



OPEN

DATA DESCRIPTOR

# *De novo* transcriptome assembly of *Acartia tonsa* adults using Nanopore long-read sequencing

Florencia Mohamed<sup>1</sup>, Marco Picone<sup>2</sup>, Greta Battaglia<sup>1</sup>, Ilenia Urso<sup>1</sup>, Elisa Camatti<sup>3</sup>,  
Alessandro Vezzi<sup>1</sup>✉ & Gabriele Sales<sup>1</sup>✉

*Acartia tonsa* is a calanoid copepod with a cosmopolitan distribution. Its ecological relevance makes it a useful bioindicator organism for assessing the toxicity of various compounds. However, transcriptomic assemblies of *A. tonsa* using long-read technologies have not yet been described. The use of long-read sequencing technologies in transcriptomics allows the study of alternative splicing, structural variations and alternative polyadenylation sites. In this study, we present a *de novo* transcriptome of *A. tonsa* adult copepods exposed to different neonicotinoids obtained from Nanopore sequences. This transcriptome (261,560 total transcripts, with 31,291 representative sequences) exhibits 88.3% completeness and an N50 of 2,580 bases, showing better results than previous assemblies of the same organism. We also performed a full annotation by sequence homology (NR database), domain identification (InterProScan) and functional classification (Gene Ontology); 54.3% of representative transcripts were annotated in at least one database. Our transcriptome represents a solid baseline for further transcriptomic studies on *A. tonsa* and, specifically, its response to currently used pesticides.

## Background & Summary

Copepods are a primary component of zooplankton assemblages in marine ecosystems. They occupy a critical trophic position in the marine food web since they graze on the phytoplankton, consume detritus, and serve as a food reserve for fish and invertebrates<sup>1</sup>. Inside this group, *Acartia tonsa* Dana (1849) is a marine, euryhaline calanoid copepod with a cosmopolitan neritic distribution, often the dominant zooplankton species in several ecosystems<sup>2</sup>. Due to its worldwide distribution, easy culturing, short generation times and ecological relevance, *A. tonsa* is a useful bioindicator organism for climate-driven ecosystem variability<sup>3</sup> and for assessing the toxicity of several compounds<sup>4–6</sup>. This organism has been successfully used in previous studies to assess the effect of neonicotinoid pesticides due to its high sensitivity caused by its nervous system's similarity with insects<sup>7,8</sup>.

Neonicotinoids (NEOs) are neurotoxic pesticides that disrupt synaptic transmission by acting as nicotinic acetylcholine receptors (nAChRs) agonists<sup>9</sup>. In the current study, we employed the first-generation NEOs acetamiprid, imidacloprid, and thiacloprid and the second-generation NEOs clothianidin and thiamethoxam.

Despite the ecological importance of *A. tonsa*, only one genomic and a few transcriptomic assemblies of this organism are available in the NCBI database<sup>2,10–13</sup>. The largest transcriptome of *A. tonsa* until the time of writing (April 2025) that includes the adult stage of this copepod was previously obtained by Jørgensen *et al.*<sup>2</sup>. These authors performed a *de novo* assembly from Illumina reads covering all life stages (embryos, nauplii/copepodites and adults), obtaining a 118,709,440-base transcriptome with 61,149 representative transcripts and 56,257 additional isoforms. However, a *de novo* transcriptome assembly of eukaryotic organisms from short RNA-Seq data is severely limited, as this type of reads does not allow accurate reconstruction of full-length transcripts and transcriptional isoforms<sup>14</sup>. For this reason, long-read sequencing technologies are increasingly being used to unravel the complex nature of transcriptomes, especially when no high-quality genome sequence is available<sup>15,16</sup>.

PacBio and Nanopore sequencing technologies achieve read lengths of around 15 kb and over 30 kb, respectively. These technologies enable the study of full-length sequences, alternative splicing, structural variation and alternative polyadenylation sites, facilitating genome annotation and gene function studies<sup>17,18</sup>.

<sup>1</sup>Department of Biology, University of Padua, Padua, Italy. <sup>2</sup>Department of Environmental Sciences, Ca' Foscari University of Venice, Venice, Italy. <sup>3</sup>Institute of Marine Science, National Research Council (CNR ISMAR), Venice, Italy. ✉e-mail: [alessandro.vezzi@unipd.it](mailto:alessandro.vezzi@unipd.it); [gabriele.sales@unipd.it](mailto:gabriele.sales@unipd.it)

In this study, we present a *de novo* transcriptome of *A. tonsa* adult copepods exposed to different NEOs obtained from Nanopore sequences. To our knowledge, this is the first report of a *de novo* transcriptome of this organism produced using a long-read sequencing strategy. Our assembly exhibits an estimated completeness of 88.3% and includes 261,560 transcripts with a mean length of 2,034 bases. It provides a solid baseline for future studies on the response of *A. tonsa* adults to different environmental stressors. Leveraging this dataset, we will study the differential expression of genes in *A. tonsa* adults exposed to different NEOs at different concentrations to identify biomarkers of stress and to support the development of early warning tools for marine ecosystem health.

## Methods

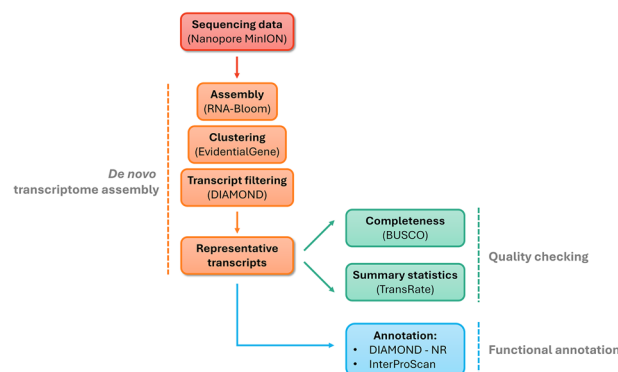
**Acartia tonsa culturing.** The biological material used for the experiments was obtained from in-house laboratory cultures kept at the Ca' Foscari University of Venice<sup>5,7,8</sup>. Stock cultures of *A. tonsa* ( $n = 600\text{--}800$  individuals) were maintained at  $20 \pm 2^\circ\text{C}$  in climatic rooms under a 16 h: 8 h light:dark photoperiod and continuous aeration using a 20‰ salinity medium prepared according to the ISO 16778 standard method<sup>19</sup>. The copepods were fed *ad libitum* with a mixture of three microalgae (*Pavlova lutheri*, *Tetraselmis suecica*, and *Tisochrysis lutea*) dosed four times per day using a peristaltic pump controlled by a timer. The *A. tonsa* cultures were cleaned three times a week by siphoning off the water from the bottom of the culture flasks (2 L glass bottles) and filtering the siphoned medium using two sieves with mesh sizes of 170  $\mu\text{m}$  and 50  $\mu\text{m}$ , respectively. This procedure allowed the separation of the adult copepods and copepodites (retained by the 170  $\mu\text{m}$  mesh size sieve), from the eggs and nauplii (passing through the 170  $\mu\text{m}$  mesh but retained by the 50  $\mu\text{m}$  mesh size sieve) and wastes (faecal pellets, aged algae, detritus). Adult copepods were reintroduced into the culture, while eggs were collected and stored separately at  $4^\circ\text{C}$  for later testing. Each stock culture was maintained for 6–7 weeks before starting new cultures.

**Experimental design.** The exposure experiments were conducted using eggs produced by in-house cultures within 24 hours before the test. The eggs were carefully counted under a dissecting microscope, and only healthy, blue-coloured eggs were selected. Approximately 400–500 eggs were then placed in 1 L Erlenmeyer glass flasks filled with the test solutions: a negative control (20‰ salinity medium) and two different concentrations of each NEO (acetamiprid, 10 and 81  $\text{ng L}^{-1}$ ; clothianidin, 8 and 62  $\text{ng L}^{-1}$ ; thiacloprid, 9 and 84  $\text{ng L}^{-1}$ , thiamethoxam, 8 and 84  $\text{ng L}^{-1}$ ; imidacloprid, 1 and 10  $\text{ng L}^{-1}$ ) showing no effects on the larval development of *A. tonsa*<sup>7,8</sup> were tested.

The exposure was performed in triplicate, under the same condition as the copepod culturing ( $T = 20 \pm 2^\circ\text{C}$ , 16 h: 8 h light:dark photoperiod). The test solution was renewed three times a week by siphoning off about half of the culture medium and replacing it with a freshly prepared medium at the same concentration. The siphoned-off medium was filtered through the 170  $\mu\text{m}$  and 50  $\mu\text{m}$  mesh size screens, and the recovered nauplii and copepodites were reintroduced into their test flask. The food was provisioned during medium renewal by using the algal mix as part of the dilution medium to prepare test concentrations. Peristaltic pumps were not used during the exposure to avoid the dilution of the test concentrations.

Experiments ended on day 14, when all exposed individuals reached adulthood and sexual maturity. For the remainder of the study, we analyzed RNA extracted from these adults. The content of each Erlenmeyer flask was filtered through the 170  $\mu\text{m}$  screen, and the retained copepods were poured into a 10 mL graduated cylinder filled with 20‰ salinity medium and kept under fluorescent light to concentrate the copepods on the surface and allow the residual food, exuviae, and faecal pellets to settle on the bottom. Adult copepods were then recovered using a 1000  $\mu\text{L}$  micropipette and placed into a second 10 mL graduated cylinder for further separation. Finally, recovered copepods were pipetted into a pre-chilled 1.5 mL RNase/DNase-free Eppendorf tube. The remaining 20‰ salinity medium was removed using a 200  $\mu\text{L}$  micropipette and 200  $\mu\text{L}$  of cold RNAProtect (Qiagen, Hilden, Germany) was added. Samples were stored at  $4^\circ\text{C}$  overnight.

**RNA extraction.** After removing RNAProtect, the samples were ground using sterilized Bel-Art® Disposable Pestles (BAF199230000, Sigma-Aldrich, MO, USA). The Eppendorf tube was partially immersed in liquid nitrogen to cool the tissue when performing the grinding. A figure representing this procedure was created with BioRender.com and is available at Figshare<sup>20</sup>. The following steps were performed at room temperature unless otherwise specified. For the lysis step, samples were resuspended in 400  $\mu\text{L}$  of RLT buffer present in the RNeasy Protect Mini Kit (cat. no. 74124, Qiagen) containing 1% of 14.3 M  $\beta$ -mercaptoethanol (Sigma-Aldrich). The micropestle was also washed with the same buffer to remove all residues. An isovolume of phenol-chloroform-IAA (Sigma-Aldrich) was then added and the samples were vortexed for 30 s and centrifuged for 10 min at max speed (16,100 g). The top aqueous phase was collected without disturbing the other layers and moved to a new RNase/DNase-free Eppendorf tube. Following this step, RNA extraction was continued by using the RNeasy Protect Mini Kit (Qiagen) according to the manufacturer's protocol for the purification of total RNA from animal cells, resumed at step 4. All buffers were prepared accordingly. Briefly: 400  $\mu\text{L}$  of freshly prepared ethanol 70% (Sigma-Aldrich) solution were added to the samples and mixed by pipetting. 700  $\mu\text{L}$  of each sample were loaded onto an RNeasy Mini spin column, followed by 15 s centrifugation at 10,000 g. Flowthrough was discarded and this step was repeated with the leftover volume. Optional DNA removal step was performed using RNase-Free DNase Set (Qiagen) following manufacturer's instructions. The protocol was resumed at step 7. Washing steps were performed, as well as the additional centrifugation step to remove leftover buffer. Elution was performed by adding 40  $\mu\text{L}$  of RNase free  $\text{H}_2\text{O}$  directly on the membrane and centrifuging for 1 min at max speed. The purity of RNA was checked with NanoDrop1000. Quantity was evaluated with Qubit™ fluorometer (ThermoFisher Scientific, MA, USA) using the Qubit™ RNA High Sensitivity (HS) kit following manufacturer's instructions. RIN values were determined using an Agilent Bioanalyzer 2100 (Agilent, CA, USA). All RIN values were higher than 7 (mean  $\pm$  SD for all samples:  $8.52 \pm 0.68$ ).



**Fig. 1** Bioinformatic workflow used in this work for the transcriptome assembly and annotation.

**Nanopore library synthesis.** All samples from the same NEO exposure experiment (control and two exposure concentrations) were pooled for library construction (five pools in total). Input RNA was obtained by pooling together equal amounts of total RNA from treated and control samples for each NEO exposure experiment, reaching the final concentration of  $60 \text{ ng } \mu\text{l}^{-1}$ . When available, samples from parental specimens were included. The cDNA libraries were obtained with the cDNA-PCR Sequencing (SQK-PCS109) kit following the protocol version PCS\_9085\_v109\_revK14Aug2019 last updated on 09/12/2020, available at the Oxford Nanopore Technologies website. One  $\mu\text{l}$  (60 ng) of pooled total RNA for each NEO experiment was used as starting material for reverse transcription and strand-switching. The 20  $\mu\text{l}$  of reverse-transcribed sample were used to make  $4 \times 50 \mu\text{l}$  PCR reactions to select full-length transcripts. The PCR conditions were set as follows: initial denaturation,  $95^\circ\text{C}$  30 s, followed by 12 cycles of denaturation  $95^\circ\text{C}$  15 s, annealing  $62^\circ\text{C}$  15 s, extension  $65^\circ\text{C}$  6 min, and a final extension  $65^\circ\text{C}$  6 min. AMPure XP beads (Beckman Coulter, CA, USA) were used for amplified-DNA purification according to the protocol. The libraries were tested for DNA size and quality, using the Bioanalyzer 2100 (Agilent), and quantity, using the Qubit<sup>TM</sup> DNA High Sensitivity kit (ThermoFisher Scientific). Libraries ranging between 35 fmol and 45 fmol were used for adapter addition. Quality checking, priming and loading of the SpotON flow cell (FLO-MIN106) were performed following manufacturer's instructions. Sequencing was performed using the MinION Mk1C technology (Oxford Nanopore Technologies, Oxford, UK) with an initial bias voltage of  $-180 \text{ mV}$ . Samples were sequenced for 72 h.

**Basecalling.** Raw Nanopore signals were converted into nucleotide sequences using the dorado basecaller (<https://github.com/nanoporetech/dorado>; version 0.5.3 + d9af343) with the Super Accurate model “dna\_r9.4.1\_e8\_sup@v3.6”. Full-length reads were selected using pypochopper (<https://github.com/epi2me-labs/pypochopper>; version 2.7.10).

**Transcriptome assembly.** The *de novo* assembly was performed as detailed in Fig. 1. Reads were assembled using the software RNA-Bloom<sup>21</sup> (version 2.0.1), a method recently extended to perform reference-free reconstruction of transcriptomes starting from long reads. The transcriptome reconstruction was performed independently on each pool of long reads (control and exposure to two NEO concentrations) coming from each NEO experiment. Individual assemblies (five in total) were then joined in a unique transcriptome using the TR2AACDS pipeline from the EvidentialGene software (<http://arthropods.eugenec.org/EvidentialGene/evigene/>) to remove redundant sequences. A final filtering step was performed using DIAMOND alignments (see “Functional annotation” section), removing transcripts whose 3 top hits matched non-eukaryotic organisms in the NR database (294 representative transcripts).

**Transcriptome completeness assessment.** The completeness of the assembly was determined using BUSCO (version 5.8.0)<sup>22</sup> against the arthropoda\_odb10 database. Only one representative transcript for each gene was considered for the analysis. General metrics for the assembly were obtained by using TransRate (version 1.0.3)<sup>23</sup>.

Indel rates in individual Nanopore reads and in the assembled transcriptome were also assessed using the BUSCO gene set as a reference. Reads were aligned to BUSCO genes using the minimap2 software<sup>24</sup> with the “splice” preset, and CIGAR operations corresponding to insertions and deletions were extracted using a custom Python script. In parallel, we evaluated the presence of indels in assembled transcripts corresponding to BUSCO single-copy orthologs using BLASTN.

**Functional annotation.** Assembled transcripts were compared with the NCBI-NR (non-redundant) protein database. Sequence alignments were computed using DIAMOND<sup>25</sup> with a sensitive preset. Results with an e-value greater than  $10^{-6}$  were discarded. Only the best hit for each transcript was considered for further analyses. We also ran InterProScan<sup>26</sup> (version 5.67–99.0) to search known functional domains and predict protein family membership.

|              | Reads generated | Reads passed | Bases passed | Estimated N50 |
|--------------|-----------------|--------------|--------------|---------------|
| Acetamiprid  | 12 M            | 9.58 M       | 9.87 Gb      | 1.25 kb       |
| Clothianidin | 14.55 M         | 11.57 M      | 10.05 Gb     | 1.01 kb       |
| Imidacloprid | 11.41 M         | 8.99 M       | 7.22 Gb      | 0.95 kb       |
| Thiacloprid  | 12.18 M         | 9.33 M       | 10.05 Gb     | 1.24 kb       |
| Thiamethoxam | 10.64 M         | 8.08 M       | 8.71 Gb      | 1.25 kb       |

**Table 1.** Main statistics of Nanopore sequencing runs.

|                                      | This work   | Reference transcriptome |
|--------------------------------------|-------------|-------------------------|
| Number of total sequences            | 261,560     | 119,439                 |
| Number of representative transcripts | 31,291      | 61,149                  |
| Smallest sequence (nt)               | 200         | 501                     |
| Largest sequence (nt)                | 34,884      | 19,191                  |
| Number of assembled bases (nt)       | 531,944,033 | 118,709,440             |
| Mean length (nt)                     | 2,033.74    | 993.89                  |
| n90                                  | 1,126       | 562                     |
| n70                                  | 1,875       | 738                     |
| n50                                  | 2,580       | 1,052                   |
| n30                                  | 3,451       | 1,608                   |
| n10                                  | 5,210       | 3,024                   |
| GC content (%)                       | 39.27       | 37.26                   |

**Table 2.** Summary statistics of the transcriptomic assembly reported in this work obtained with TransRate. Results were compared with the *de novo* transcriptome of *A. tonsa* previously described by Jørgensen *et al.*<sup>2</sup>.

## Data Records

All reads generated in this work were deposited in the NCBI SRA repository<sup>27</sup> under BioProject PRJNA1104241. The *de novo* transcriptome assembly was deposited at GenBank under the accession prefix GKXQ<sup>28</sup>. Tables with the complete data of the functional annotation and amino acid sequences derived from the transcriptome are made accessible on Figshare<sup>20</sup>.

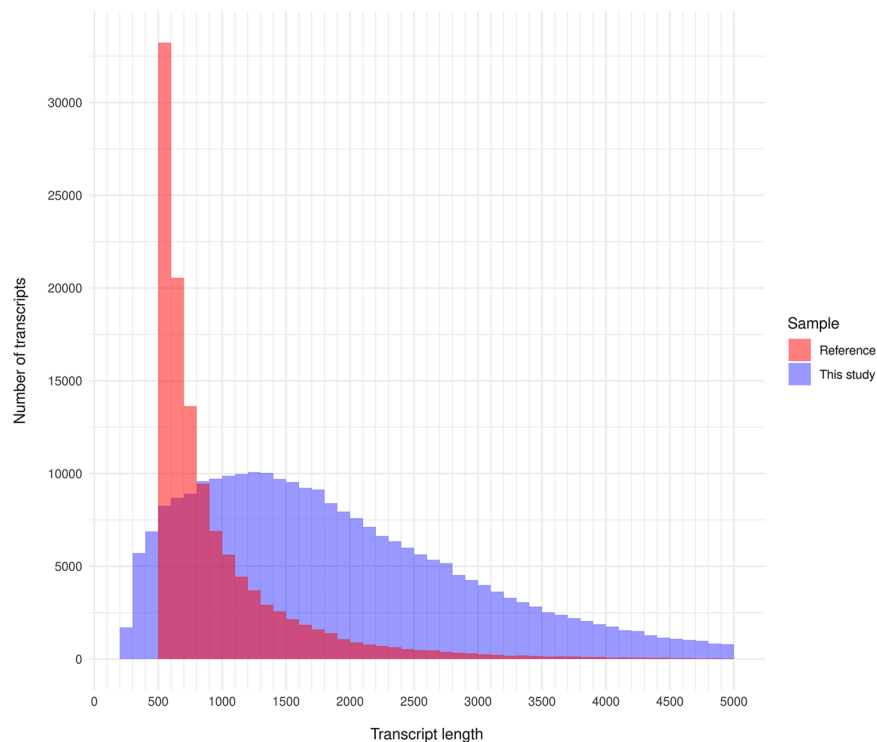
## Technical Validation

**Sequencing statistics.** The main statistics of Nanopore sequencing runs for each NEO are summarized in Table 1. More than 8 M reads passed the quality requirements in all cases, showing an average N50 of  $1.14 \pm 0.15$  kb.

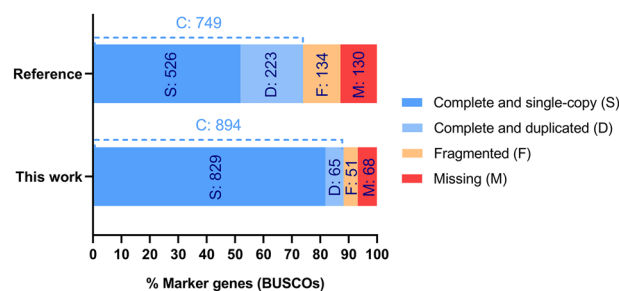
**Transcriptome assembly characteristics.** The main properties of the *de novo* transcriptome assembly (obtained with the TransRate software) are detailed in Table 2. A total of 261,560 transcripts were obtained, where 31,291 sequences were considered as representative transcripts. The mean sequence length was 2,034 bases, over two times the length obtained in the Jørgensen *et al.*<sup>2</sup> transcriptome (994 bases). The N50 was also higher (2,580 compared to 1,052 bases), whereas the GC content was similar between the two transcriptomes (37 vs. 39%). The distribution of all transcripts according to their length (Fig. 2) also showed longer sequences in the transcriptome obtained in this study, with 77.26% of transcripts being longer than 1000 bases.

The completeness of the transcriptome was assessed and compared with a previous transcriptome of *A. tonsa*<sup>2</sup> using BUSCO with representative transcripts of each dataset as input (Fig. 3). The single-copy BUSCO markers are broadly used for completeness measurement, since they are expected to be found once in a genome<sup>21</sup>. Results showed 88.3% of completeness for our transcriptome (894 out of 1,013 arthropod lineage marker genes), of which 81.8% corresponded to single-copy genes and 6.4% to duplicated ones. In contrast, the reference transcriptome<sup>2</sup> exhibited a lower completeness (73.9%) with a low percentage of single-copy markers (51.9%) and a higher percentage of duplicated BUSCOs (22.0%). The high percentage of BUSCO genes that were present in one copy in the representative transcriptome of this work indicates a nearly complete transcriptome with low redundancy. Furthermore, the amino acid sequence identity of BUSCO genes identified in our *de novo* transcriptome assembly against those in the Arthropoda\_odb10 database was  $66.28\% \pm 14.07$ , being slightly higher than the identity between BUSCO genes of the reference transcriptome and the same database ( $61.64\% \pm 15.10$ ).

**Assessment of indels in the dataset.** While the length of Nanopore reads provide significant advantages for transcriptome assembly, they are known to have a higher rate of insertions and deletions compared to alternative technologies<sup>29</sup>. We investigated this tradeoff in our data. In the absence of a reliable genome assembly for comparison, we reasoned that BUSCO genes could serve as a reference for assessing indel frequency. These genes were selected for their presence as near-universal single-copy orthologs, making them highly conserved across species within our taxonomic group. While BUSCO genes were primarily selected for assessing genome/



**Fig. 2** Length distribution of assembled transcripts from this work and from Jørgensen *et al.*<sup>2</sup>.



**Fig. 3** BUSCO completeness analysis. The representative transcripts from this work and from Jørgensen *et al.*<sup>2</sup> (reference) were analyzed using the Arthropoda dataset odb10 (total of marker genes: 1013). The number of complete genes (C) (single copy: S; duplicated: D), fragmented genes (F) and missing genes (M) found for each transcriptome are detailed in the figure.

transcriptome completeness, we leverage their highly conserved nature to provide an estimate of indel frequency, acknowledging that this might only provide an approximate estimate of sequence accuracy.

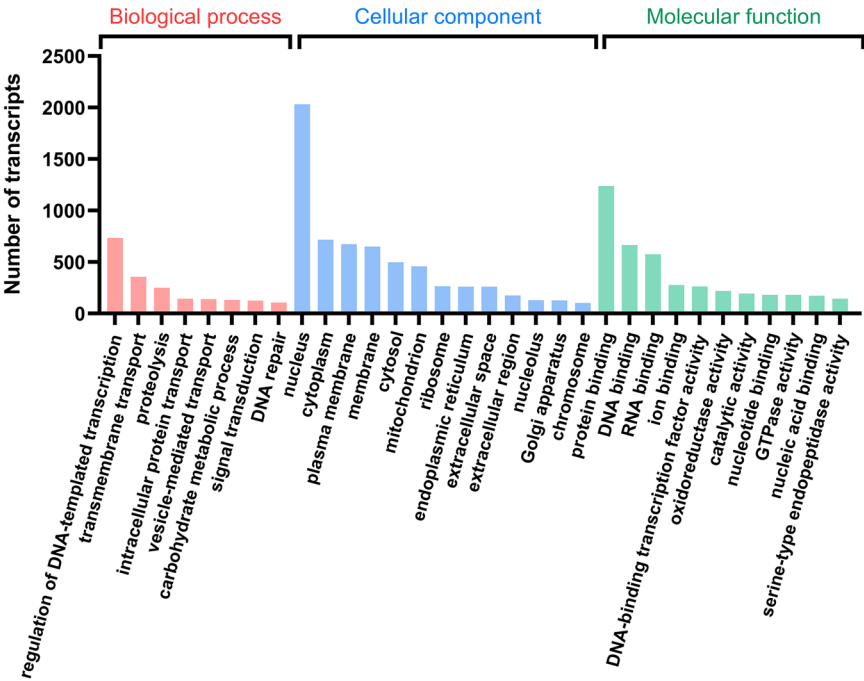
Indel rates in single Nanopore reads were first assessed. Overall, we observed that all sequencing runs were affected by similar indel rates, with a median value of 2.5%. This is consistent with previously reported indel rates for Nanopore sequencing<sup>29</sup>.

To assess the quality of our transcriptome, we compared assembled transcripts corresponding to BUSCO single-copy orthologs using the BLASTN software. Out of the 1,395 matched transcripts, 1,182 (85%) showed no detectable indels. The mean indel rate across all matched transcripts was 0.1%, representing a substantial reduction from the read-level rate. This suggests that our assembly pipeline, incorporating both RNA-Bloom and EvidentialGene, effectively reduced indel rates.

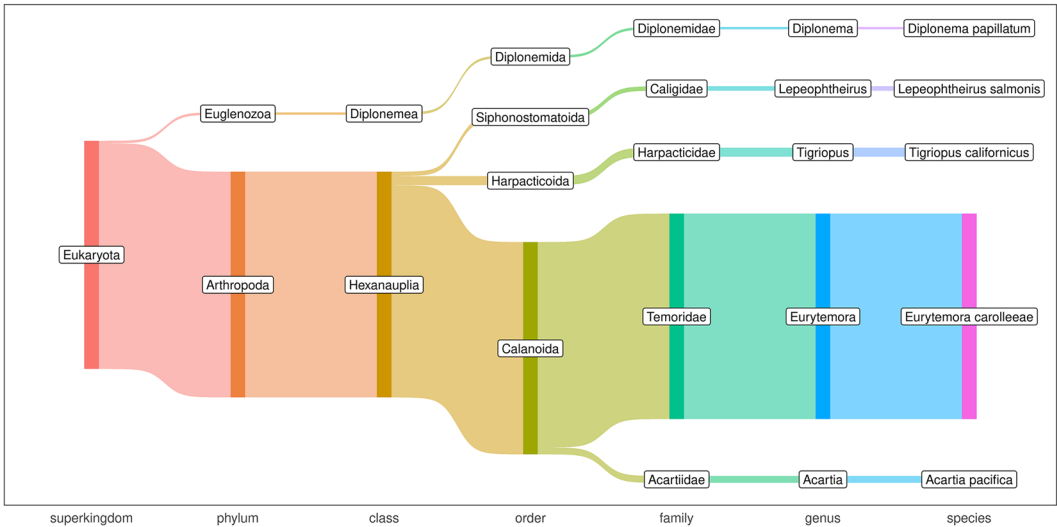
**Functional annotation.** Assembled transcripts were annotated using the different resources included in the InterProScan database (such as Metacyc, Reactome and the Gene Ontology) and searching for sequence homology with proteins stored in the NCBI-NR database. According to annotation statistics (Table 3), 17,005 out of 31,291 representative transcripts were annotated in at least one database, whereas 8,381 were annotated by all databases. The frequency of GO subsets (also known as GO slims) classified in Biological Processes (BP), Cellular Component (CC) or Molecular Function (MF) is shown in Fig. 4. From BP categories, the most frequent GO slim was “regulation of DNA-templated transcription” (734 representative transcripts). Inside the CC ontology,

|                                     | Number of transcripts | Percentage (%) |
|-------------------------------------|-----------------------|----------------|
| Total of representative transcripts | 31,291                | 100            |
| Annotated with GO                   | 12,341                | 39.44          |
| Annotated with MetaCyc              | 10,758                | 34.38          |
| Annotated with Reactome             | 13,468                | 43.04          |
| Annotated with NR-NCBI              | 15,771                | 50.4           |
| Annotated in all databases          | 8,381                 | 26.79          |
| Annotated at least in one database  | 17,005                | 54.35          |

**Table 3.** Percentages of representative transcripts annotated in different databases.



**Fig. 4** Annotation of representative transcripts against the Gene Ontology (GO) database. Bars represent the distribution of representative transcripts in GO slims classified into three functional categories: biological process, cellular component, and molecular function. Only GO slims with more than 100 counts are shown.



**Fig. 5** Taxonomic distribution of the 5 most frequent species among NR hits. The width of the branches represents the frequency of each taxon identified by the DIAMOND search.



more than 2000 transcripts were classified in the “nucleus” category, whereas in the case of the MF group, the top 3 GO slims were protein binding (1238 transcripts), DNA binding (666 transcripts) and RNA binding (575 transcripts). Finally, a homology search of predicted amino acid sequences against the NR database revealed significant matches, primarily with the copepods *Eurytemora carolleeae* (12,447 representative transcripts), *Tigriopus californicus* (550 transcripts) and *Acartia pacifica* (416 transcripts) (Fig. 5).

### Code availability

All software with respective versions and parameters used in this work are listed in the Methods section. Software without associated parameters were used with default settings.

Received: 30 July 2024; Accepted: 13 June 2025;

Published online: 01 July 2025

### References

1. Turner, J. T. The importance of small planktonic copepods and their roles in pelagic marine food webs. *Zool. Stud* **43**, 255–266 (2004).
2. Jørgensen, T. S. *et al.* 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**, 1440–1450 (2019).
3. Smith, J. A. C. *et al.* *Acartia arbruta* (previously *A. tonsa*) in British Columbia: a bioindicator of climate-driven ecosystem variability in the northeast Pacific Ocean. *J Plankton Res* **43**, 546–564 (2021).
4. Wollenberger, L., Breitholtz, M., Kusk, K. O. & Bengtsson, B.-E. Inhibition of larval development of the marine copepod *Acartia tonsa* by four synthetic musk substances. *Sci Total Environ* **305**, 53–64 (2003).
5. Picone, M. *et al.* Impacts of exhaust gas cleaning systems (EGCS) discharge waters on planktonic biological indicators. *Mar Pollut Bull* **190**, 114846 (2023).
6. Koski, M., Stedmon, C. & Trapp, S. Ecological effects of scrubber water discharge on coastal plankton: Potential synergistic effects of contaminants reduce survival and feeding of the copepod *Acartia tonsa*. *Mar Environ Res* **129**, 374–385 (2017).
7. Picone, M. *et al.* Inhibition of Larval Development of Marine Copepods *Acartia tonsa* by Neonicotinoids. *Toxics* **10**, 158 (2022).
8. Picone, M. *et al.* Long-term effects of neonicotinoids on reproduction and offspring development in the copepod *Acartia tonsa*. *Mar Environ Res* **181**, 105761 (2022).
9. Goulson, D. An overview of the environmental risks posed by neonicotinoid insecticides. *J Appl Ecol* **50**, 977–987 (2013).
10. Acebal, M. C., Dalggaard, L. T., Jørgensen, T. S. & Hansen, B. W. Embryogenesis of a calanoid copepod analyzed by transcriptomics. *Comp. Biochem. Physiol. Part D Genomics Proteomics* **45**, 101054 (2023).
11. Nilsson, B., Jepsen, P. M., Bucklin, A. & Hansen, B. W. Environmental stress responses and experimental handling artifacts of a model organism, the copepod *Acartia tonsa* (Dana). *Front Mar Sci* **5**, 156 (2018).
12. Zhou, C. *et al.* *De novo* transcriptome assembly and differential gene expression analysis of the calanoid copepod *Acartia tonsa* exposed to nickel nanoparticles. *Chemosphere* **209**, 163–172 (2018).
13. Raghavan, V., Eichele, G., Larink, O., Karin, E. L. & Söding, J. RNA sequencing indicates widespread conservation of circadian clocks in marine zooplankton. *NAR Genom Bioinform* **5**, lqad007 (2023).
14. Steijger, T. *et al.* Assessment of transcript reconstruction methods for RNA-seq. *Nat Methods* **10**, 1177–1184 (2013).
15. Ebenezer, S. *et al.* Full-length transcriptome from different life stages of cobia (*Rachycentron canadum*, Rachycentridae). *Sci Data* **10**, 97 (2023).
16. Lin, J. *et al.* Nanopore-based full-length transcriptome sequencing of Muscovy duck (*Cairina moschata*) ovary. *Poult Sci* **100**, 101246 (2021).
17. Oikonomopoulos, S. *et al.* Methodologies for transcript profiling using long-read technologies. *Front Genet* **11**, 606 (2020).
18. Hayrabedian, S., Kostova, P., Zlatkov, V. & Todorova, K. Single-cell transcriptomics in the context of long-read nanopore sequencing. *Biotechnol Bioinform Equip* **35**, 1439–1451 (2021).
19. ISO 16778:2015 - Water quality — Calanoid copepod early-life stage test with *Acartia tonsa*. <https://www.iso.org/standard/57698.html> (2015).
20. Mohamed, F. *et al.* *Acartia tonsa* transcriptome annotation. *Figshare* <https://doi.org/10.6084/m9.figshare.26405008> (2025).
21. Nip, K. M. *et al.* Reference-free assembly of long-read transcriptome sequencing data with RNA-Bloom2. *Nat Commun* **14**, 2940 (2023).
22. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).
23. Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J. M. & Kelly, S. TransRate: reference-free quality assessment of *de novo* transcriptome assemblies. *Genome Res* **26**, 1134–1144 (2016).
24. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
25. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* **18**, 366–368 (2021).
26. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
27. NCBI Sequence Read Archive <http://identifiers.org/ncbi/insdc.sra:SRP503798> (2025).
28. Mohamed, F. *et al.* TSA: *Acartia tonsa*, transcriptome shotgun assembly. *GenBank* <http://identifiers.org/ncbi/insdc:GKXQ00000000> (2025).
29. Dohm, J. C., Peters, P., Stralis-Pavese, N. & Himmelbauer, H. Benchmarking of long-read correction methods. *NAR Genom Bioinform* **2**, lqaa037 (2020).

### Acknowledgements

Scientific activity performed in the Research Programme Venezia2021, coordinated by CORILA, with the contribution of the Provveditorato for the Public Works of Veneto, Trentino Alto Adige and Friuli Venezia Giulia. F. M. was supported by the MIUR project – “Dipartimenti di Eccellenza 2018–2022” entitled “The signal in Biology - from cell to ecosystems”. G.S. was funded by the BIRD PRID SEED of the University of Padua with the code SALE\_VIRD21\_01.

### Author contributions

E.C., M.P. and A.V. conceptualized the study. M.P. was responsible for *Acartia tonsa* culturing and NEOs exposure. G.B. carried out all the molecular protocols and Nanopore sequencing. G.S., I.U., A.V. and F.M. processed and investigated the data. G.S., F.M., G.B. and M.P. drafted the article. All authors contributed to the final writing of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to A.V. or G.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025