd.2022.101054

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

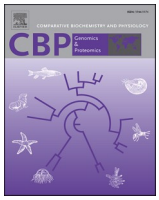
Citation for published version (APA):
Cifuentes Acebal, M., Dalgaard, L. T., Jørgensen, T. S., & Hansen, B. W. (2023). Embryogenesis of a calanoid copepod analyzed by transcriptomics. *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics*, *45*, [101054]. https://doi.org/10.1016/j.cbd.2022.101054

Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics 45 (2023) 101054

Contents lists available at ScienceDirect

Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics

journal homepage: www.elsevier.com/locate/cbpd



Embryogenesis of a calanoid copepod analyzed by transcriptomics

Miguel Cifuentes Acebal [a], Louise Torp Dalgaard [a], Tue Sparholt Jørgensen [a,b,c], Benni Winding Hansen [a,*]

[a] Department of Science and Environment, Roskilde University, Universitetsvej 1, DK-4000 Roskilde, Denmark
[b] Department of Environmental Science - Environmental Microbiology and Biotechnology, Aarhus University, Frederiksborgvej 399, DK-4000 Roskilde, Denmark
[c] The Novo Nordisk Foundation Center for Biosustainability (DTU Biosustain) at the Technical University of Denmark, Building 220, Kemitorvet, DK-2800 Kgs. Lyngby, Denmark[1]

ARTICLE INFO

Edited by Chris Martyniuk

Keywords:
Enrichment of pathways
Subitaneous embryogenesis, comparative genomics
Transcriptome assembly
Invertebrate genomics

ABSTRACT

The calanoid copepod *Acartia tonsa* (Dana) has attracted interest because of its use as a copepod model organism as well as its potential economic role as live fish larval feed. While the adult genome and transcriptome of *A. tonsa* has been investigated, no studies have been performed investigating the genome-wide transcriptional changes during the normal subitaneous embryogenesis. Thus, the aim of the current study was to investigate said transcriptional changes throughout *A. tonsa* embryonic development.

RNA extraction and de novo transcriptome assembly for the subitaneous embryogenesis of the copepod was conducted. The assembly includes for the first-time samples describing quiescent development and overall helps establishing a framework for future studies on the molecular biology of our species of interest. Among the findings reported, sequences annotated to well-known developmental genes, were identified. At the same time are described the molecular changes and gene expression levels throughout the entire 42 h the embryonic development lasts.

In conclusion, here we present the most complete genome-wide transcriptional map of early copepod embryonic development to date, enabling further use of *A. tonsa* as a model organism for crustacean development. Keywords: enrichment of pathways; subitaneous embryogenesis, comparative genomics; transcriptome assembly; invertebrate genomics.

1. Introduction

The calanoid copepod *A. tonsa* has recently attracted interest to its molecular biology. As the literature shows, it is one of the few copepods which genome and transcriptome has been sequenced out of the ~14,000 known species (Jørgensen et al., 2019c). The increasing interest in the species is especially noticeable during embryogenesis, where several studies to some extents have described both subitaneous and quiescent development (Nilsson et al., 2014; Nilsson and Hansen, 2018; Jørgensen et al., 2019a). The target species is among the approximately 50 calanoids from the Centropagid superfamily described to produce resting eggs wherein the embryo enters arrested development. This can either be maternally programmed (diapause) or stimulated by adverse environmental conditions directly on the embryo (quiescence) (Holm et al., 2018). Nevertheless, the knowledge on the subject is still limited and additional information on the biology behind *A. tonsa* embryogenesis would be highly beneficial. The reason behind the growing interest is the capability of the free spawning *A. tonsa* of inducing embryonic arrest when adverse environmental conditions occur and the ecological and industrial implications this supposes (Holm et al., 2018; Hansen, 2019).

To support the expanding knowledge in copepod developmental processes, it is important to understand the timing of the normal subitaneous development. Achieving a good understanding of the mechanisms of subitaneous development works will enable further studies of stress induced development such as quiescence. Both subitaneous and quiescent embryological processes have been theorized to follow the same pathways most of the time and be differentially regulated just at certain developmental stages (Nilsson and Hansen, 2018; Acebal et al., 2022). At the same time, expanding the knowledge on subitaneous

Abbreviations: DE, differential expression; DET, differentially expressed transcripts; GO, Gene Ontology; BUSCO, Basic Universal Single Copy Orthologs.

* Corresponding author at: Universitetsvej 1, DK-4000 Roskilde, Denmark.
E-mail address: bhansen@ruc.dk (B.W. Hansen).

[1] Present affiliation.

https://doi.org/10.1016/j.cbd.2022.101054
Received 7 September 2022; Received in revised form 22 November 2022; Accepted 6 December 2022
Available online 12 December 2022

Table 1
Acartia tonsa. Description of the samples used to acquire RNA-seq libraries and data from the raw reads. In italics and an asterisk (*) are the samples that were not included in the final assembly. After removal of the samples 26 libraries (5*3 hours, 5*7 hours, 4*14 hours, 4*24 hours, 5*42 hours, 3*quiescent) were used for the assembly of the transcriptome (the 3 quiescent samples are not shown here as they were part of another experiment run simultaneously) and reported in Acebal et al. (In review).

Sample name	Timepoint in development	Technical replicate	RNA concentration (ng/μL)	% duplicate reads	% GC	RAW reads (10^6)	Strandedness (%F)
3A	3 hours	1	37.4	69.7 %	44 %	33.2	67.4 %
3B	3 hours	2	34.9	71.3 %	45 %	39.5	69.8 %
3C	3 hours	3	20.0	65.6 %	47 %	23.1	68.2 %
3D	3 hours	4	37.7	61.9 %	45 %	22.2	70.4 %
3E	3 hours	5	34.4	63.8 %	46 %	22.6	70.2 %
7A	7 hours	1	39.8	68.6 %	45 %	27.5	64.6 %
7B	7 hours	2	31.7	71.1 %	45 %	25.2	62.5 %
7C	7 hours	3	27.1	65.4 %	46 %	23.9	64.3 %
7D	7 hours	4	34.5	64.0 %	44 %	23.8	66.0 %
7E	7 hours	5	14.7	66.8 %	46 %	20.0	66.9 %
14A	14 hours	1	34.3	67.1 %	45 %	26.7	64.0 %
14B	14 hours	2	38.4	67.8 %	45 %	30.1	69.0 %
14C	14 hours	3	18.7	63.5 %	44 %	23.4	67.5 %
*14D	14 hours	4	57.0	82.7 %	45 %	22.8	60.0 %
14E	14 hours	5	51.0	58.5 %	45 %	19.3	72.3 %
24A	24 hours	1	13.7	63.4 %	44 %	24.2	83.9 %
24B	24 hours	2	15.9	63.8 %	45 %	22.3	80.3 %
24C	24 hours	3	12.0	65.2 %	46 %	17.5	75.3 %
24D	24 hours	4	15.2	73.9 %	45 %	23.2	67.9 %
*24E	24 hours	5	15.9	73.0 %	46 %	22.9	67.0 %
42A	42 hours	1	15.6	66.1 %	45 %	23.3	77.0 %
42B	42 hours	2	14.5	78.6 %	42 %	18.5	78.9 %
42C	42 hours	3	14.0	79.0 %	43 %	21.6	83.4 %
42D	42 hours	4	12.9	69.5 %	44 %	22.0	72.0 %
42E	42 hours	5	20.9	69.7 %	46 %	20.4	72.0 %

development and identifying the key sequences behind it, will pose a better understanding on the lifecycle of calanoid copepods and the molecular cues that the embryos may undergo and how the environment can affect them. To achieve these goals, we performed a whole transcriptome analysis of the normal subitaneous development of *Acartia tonsa*, combined with differential expression (DE) analysis and a pathway enrichment analysis, these tools should show the most important transcripts and processes that characterize embryonic development of *A. tonsa*.

2. Material and methods

2.1. Cultures

Copepods belonging to the species *A. tonsa* (*DFH-ATI*) first obtained in Øresund (N 56″/E12″; Denmark) in 1981 and in continuous culture ever since were used for this project. The strain has been maintained under constant salinity, temperature, and light conditions for >40 years (Støttrup et al., 1986). Moreover, it has been used before for developmental research (Nilsson and Hansen, 2018; Jørgensen et al., 2019a, 2019c). A copepod culture was set up in Roskilde University prior to the project and kept in similar environmental conditions as the ones above mentioned (salinity 32; 16.9 °C; no light). The culture was fed daily ad libitum with the monoculture of the microalgae *Rhodomonas salina* (>800 μg C L^{-1} sensu Berggreen et al., 1988) which was kept in 20 L plastic bags with F/2 media (Guillard, 1975). The copepod culture was kept in 60 L flat-bottomed polyethylene tanks. The seawater used in the experiments was UV light treated and filtered through a 0.2 μm pore size filter.

2.2. Library preparation

For the preparation of the 26 RNA libraries (Table 1), eggs were obtained from the bottom of the tank containing the copepod culture by first thoroughly cleaning it. After this, the samples were collected from the bottom every hour by cleaning the tank in a similar fashion as the one described here (Jørgensen et al., 2019a). The method selected has been used in the cited literature as a reliable method to obtain the most synchronous in age samples possible. The eggs were then filtered using a 54 μm mesh and allowed to continue development until the target timepoint was reached: 3, 7, 14, 24, and 42 h after which they were collected (16–32 cell stage, gastrulation, organogenesis, limb bud formation, and final nauplii, respectively) (Supplementary Table I). With this method, the eggs obtained are not a snapshot but rather represent a 1-hour range in the embryo development, the hour here shown represents the maximum age of eggs in said sample (e.g., 3 h is the eggs with an age ranging between 2 and 3 h). Subitaneous eggs were allowed to develop normally in the same environment as the main culture. Once incubation was over, the eggs were collected in an Eppendorf tube and later divided in five technical replicates per timepoint that contained an equal number of eggs per sample (Supplementary Table I). The eggs in the replicates were collected together and later divided into subsamples, thus they cannot be considered as true biological replicates for its origin is the same population at the same time. The number of eggs per replicate was counted under the dissection microscope at 25–40× magnification and transferred to 1.5 mL Eppendorf tubes. To pellet the eggs at the bottom of the Eppendorf tubes, the samples were centrifuged (10,000 rpm, 30 s), and the supernatant was removed. Finally, to preserve the RNA present within the samples by stopping RNAase activity 50 μL of RNA later (ThermoFisher Scientific, Massachusetts, USA) were added to each tube.

After all samples had been collected, RNA was extracted using RNAEasy mini kit (Qiagen GmbH, Hilden, Germany) in 50 μL elutions and following manufacturer's protocol. The RNA yield per sample was measured in a Qubit fluorometer (ThermoFisher Scientific, Massachusetts, USA) and from here 10 μL of RNA were selected for library preparation for all the samples when the RNA concentration was ≥20 ng/μL and 20 μL when the RNA concentration was <20 ng/μL. The libraries were made using the TruSeq stranded mRNA kit (Illumina, California, USA) following half volumes of manufacturers protocol (Combs and Eisen, 2015). The samples were sequenced on an Illumina NextSeq500 (Illumina, California, USA) using a 75-cycle kit, giving a 1 × 75 single-read length. In a complementary experiment, samples corresponding to quiescent eggs were prepared, the results from these samples are

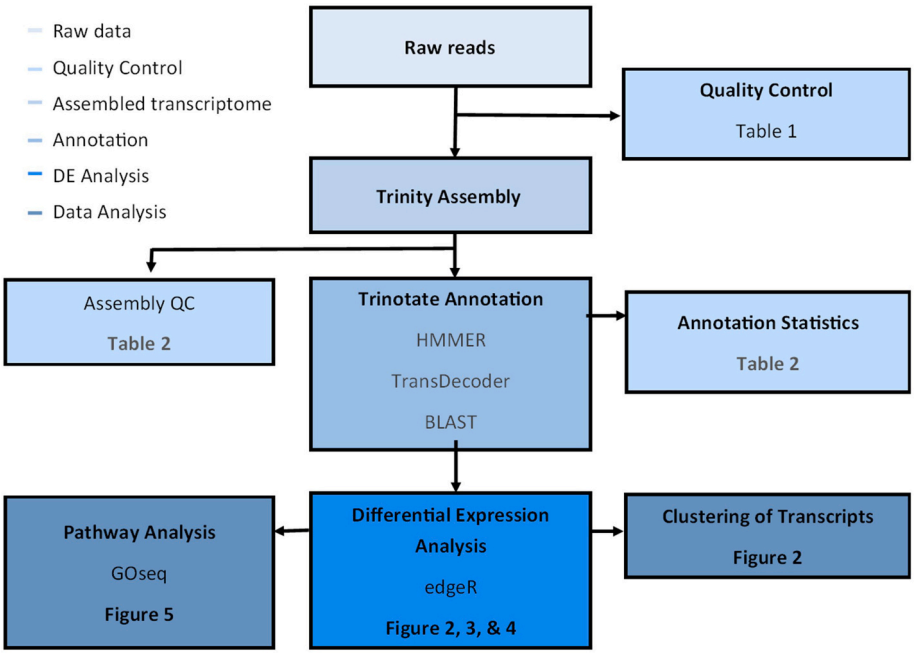


Fig. 1. *Acartia tonsa*. Flowchart of the bioinformatics analysis using the Trinity-Trinotate pipeline. The samples here labeled as raw data/reads were used for the de novo transcriptome assembly are described in Table 1. Each shade of blue represents a different step in the data analysis and its meaning can be seen at the top-left corner.

discussed in a sister publication. RNA-seq libraries of these were sequenced in parallel with the subitaneous samples, and the analysis are described elsewhere (Acebal et al., 2022, manuscript submitted).

2.3. Bioinformatics analysis

The raw RNAseq libraries were uploaded to the Galaxy server in Freiburg (Afgan et al., 2018; Freiburg Galaxy Team, 2021) where the more resource demanding analysis was conducted. The guidelines proposed by the Galaxy team were followed (Bretaudeau et al., 2021). A quality control (QC) report on the libraries was produced and those that showed aberrant values were removed. Any sample showing a different trend in the number of reads ($<10^7$ reads) or the duplicate count (>80 %), since these two values appeared to be the most consistent through files, were considered aberrant. Additional QC data including Phred scores, diversity per base and others, is available as Supplementary material I. In short, all metrics were satisfactory, with the Phred score above 30 in all libraries and nucleotides. Although it is common to find contaminant adaptor content in RNA-seq data, no such contaminants could be found in the reads after several QC analysis with different parameters. The workflow of the analysis can be seen in Fig. 1.

Samples were merged and assembled using the frequently used Trinity software (version v.2.13.2) (Grabherr et al., 2011; Galachyants et al., 2019; Hölzer and Marz, 2019) (Fig. 1). Trinity was run with default settings for a forward stranded library (single-end, forward strand specific library, minimum contig length: 200, no genome guided mode, minimum count for k-mers to be considered: 1) and a transcriptome was produced. Before continuing in the pipeline, a new QC analysis was performed calculating the N50 values and a Benchmarking Universal Single-Copy Orthologs (BUSCO) (Simao et al., 2015) analysis with the arthropoda_odb10 database.

The assembly was used as an input for Trinotate to obtain a functional annotation of the transcriptome as described (Bryant et al., 2017; "Home Trinotate/Trinotate.github.io Wiki GitHub," 2021) (Fig. 1). The top Blastx and Blastp hits against the Uniprot database (Bateman et al., 2021) were also submitted and the Pfam domains identified by HMMER (Fig. 1). Additionally, TransDecoder (Haas et al., 2013) predicted coding regions were also submitted to produce the annotation. Trinotate also retrieves the Gene Ontology (GO) terms from the Uniprot database (Bateman et al., 2021) using the best Blast hits giving information regarding the known roles of said genes in other organisms. A summarized report on both the transcriptome and the annotation (Table 2) was produced with trinotateR (Stubben, 2016).

With the annotated transcriptome a DE analysis was performed using edgeR (Robinson et al., 2010). This was achieved by comparing the subitaneous samples (3, 7, 14, 24, 42 h) in a pairwise analysis against each other. Subitaneous samples were also compared against each other as part of the analysis. The analysis was done in two steps. First, non-normalized reads were used to produce the initial DE results. The second step was to extract and cluster those transcripts which were significantly DE according to the false discovery rate (FDR) and fold change values using normalized values. After the analysis was performed, a correlation matrix showing the similarity between samples was drawn by edgeR. The transcripts found to be DE were clustered according to a 50 % similarity cutoff value in the hierarchical tree produced by edgeR and shown in Supplementary Fig. I. At the same time, a GO enrichment was also performed using the GOSeq package for R (Young et al., 2010) also included with Trinity.

The results from the pathway analysis were manually reviewed and individual pathways were selected according to relevance and the adjusted p-value. At the same time pathways where only one gene was present were also discarded. Then the most significant biological process (BP) pathways per pairwise comparison were chosen and plotted together to evaluate the organism's behavior during subitaneous embryogenesis. Molecular function and cellular component terms were also analyzed (data not shown).

2.4. Software and statistical methods

All the analysis was conducted in a Unix environment either in the Galaxy Europe servers or in a personal computer when the processing power required was not excessive. The software was run using the default settings unless stated otherwise. The Trinity-Trinotate pipeline (Fig. 1) was used to assemble and annotate the RNA libraries. The DE

Table 2
Acartia tonsa. Summary table showing different metrics used to evaluate the quality of the three different transcriptome assemblies of the calanoid copepod *A. tonsa* (top). Data regarding the annotation of the transcriptome **GSE210554** (bottom).

Transcriptome summary						
Acartia tonsa transcriptome assembly and Trinity version	GSE210554 v.2.13.2		HAGX01 v.2.5.1		GFWY00000000.1 v.2.3.2	
Reference	This study		Jørgensen et al., 2019c		Nilsson et al., 2018	
Assembly statistics	**General metric**					
Total 'genes':	131,359		61,149		27,171	
Total transcripts:	252,212		117,406		60,662	
Percent GC:	38.34		37.26		38.49	
Stats based on ALL transcript contigs:						
Contig N90:	274		562		497	
Contig N50:	1233		1052		1874	
Contig N50 (>2kbp transcripts):	3311		2929		3115	
Contig N10:	4382		3024		4468	
Median contig length:	370		739		790	
Average contig:	714.68		993.89		1222.45	
Total assembled bases:	180,251,909		118,709,440		74,163,142	
Type of BUSCO (Database arthropoda_odb10)	**Number of BUSCO**	**Percentage of BUSCO**	**Number of BUSCO**	**Percentage of BUSCO**	**Number of BUSCO**	**Percentage of BUSCO**
Complete BUSCOs (C)	971	95.8 %	861	85.0 %	929	91.7 %
Complete and single-copy BUSCOs (S)	147	14.5 %	442	43.6 %	382	37.7 %
Complete and duplicated BUSCOs (D)	824	81.3 %	419	41.4 %	547	54.0 %
Fragmented BUSCOs (F)	25	2.5 %	91	9.0 %	22	2.2 %
Missing BUSCOs (M)	17	1.7 %	61	6.0 %	62	6.1 %
Total BUSCO groups searched	1,013	100 %	1,013	100 %	1013	100 %

Annotation summary		
Annotation feature	**Unique**	**Total**
Genes	131,359	254,118
Transcripts	252,212	254,118
Top BLASTX hit Swissprot	63,237	71,288
Top BLASTP hit Swissprot	37,827	52,258
Pfam	34,556	51,987
SignalP	2700	8029
Keggs mapped	19,163	61,195
Eggnogs mapped	408	1260
GOs mapped from BLASTX hits	17,363	70,074
GOs mapped from BLASTP hits	13,464	51,216
Gos mapped from Pfam hits	1786	31,172
Total number GO Terms	32,613	152,462

analysis was performed using the *edgeR* package part of Bioconductor project setting significance at alpha level 0.05 and Fold change of 2 and using the gene mode. Fold change was calculated as Log2 (FPKM+1) and was centered around the mean expression levels in the transcriptome. FPKM is a measure that stands for fragments per kilobase of exon per million mapped fragments and is used as a within sample normalization method (Mortazavi et al., 2008; Trapnell et al., 2010). For the enrichment analysis GOSeq, also a part of Bioconductor project (Gentleman et al., 2004), was used. GOSeq as well as *edgeR* use the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) to calculate the FDR.

For the QC analysis, FastQC, BUSCO, *trinitystats.pl* (Trinity), *infer_experiment.py* (RSeQC) scripts, and TrinotateR were used. MultiQC was used for visualization (Andrews et al., 2010; Grabherr et al., 2011; Wang et al., 2012; Simao et al., 2015; Ewels et al., 2016; Stubben, 2016). Finally, the transcripts were analyzed using Blastx and Blastp searches against Swissprot database, Pfam search was done using HMMER and signal peptides were identified with SignalP 4.1 (Altschul et al., 1990; Bateman et al., 2021; "HMMER," n.d.; Mistry et al., 2021; Nielsen et al., 2019). Plotting was done with GIMP, R tidyverse and Bioconductor packages; elements from Flaticon.com, and Servier Medical Art licensed under a Creative Commons Attribution 3.0 Unported License were used for illustration (Flaticon, 2022; Gentleman et al., 2004; LesLaboratories Servier SAS, 2020; The GIMP Development Team, n.d.; Wickham et al., 2019).

Finally, the reason behind performing a de novo transcriptome assembly using Trinity instead of a mapped alignment is twofold: On the one hand, and even though the authors are aware of the benefits that using an already existing public assembly to map the reads supposes, the number of reads that were mapped to annotated features was very low when this approach was used. Only around 20 % of the reads were mapped if averaging all the libraries (Supplementary Table II). The RNAseq mapping approach was performed using HISAT2 as the mapping software and Subread featureCounts to count the mapped reads that were assigned to annotated features (Liao et al., 2014; Zhang et al., 2021). On the other hand, while the publicly available assembly does an excellent job in describing the ontogeny of *A. tonsa*, it does not contain any stress-induced life stages such as quiescence or diapause. This could be a reason for bias, since the current transcriptome was used for a complementary study regarding quiescence (Acebal et al., 2022) and

misrepresentation could lead to misleading conclusions, as some key sequences in quiescence may be underrepresented or not shown in the public assembly under the accession number: **HAGX01**. These two circumstances combined strongly advocated for a de novo assembly approach as a more reliable approach.

3. Results

3.1. Quality control

The QC analysis of the RNA-seq raw data is shown in Table 1 and overall similar values can be observed through the different samples. The analysis showed dissonant values of raw reads ($<10^7$ reads) (14D) for one sample. Additionally, another library was too dissimilar to the rest of the libraries when further analyzed (24E). These were not included in the transcriptome assembly. Duplication remained consistent through the samples with high values that could be caused by highly transcribed constitutive genes. Regarding the strandedness of the samples it was found not to be optimal, with values ranging between what is expected for unstranded libraries (50 %) and what is expected for a stranded one (>95 %) (Table 1).

The previously mentioned samples (14D & 24E In italics and marked with an asterisk in Table 1) were removed from further analysis. Additionally, a third library (42D), did not group together with the rest of 42 h replicates and instead was shown to be closer to 7 h in the correlation plot (Supplementary Fig. I). It was decided to be included as it shared similarity to other libraries in the study. Moreover, the per base sequence quality scores were good, with all values above 30 Phred score (Supplementary material I).

Once trimming of aberrant samples was performed, the transcriptome was assembled and annotated. The QC and meta data for the transcriptome assembly are shown in Table 2 where they are compared against the two other publicly available *A. tonsa* transcriptomes. A total of 131,359 putative genes were reported by Trinity out of which 63,000 had a match against the Blastx analysis and 38,000 against Blastp. Other significant annotations were 35,000 Pfam hits; 3000 against SignalP; and 33,000 mapped unique GO terms. The GC content of the assembly was 38.3 % throughout the transcriptome and a N50 value of 1230 based on all isoforms and 750 when considering only the longest isoform per putative gene (Table 2). Additionally, if only transcripts bigger than 2kbp were considered, N50 yields the best results of the three assemblies. The BUSCO analysis complements this data reporting a 95.8 % of complete BUSCO genes (Table 2), while the fragmented BUSCOs account for 2.5 % and missing for 1.7 %. The correlation matrix drawn by Trinity (Supplementary Fig. I) shows the replicates to be more similar between them than to other timepoints, except for the case of 42D, which is clustered closer to the 7 hour and 3 hour samples. The QC values here reported show similar values of N50 (Roncalli et al., 2018; Jørgensen et al., 2019c) and BUSCO completeness (Jørgensen et al., 2019b, 2019c) to other copepod transcriptomes. Therefore, the assembly was adequate for further analysis. The complete assembly and raw sequence libraries are deposited in gene expression omnibus (GEO) (Barrett et al., 2013) under accession number **GSE210554**. If compared to **HAGX01** and **GFWY00000000.1**, GC content is remarkably close within the three assemblies, with just 1.23 % difference between the highest and the lowest value. The genes and transcripts show high variability between assemblies although there is always approximately twice as many transcripts as there are genes. The high variability between assemblies is likely due to differences in the software and parameters used to construct them, as well as the life stages represented, and the different effort in reducing the fragmented genes. Moreover, the current assembly is the one that retrieves the best results for a BUSCO analysis in all categories (More complete BUSCOS, less missing and fragmented genes). At the same time, it is also the assembly with the highest number of duplicate genes. This is likely because it contains more than twice the number of transcripts as **HAGX01**; probably because some transcripts are recognized by Trinity as more than one, effectively duplicating the final number. Additionally, an unknown extent of the transcripts present in the current assembly could be caused by splice variants of the sequences. Moreover, when the biggest transcripts are considered (Bigger than 2 kb, Table 2), it can be seen how the current assembly has a larger number of long fragments of the three assemblies.

3.2. Cluster analysis

Five embryological timepoints were selected and analyzed. Given the large number of changes occurring in the developing embryo through the approximately two days it takes for the egg to hatch, some criteria had to be established for selection of stages. The timeline shown in this study was chosen to represent well-known embryological stages in Arthropoda, among these, *A. tonsa* and *Drosophila melanogaster*: The 16–32 cell stage (3 h), gastrulation (7 h), organogenesis (14 h & 24 h), limb bud formation (24 h), and final nauplii (42 h); in model developmental organisms so that any similarities/differences in the gene expression pattern could easily be noted and described. This was possible thanks to the description of embryonic stages in *A. tonsa* done previously (Nilsson and Hansen, 2018).

The transcriptome analysis revealed a total of 9502 assembled transcripts divided in 7 clusters (53.4 %, 21.2 %, 19.9 %, 1 %, 4 %, 0.6 %, 0.5 % of the transcripts, respectively) according to a 50 % similarity cutoff value to be DE between any two of the time points (Fig. 2). Among these transcripts, some were annotated to developmental genes or proteins similar to those in other organisms (*Drosophila melanogaster* and Humans mostly): protein Wnt-8 (*wnt8*), Spaetzle (*spz*), protein lethal (2) essential for life (*l(2)efl*), protein roadkill (*rdx*), or protein scarface (*scaf*) are some examples. The seven clusters before mentioned can be merged in three groups regarding their observed behavior (Fig. 2).

3.2.1. Group I

This group (clusters A, B, C) is a heterogeneous mixture of sequences caused by the high density of transcripts (8976). One of the most notable observations is the prominence of 14 h timepoint in clusters A and C while the surrounding timepoints (7 h, 24 h) seem to have decreased expression values. Among the transcripts observed in this group, sequences annotated to chitin metabolism genes (pupal cuticle protein Edg-84A (*cup8*)), developmental genes (*l(2)efl*, *wnt8*), and transcription factors (upstream activation factor subunit UAF30 (*uaf30*)) can be observed.

3.2.2. Group II

The second group is formed by the three least populated clusters (clusters D, E, F) (199 transcripts), and is characterized by having increased expression values as embryogenesis progresses. Clusters D and E show a similar behavior pattern with the exception that the fold changes in expression are more profound when looking to cluster E. As the analysis of the previous group had also shown, an upregulation of sequences occurs at 14 h together with downregulation at 7 and 24 h. Regarding the last cluster in this group, F, an active downregulation of sequences is observed for its 61 sequences at both 3 and 7 h. The next timepoints then show increasing expression values that reach almost 4-fold change at 14 h. After this moment, the sequences reach a plateau and stay upregulated for the rest of the developmental stages. Most of the sequences in this group are not annotated to known features, even after using less stringent search parameters.

3.2.3. Group III

The last cluster (G) is the only occurrence of transcripts starting with high expression values (4-fold from baseline) and later decreasing until reaching baseline or even expression values that indicate a downregulation. The set of transcripts starts being highly expressed at the initial phases of embryonic development and it is followed by an

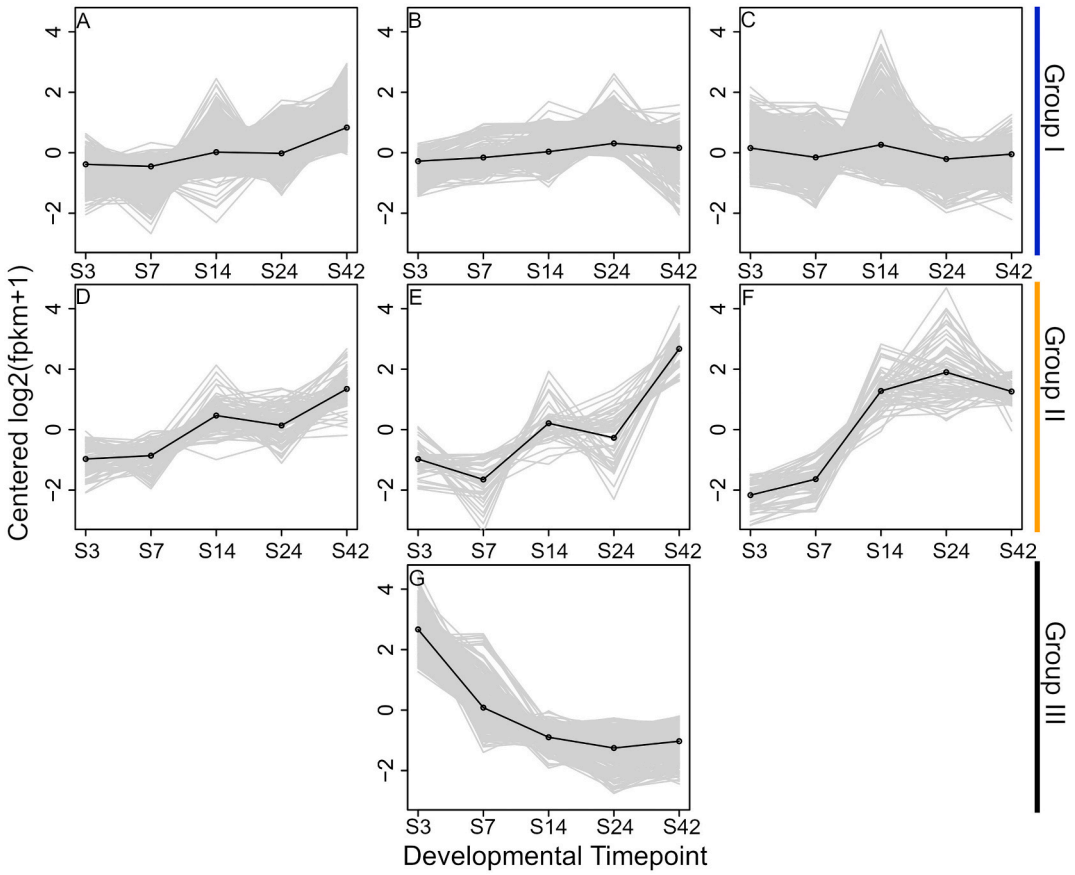


Fig. 2. *Acartia tonsa*. Clusters resulting from the DE analysis performed by edgeR. The x axis represents the timepoints studied. These being: 16–32 cell stage (S3), gastrulation (S7), organogenesis (S14), limb bud formation (S24), final nauplii (S42). The grey lines in the clusters represent each one of the transcripts included in it and the black line shows the average expression for each one of the timepoints. Finally, the expression change is plotted on the y axis as a log2(FPKM+1) change. On the right side the grouping of the clusters according to their behavior can be seen. With Group I (blue) conformed by clusters A, B, and C; Group II (orange): clusters D, E, and F; Group III (black): cluster G.

important decrease in the expression occurring mainly at 7 h and to a lesser degree at 14 h. After this, the expression is stabilized at reduced values. The behavior is opposite to the one observed in cluster F. Among the sequences found in this group are those annotated to genes/proteins associated with survival of the early embryo (*rdx*), the cell (*dapk1*), and transcriptional activators (*srfb1*).

The complete list of transcripts included in the 7 clusters is included as Supplementary Table III (Excel file).

3.3. Transcript analysis

We plotted the differentially expressed (DE) transcripts in Volcano plots (Fig. 3). For the volcano plots the time points involved in the biggest changes in expression values according to the cluster analysis (3 h, 14 h, and 42 h) were selected. Finally, a fourth volcano plot was done to explore the differences between 14 h and 24 h, as the latter seemed to have lower expression values than 14 h in most sequences according to the clusters (Fig. 2). To achieve a better understanding of subitaneous development, 12 representative transcripts annotated to relevant genes or proteins from other organisms, significantly DE at some stages of development were selected for labeling: Fig. 3.

The volcano plots further illustrate the observations of the cluster analysis, now more clearly as fold change is easier to observe and individual transcripts can be tracked. Fig. 3A and B illustrates the highest DE that occurs throughout the analyzed embryological timepoints. These are between 42 h and 3 h (Fig. 3A) (34,873 transcripts); as well as 3 h and 7 h (Fig. 3B) (30,657 transcripts). The analysis shows that most of the transcripts are upregulated at the later stages of development (such as sequestosome (*sqstm*), *scarf*, *cup8*, *spz2*, *l2efl*, nucleoredoxin like 2 (*nxnl2*)) and only few transcripts are upregulated at 3 h (death-associated protein kinase 1 (*dapk1*), serum response factor-binding protein 1 (*srfb1*), *rdx*, *uaf30*, *wnt8*).

Fig. 3C represents the pairwise analysis of timepoints 24 h and 14 h (31,630 transcripts). In this case most of the selected sequences are either not DE at all or not in a significant way. The only DE transcript at 24 h is spz2 while *nxnl2*, *scarf*, *sqstm*, and *l(2)efl* remain upregulated at 14 h. The plot reveals that few sequences are upregulated at 24 h, and the fold change of those is low, with just four sequences being >5-fold upregulated. This is a big contrast with the high count of transcripts meeting this criterion in Fig. 3A, B, D. The behavior for 14 h in the plot is similar to that observed in the previous plot (Fig. 3B) with *sqstm* and *l(2)efl* being among the most upregulated transcripts. With this plot (Fig. 3C) the observation that had been done with the cluster analysis (Fig. 2D & E) regarding the decrease of expression values at 24 h is further analyzed and confirmed.

Fig. 3D shows that 14 h and 42 h share almost the same expression pattern regardless of the time-gap observed. Most of the sequences behave in the same way among both samples with a minimum number of DE transcripts shown, among which none of the selected sequences are found. Moreover, the average p-values for the significant DE transcripts are much lower than in previous analysis.

Finally, the same transcripts highlighted in the volcano plots were plotted individually in longitudinal plots (Fig. 4), which illustrate the expression patterns previously established by the clustering analysis and that changes in the expression pattern varied between 2 and 4-fold up- or down-regulation. With Fig. 4 it is also possible to study the standard

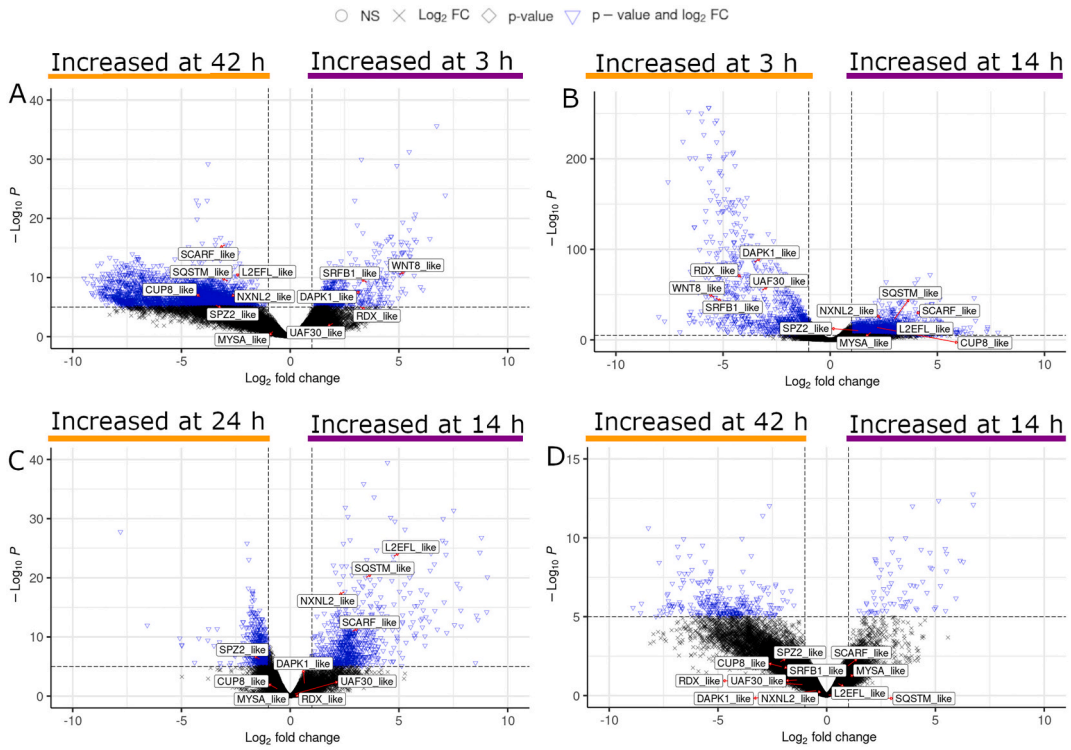


Fig. 3. *Acartia tonsa*. Volcano plots showing the DE sequences between timepoints. A: Final nauplii vs 16–32 cell stage, B: 16–32 cell stage vs organogenesis, C: limb bud formation vs organogenesis, D: final nauplii vs organogenesis. The same selected sequences as in Fig. 2 are labeled in the four plots. If a sequence is not labeled in a certain plot, it is because it was not DE in that pairwise analysis. The blue triangles represent the sequences which were significant in both p-value and fold change. The non-significant sequences are plotted in black. The y axis represents the significance level of each observation in a logarithmic scale. Note that the figures on the right side (B & D) have a different scale in this axis to properly illustrate the significance. The x axis shows the fold change in the sequences in a log2 scale.

error associated to the replicates and it can be observed that these are minimal for most transcripts and timepoints. Four of the transcripts (*sqstm*, *nxnl2*, *l(2)efl*, *scarf*) also display the significant decrease in the expression that seems to occur at 24 h. Finally, it can be observed that most of the transcripts start with low expression and only begin to gain higher values after organogenesis initiates at 14 h.

3.4. Enrichment analysis

The analysis of enriched biological processes (Fig. 5) shows that many of the enriched pathways belong to the final stages of embryological development. The most significant processes enriched at the initial stages of embryogenesis (4–16 cell stage (3 h), gastrulation (7 h)) are related to wnt signaling and an increase in the transcription machinery (Fig. 5). These are driven by different transcripts annotated to the wnt family and transcription initiators/factors (*uaf30*, *srfb1*), whose behavior are shown in Fig. 4.

The next embryological stages (14 h and 24 h) participate in organogenesis and limb bud formation and are dominated by muscle formation processes (Fig. 5). Several transcripts annotated to microfibrils, and microtubules components are found at these two time points as the main components for the enrichment analysis. Among the transcripts the heavy chains of muscle myosin (*mysa_drome* in Fig. 4) seems to be one of the main drivers of this enrichment.

The stage of final nauplii ready to hatch (42 h) shows many of the enriched processes observed in the embryogenesis like 14 h (Fig. 5). Among the enriched processes a wide variety of them can be observed with no specific tissue/process being enriched. The prominent observed processes are cell to cell communication, chitin metabolism processes (*cup8*), fat cell proliferation (*sqstm*), establishment of location and ion transport, reflecting the more specialized functions of the developing nauplii.

4. Discussion

To our knowledge, this is the first time an assembled transcriptome of the embryological development of a copepod species has been used to describe the process on a molecular scale. This constitutes a potentially powerful tool to bring light onto the otherwise unknown process of embryonic development in the model calanoid *Acartia tonsa*.

Some limitations may be raised regarding the raw reads used to assemble the transcriptome and its quality, especially due to the observed strandedness, as well as the high percentage of duplication together with the low percentage of GC content that can be observed throughout the samples. These values are not necessarily problematic as these are also dependent on the animal species and can range broadly between different animal species.

Regarding the strange values observed for strandedness, this should not be a primary reason for concern, as the large genome size of *A. tonsa* leads to a small coverage of non-mRNA areas. Eventually, this is translated into a small number of false transcripts originated as an artifact.

Moreover, the GC values remain constant throughout the different libraries (Table 1) and show a high value in contrast with the 30 % GC content of *A. tonsa* genome. Although the 10 % decrease in the assembly (Table 2) could indicate DNA contamination in the assembly the values remain higher to those of the genome and remain like those reported in the other available transcriptomes (Nilsson et al., 2018; Jørgensen et al., 2019c) for *A. tonsa* favors the idea that the observation is representing a biological truth. Although RNA degradation could also be considered as the cause of this observation it is not considered in this article as that would have affected other metrics as well and this is not the case.

For the quality of the assembly itself (Table 2), two measurements were used as main indicators: N50, which showed low values indicating a long total assembly length; and BUSCO which not only yielded very good results but was also in accordance, and outperformed, previously

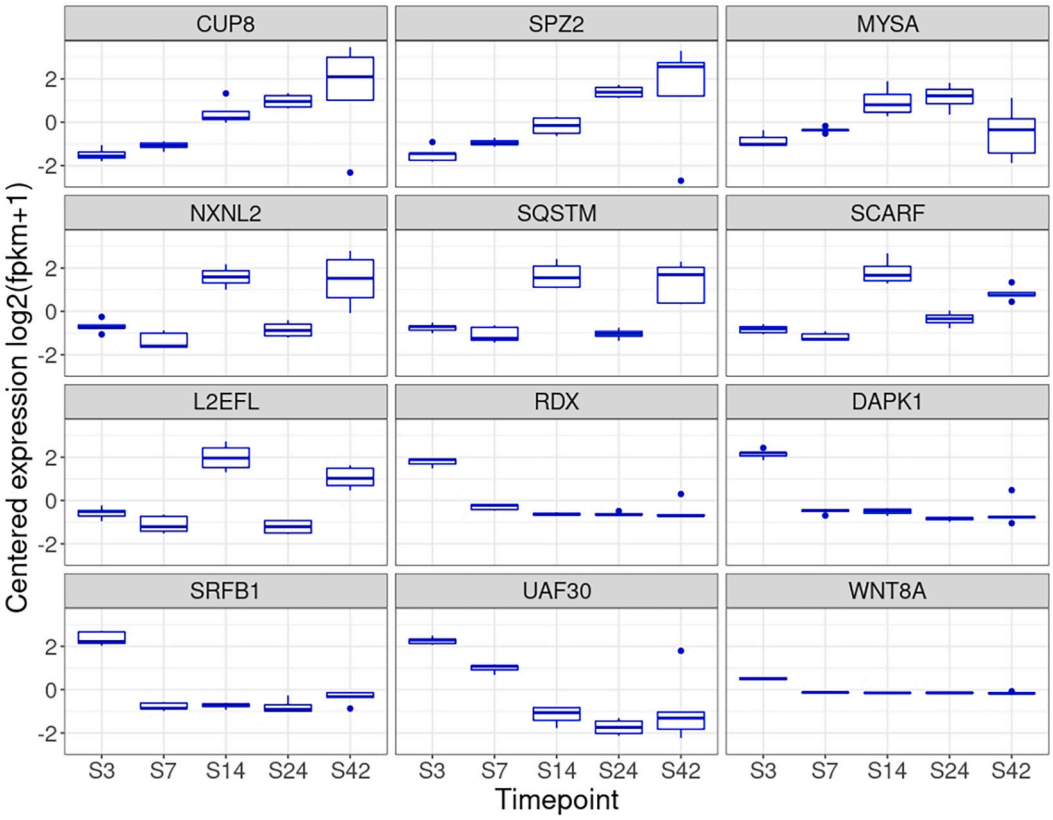


Fig. 4. *Acartia tonsa*. Individual boxplots of the selected sequences with the standard error for each timepoint; 16–32 cell stage (S3), gastrulation (S7), organogenesis (S14), limb bud formation (S24), final nauplii (S42). The expression value is calculated as log2(FPKM+1). On top of each of the sequences is shown the name of the gene/protein/feature they are annotated to.

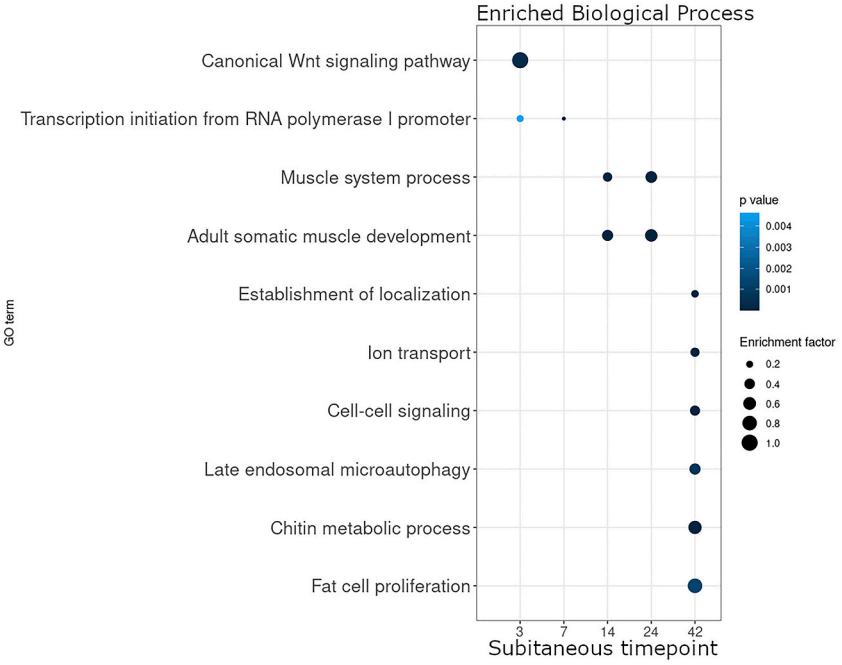


Fig. 5. *Acartia tonsa*. Bubble plot showing the 10 most significant biological processes enriched in subitaneous development and their enrichment factor through the five timepoints studied which represent: 16–32 cell stage, gastrulation, organogenesis, limb-bud formation, and final nauplii ready to hatch, respectively. The color of the bubbles shows the p-value associated to the process and the size represents the enrichment factor of the biological factor compared to the one expected from its representation across the transcriptome.

published research, showing 5 % less missing and 10 % and 5 % more complete BUSCOs (**HAGX01** & **GFWY00000000.1**, respectively) (Nilsson et al., 2018; Jørgensen et al., 2019c). Regarding N50 values, it must be commented that in transcriptomics, this metric is not an indicator as strong as it is on genomes and therefore can be misleading. Nevertheless, we decided to include it as it is still reported commonly in transcriptome assemblies (e.g., Li et al., 2014). Median contig length is another metric that could be considered problematic for the assembly, especially when

compared with bigger values of the other two assemblies (Table 2). The median being driven down does not mean however that there is a smaller amount of long contigs, as it can be seen on the N10 of the same table, which is 1000 nucleotides bigger than for **HAGX01**, the other transcriptome that has embryogenesis representation. The median deviation then, should not be a reason for concern. The cumulative length report further supports this statement showing that there are indeed more long transcripts in the current assembly (**GSE2105554**) than in the previous ones (Supplementary Fig. II). Overall, and although 1x75nt transcriptomes are not generally used for a reference assembly, the quality of the assembly is offset by a greater sequence diversity and possibly by the life stages included in this transcriptome but missing or rare in the other two. Thus, the assembly is considered adequate to explain the molecular changes occurring throughout the different embryonic developmental stages.

The analysis in the DE transcripts has shown that sequences annotated to widely studied developmental genes from other metazoans can be found in *A. tonsa* embryogenesis as it is the case for *wnt8*, spaetzle, protein lethal, protein roadkill or scarface (Kurzik-Dumke and Lohmann, 1995; Kent et al., 2006; Grabherr et al., 2011; Parthier et al., 2014; Kushnir et al., 2017). Nevertheless, and as expected for any biological process as embryogenesis, developmental genes are not the only type of transcript observed in the transcriptome. Other differentially regulated transcripts were observed with their function retrieved from Uniprot site (Bateman et al., 2021): myosin heavy chains (*mysa*), transcriptional regulators (*srfb1*, *nxn*), or chaperones (*sqstm*) also seem to be greatly regulated at some timepoints.

Most of the above-mentioned transcripts share a trend of increased expression as embryogenesis progresses. Only few transcripts start high and later decrease their expression, among these sequences it is of notice to focus on clusters F and G, which seem to have an opposite behavior. This suggests some sort of co-regulation of the sequences, probably influenced by developmental regulators. The sequences in cluster G may be important during the early stages of embryogenesis and not necessary for the later. This idea is partially supported by the presence of a sequence annotated to the protein roadkill (rdx), which via protein ubiquitination engages in segment polarity and hedgehog signaling in *D. melanogaster* development. Roadkill is only upregulated in the early stages of embryogenesis in *A. tonsa*, indicating a similar function as in the fruit fly (Kent et al., 2006). Other sequences upregulated at the early stages of embryogenesis are those annotated to the Wnt family involved in segment polarity. These two transcripts being upregulated in the initial stages suggest that the initial segmentation of the embryo, probably establishment of the body axis, occurs during the 4–16 cell stages.

The 7 h time point is the moment that marks both gastrulation and the beginning of differences between subitaneous and quiescent development (Nilsson and Hansen, 2018; Acebal et al., 2022). Nevertheless, there are no characteristic processes associated with this time point, neither does it show a distinct behavior from the 3 h time point which suggests that the regulation could be occurring at a post transcriptional level. Additionally, the resolution of the data might be not clear enough, and the transcripts may last less than the 1-hour sampling timespan.

Among all the clusters, the biggest change in regulation of the transcripts, occurs between gastrulation and organogenesis. The later supposes a major change in behavior of the transcripts by either augmenting or decreasing their expression. The change is stabilized afterwards and while some variability can be observed at limb bud formation stage, most of the sequences do not show this behavior. For this reason, we propose organogenesis to be a vital moment for the embryonic development of *A. tonsa* at least transcription wise. Nevertheless, the enrichment analysis does not seem to support this observation as it only shows enrichment of muscle/cytoskeleton processes. In this case, we hypothesize that the large number of transcripts with this annotation could be affecting the enrichment calculations and although they dominate as the most enriched processes this does not necessarily reflect the total biological significance as both the volcano plots and the individual transcripts show that in this dataset there are few gene expression differences between 14 and 42 h (organogenesis and final nauplius).

Even though only few transcripts are decreased at the limb bud stage, some of the selected sequences such as nxnl2, l(2)efl or sqstm, display a significant downregulation at 24 h even if they are upregulated at both the previous and next timepoint. The reason for the drastic downregulation of specific transcripts at 24 h is unknown and very striking as these transcripts usually gain back the previous values once the next stage is reached (nxnl2, scarf, l(2)efl, and sqstm). This observation has raised our interest for its behavior, and we recommend it to be followed up in future studies to acquire a better knowledge on the processes that lead to this scenario in the embryo expression pattern, a higher resolution (e.g., more timepoints) could help elucidate the matter.

Finally, the last stage of embryonic development analyzed is the final nauplii ready to hatch (42 h). At this time point embryonic development has reached its conclusion and soon a new stage of development will begin, the free-living copepod. The behavior of most of the genes observed in the previous stage is consolidated and perhaps the most interesting observation is the presence of chitin binding domains in the DE sequences and the emergence of transcripts encoding for proteins involved in the formation of the cuticle in pupal stages of *D. melanogaster* as it is the case for Cup8 (Apple and Fristrom, 1991; Karouzou et al., 2007).

Overall, we can observe a progression in *A. tonsa* embryonic development from an early stage with an increase in transcriptional machinery and some initial cell fate and axis determination followed by the genesis of different vital structures in the embryo such as limbs or muscles. At the same time, a bigger number of transcripts is present at later embryological stages than at the early ones. The observed behavior for the enrichment analysis where many processes are only present at one timepoint, could have been caused due to the lack of annotation of key genes in said processes (e.g., wnt family) thus only receiving a signal at certain moments.

Finally, the data resulting from the analysis will help establishing the embryological stages defined and visualized by Nilsson and Hansen more firmly (Nilsson and Hansen, 2018). Moreover, we hope to lay the foundation for future research in *A. tonsa* embryonic development. The present and future research will allow a comparison of *A. tonsa* and other copepods embryonic development with other crustaceans like *Daphnia magna* or other arthropods used in traditional developmental studies like the fruit fly *Drosophila melanogaster*. Such studies would increase the use of *A. tonsa* as a model organism for developmental biology in crustaceans or even Arthropoda. As a model organism, *Acartia* offers the knowledge already available at the ecological level, together with the cheap maintenance of cultures, short life cycle and embryogenesis, and the high economic prospect of copepods as live feed in aquaculture that has risen in the past years. At the same time, we recommend to further focus on research into some of the interesting results observed such as the downregulation of transcripts relevant for limb bud formation (24 h), previously unappreciated in terms of transcription during the period of gastrulation in the embryonic development.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cbd.2022.101054.

Conflict of interests

The authors have no declaration of interest and no conflicts with anyone.

Data availability

Data will be made available on request.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The research has followed the EU directive 2010/63/EU for animal experiments. If desired, access to data can be obtained by contacting the corresponding author.

References

Acebal, M.C., Daalgard, L.T., Hansen, B.W., Jørgensen, T.S., 2022. Transcriptome comparison between subitaneous and quiescent embryogenesis in the calanoid copepod Acartia tonsa. Pending publication.

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., Blankenberg, D., 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 46, W537–W544. https://doi.org/10.1093/NAR/GKY379.

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.

Andrews, S., et al., 2010. FastQC: a quality control tool for high throughput sequence data, 2010. Https://Www.Bioinformatics.Babraham.Ac.Uk/Projects/Fastqc/.

Apple, R.T., Fristrom, J.W., 1991. 20-Hydroxyecdysone is required for, and negatively regulates, transcription of Drosophila pupal cuticle protein genes. Dev. Biol. 146, 569–582. https://doi.org/10.1016/0012-1606(91)90257-4.

Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C.L., Serova, N., Davis, S., Soboleva, A., 2013. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Res. 41, D991–D995. https://doi.org/10.1093/NAR/GKS1193.

Bateman, A., Martin, M.J., Orchard, S., Magrane, M., Agivetova, R., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bursteinas, B., Bye-A-Jee, H., Coetzee, R., Cukura, A., Silva, A.Da., Denny, P., Dogan, T., Ebenezer, T.G., Fan, J., Castro, L.G., Garmiri, P., Georghiou, G., Gonzales, L., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., Insana, G., Ishtiaq, R., Jokinen, P., Joshi, V., Jyothi, D., Lock, A., Lopez, R., Luciani, A., Luo, J., Lussi, Y., MacDougall, A., Madeira, F., Mahmoudy, M., Menchi, M., Mishra, A., Moulang, K., Nightingale, A., Oliveira, C.S., Pundir, S., Qi, G., Raj, S., Rice, D., Lopez, M.R., Saidi, R., Sampson, J., Sawford, T., Speretta, E., Turner, E., Tyagi, N., Vasudev, P., Volynkin, V., Warner, K., Watkins, X., Zaru, R., Zellner, H., Bridge, A., Poux, S., Redaschi, N., Aimo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M.C., Bolleman, J., Boutet, E., Breuza, L., Casals-Casas, C., de Castro, E., Echioukh, K.C., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Estreicher, A., Famiglietti, M.L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Keller, G., Kerhornou, A., Lara, V., Le Mercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T.B., Paesano, S., Pedruzzi, I., Pilbout, S., Pourcel, L., Pozzato, M., Pruess, M., Rivoire, C., Sigrist, C., Sonesson, K., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Wu, C.H., Arighi, C.N., Arminski, L., Chen, C., Chen, Y., Garavelli, J.S., Huang, H., Laiho, K., McGarvey, P., Natale, D.A., Ross, K., Vinayaka, C.R., Wang, Q., Wang, Y., Yeh, L.S., Zhang, J., da Silva, A., Denny, P., Dogan, T., Ebenezer, T.G., Fan, J., Castro, L.G., Garmiri, P., Georghiou, G., Gonzales, L., Hatton-Ellis, E., Hussein, A., Ignatchenko, A., Insana, G., Ishtiaq, R., Jokinen, P., Joshi, V., Jyothi, D., Lock, A., Lopez, R., Luciani, A., Luo, J., Lussi, Y., MacDougall, A., Madeira, F., Mahmoudy, M., Menchi, M., Mishra, A., Moulang, K., Nightingale, A., Oliveira, C.S., Pundir, S., Qi, G., Raj, S., Rice, D., Lopez, M.R., Saidi, R., Sampson, J., Sawford, T., Speretta, E., Turner, E., Tyagi, N., Vasudev, P., Volynkin, V., Warner, K., Watkins, X., Zaru, R., Zellner, H., Bridge, A., Poux, S., Redaschi, N., Aimo, L., Argoud-Puy, G., Auchincloss, A., Axelsen, K., Bansal, P., Baratin, D., Blatter, M.C., Bolleman, J., Boutet, E., Breuza, L., Casals-Casas, C., de Castro, E., Echioukh, K.C., Coudert, E., Cuche, B., Doche, M., Dornevil, D., Estreicher, A., Famiglietti, M.L., Feuermann, M., Gasteiger, E., Gehant, S., Gerritsen, V., Gos, A., Gruaz-Gumowski, N., Hinz, U., Hulo, C., Hyka-Nouspikel, N., Jungo, F., Keller, G., Kerhornou, A., Lara, V., Le Mercier, P., Lieberherr, D., Lombardot, T., Martin, X., Masson, P., Morgat, A., Neto, T.B., Paesano, S., Pedruzzi, I., Pilbout, S., Pourcel, L., Pozzato, M., Pruess, M., Rivoire, C., Sigrist, C., Sonesson, K., Stutz, A., Sundaram, S., Tognolli, M., Verbregue, L., Wu, C.H., Arighi, C.N., Arminski, L., Chen, C., Chen, Y., Garavelli, J.S., Huang, H., Laiho, K., McGarvey, P., Natale, D.A., Ross, K., Vinayaka, C.R., Wang, Q., Wang, Y., Yeh, L.S., Zhang, J., 2021. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. 49, D480–D489. https://doi.org/10.1093/NAR/GKAA1100.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57, 289–300. https://doi.org/10.1111/J.2517-6161.1995.TB02031.X.

Berggreen, U., Hansen, B., Kiørboe, T., 1988. Food size spectra, ingestion and growth of the copepodAcartia tonsa during development: Implications for determination of copepod production. Mar. Biol. 99, 341–352. https://doi.org/10.1007/BF02112126.

Bretaudeau, A., Le Corguillé, G., Corre, E., Liu, X., 2021. De novo transcriptome assembly, annotation, and differential expression analysis (Galaxy Training Materials) [WWW Document]. URL. https://training.galaxyproject.org/training-material/topics/transcriptomics/tutorials/full-de-novo/tutorial.html (accessed 2.15.22).

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. R., 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/ATTACHMENT/B24F06B8-7AC7-45AF-B605-0DA3B55AE55E/MMC11.ZIP.

Combs, P.A., Eisen, M.B., 2015. Low-cost, low-input RNA-seq protocols performnearly as well as high-input protocols. PeerJ 2015, e869. https://doi.org/10.7717/PEERJ.869/SUPP-5.

Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. {MultiQC}: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354.

Flaticon, 2022. Iconos vectoriales y stickers - PNG, SVG, EPS, PSD y CSS [WWW Document]. URL. https://www.flaticon.es/ (accessed 7.29.22).

Freiburg Galaxy Team, 2021. Galaxy Europe [WWW Document]. https://usegalaxy.eu/.

Galachyants, Y.P., Zakharova, Y.R., Volokitina, N.A., Morozov, A.A., Likhoshway, Y.V., Grachev, M.A., 2019. De novo transcriptome assembly and analysis of the freshwater araphid diatom Fragilaria radians, Lake Baikal. Sci. Data 6, 1–11. https://doi.org/10.1038/s41597-019-0191-6.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A.J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J.Y., Zhang, J., 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5, 1–16. https://doi.org/10.1186/GB-2004-5-10-R80.

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., Mauceli, 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.

Guillard, R.R.L., 1975. Culture of phytoplankton for feeding marine invertebrates. Cult. Mar. Invertebr. Anim. 29–60 https://doi.org/10.1007/978-1-4615-8714-9_3.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8 (8), 1494–1512. https://doi.org/10.1038/nprot.2013.084.

Hansen, B.W., 2019. Copepod embryonic dormancy: “An egg is not just an egg”. Biol. Bull. 237, 145–169. https://doi.org/10.1086/705546.

HMMER [WWW Document]. n.d. URL. http://hmmer.org/ (accessed 3.25.22).

Holm, M.W., Kiørboe, T., Brun, P., Licandro, P., Almeda, R., Hansen, B.W., 2018. Resting eggs in free living marine and estuarine copepods. J. Plankton Res. 40, 2–15. https://doi.org/10.1093/PLANKT/FBX062.

Hölzer, M., Marz, M., 2019. De novo transcriptome assembly: a comprehensive cross-species comparison of short-read RNA-Seq assemblers. Gigascience 8. https://doi.org/10.1093/GIGASCIENCE/GIZ039.

Home · Trinotate/Trinotate.github.io Wiki · GitHub [WWW Document]. URL. https://github.com/Trinotate/Trinotate.github.io/wiki (accessed 3.15.22).

Jørgensen, T.S., Jepsen, P.M., Petersen, H.C.B., Friis, D.S., Hansen, B.W., 2019a. Eggs of the copepod Acartia tonsa Dana require hypoxic conditions to tolerate prolonged embryonic development arrest. BMC Ecol. 19, 1–9. https://doi.org/10.1186/S12898-018-0217-5/FIGURES/2.

Jørgensen, T.S., Nielsen, B.L.H., Petersen, B., Browne, P.D., Hansen, B.W., Hansen, L.H., 2019b. The whole genome sequence and mRNA transcriptome of the tropical cyclopoid copepod Apocyclops royi. G3 9, 1295–1302. https://doi.org/10.1534/G3.119.400085 (Bethesda).

Jørgensen, T.S., Petersen, B., Petersen, H.C.B., Browne, P.D., Prost, S., Stillman, J.H., Hansen, L.H., Hansen, B.W., 2019c. The genome and mRNA transcriptome of the cosmopolitan calanoid copepod Acartia tonsa Dana improve the understanding of copepod genome size evolution. Genome Biol. Evol. 11, 1440. https://doi.org/10.1093/GBE/EVZ067.

Karouzou, M.V., Spyropoulos, Y., Iconomidou, V.A., Cornman, R.S., Hamodrakas, S.J., Willis, J.H., 2007. Drosophila cuticular proteins with the R&R Consensus: annotation and classification with a new tool for discriminating RR-1 and RR-2 sequences. Insect Biochem. Mol. Biol. 37, 754–760. https://doi.org/10.1016/J.IBMB.2007.03.007.

Kent, D., Bush, E.W., Hooper, J.E., 2006. Roadkill attenuates Hedgehog responses through degradation of Cubitus interruptus. Development 133, 2001–2010. https://doi.org/10.1242/DEV.02370.

Kurzik-Dumke, U., Lohmann, E., 1995. Sequence of the new Drosophila melanogaster small heat-shock-related gene, lethal(2) essential for life [l(2) efl], at locus 59F4,5. Gene 154, 171–175. https://doi.org/10.1016/0378-1119(94)00827-F.

Kushnir, T., Mezuman, S., Bar-Cohen, S., Lange, R., Paroush, Z., Helman, A., 2017. Novel interplay between JNK and Egfr signaling in Drosophila dorsal closure. PLoS Genet. 13 https://doi.org/10.1371/JOURNAL.PGEN.1006860.

LesLaboratories Servier SAS, 2020. SMART Servier Medical Art [WWW Document]. URL. SMART Servier Med. Art. (accessed 729.22). https://smart.servier.com/

Li, B., Fillmore, N., Bai, Y., Collins, M., Thomson, J.A., Stewart, R., Dewey, C.N., 2014. Evaluation of de novo transcriptome assemblies from RNA-Seq data. Genome Biol. 15, 553. https://doi.org/10.1186/S13059-014-0553-5.

Liao, Y., Smyth, G.K., Shi, W., 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. https://doi.org/10.1093/BIOINFORMATICS/BTT656.

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D., Bateman, A., 2021. Pfam: the protein families database in 2021. Nucleic Acids Res. 49, D412–D419. https://doi.org/10.1093/NAR/GKAA913.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5, 621–628. https://doi.org/10.1038/nmeth.1226.

Nielsen, H., Tsirigos, K.D., Brunak, S., von Heijne, G., 2019. A brief history of protein sorting prediction. Protein J. 38, 200–216. https://doi.org/10.1007/S10930-019-09838-3.

Nilsson, B., Hansen, B.W., 2018. Timing of embryonic quiescence determines viability of embryos from the calanoid copepod, Acartia tonsa (Dana). PLoS One 13, 1–16. https://doi.org/10.1371/journal.pone.0193727.

Nilsson, B., Jepsen, P.M., Bucklin, A., Hansen, B.W., 2018. Environmental stress responses and experimental handling artifacts of a model organism, the Copepod Acartia tonsa (Dana). Front. Mar. Sci. 5, 156. https://doi.org/10.3389/FMARS.2018.00156/BIBTEX.

Nilsson, B., Jepsen, P.M., Rewitz, K., Hansen, B.W., 2014. Expression of hsp70 and ferritin in embryos of the copepod Acartia tonsa (Dana) during transition between subitaneous and quiescent state. J. Plankton Res. 36, 513–522. https://doi.org/10.1093/plankt/fbt099.

Parthier, C., Stelter, M., Ursel, C., Fandrich, U., Lilie, H., Breithaupt, C., Stubbs, M.T., 2014. Structure of the Toll-Spätzle complex, a molecular hub in Drosophila development and innate immunity. Proc. Natl. Acad. Sci. U. S. A. 111, 6281–6286. https://doi.org/10.1073/PNAS.1320678111/-/DCSUPPLEMENTAL.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139. https://doi.org/10.1093/BIOINFORMATICS/BTP616.

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.

Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 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.

Støttrup, J.G., Richardson, K., Kirkegaard, E., Pihl, N.J., 1986. The cultivation of Acartia tonsa Dana for use as a live food source for marine fish larvae. Aquaculture 52, 87–96. https://doi.org/10.1016/0044-8486(86)90028-1.

Stubben, C., 2016. trinotateR.

The GIMP Development Team, n.d.The GIMP Development Team , n.d. GIMP.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28, 511–515. https://doi.org/10.1038/nbt.1621.

Wang, L., Wang, S., Li, W., 2012. RSeQC: quality control of RNA-seq experiments. Bioinformatics 28, 2184–2185. https://doi.org/10.1093/BIOINFORMATICS/BTS356.

Wickham, H., Averick, M., Bryan, J., Chang, W., D', L., Mcgowan, A., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Lin Pedersen, T., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H., 2019. Welcome to the Tidyverse. J. Open Source Softw. 4, 1686. https://doi.org/10.21105/JOSS.01686.

Young, M.D., Wakefield, M.J., Smyth, G.K., Oshlack, A., 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 11, 1–12. https://doi.org/10.1186/GB-2010-11-2-R14/TABLES/4.

Zhang, Y., Park, C., Bennett, C., Thornton, M., Kim, D., 2021. Rapid and accurate alignment of nucleotide conversionsequencing reads with HISAT-3N. Genome Res. 31, 1290–1295. https://doi.org/10.1101/GR.275193.120.