The transcriptome of the marine calanoid copepod *Temora longicornis* under heat stress and recovery

Ilias Semmouri\textsuperscript{a}, Jana Asselman \textsuperscript{a}, Filip Van Nieuwerburgh\textsuperscript{b}, Dieter Deforce\textsuperscript{b}, Colin R. Janssen\textsuperscript{a}, Karel A.C. De Schamphelaere\textsuperscript{a}

\textsuperscript{a} Ghent University, Laboratory of Environmental Toxicology and Aquatic Ecology, Faculty of Bioscience Engineering, 9000 Ghent, Belgium

\textsuperscript{b} Ghent University, Laboratory of Pharmaceutical Biotechnology, Faculty of Pharmaceutical Sciences, 9000 Ghent, Belgium

contact corresponding author: ilias.semmouri@ugent.be

Present address: Ghent University, Laboratory of Environmental Toxicology and Aquatic Ecology, Campus Coupure, Building F – 2nd Floor, Coupure Links 653 – B-9000 Ghent – Belgium
Abstract

Understanding the impacts of global change in zooplankton communities is crucial, as alterations in the zooplankton communities can affect entire marine ecosystems. Despite the economic and ecological importance of the calanoid copepod *Temora longicornis* in the Belgian part of the North Sea, molecular data is still very limited for this species. Using HiSeq Illumina sequencing, we sequenced the whole transcriptome of *T. longicornis*, after being exposed to realistic temperatures of 14 and 17 °C. After both an acute (1 day) and a more sustained (5 days) thermal exposure to 17 °C, we investigated gene expression differences with animals exposed to 14 °C, which may be critical for the thermal acclimation and resilience of this copepod species. We also studied the possibility of a short term stress recovery of a heat shock. A total of 179,569 transcripts were yielded, of which 44,985 putative ORF transcripts were identified. These transcripts were subsequently annotated into roughly 22,000 genes based on known sequences using Gene Ontology (GO) and KEGG databases. *Temora* only showed a mild response to both the temperature and the duration of the exposure. We found that the expression of 27 transcripts varied significantly with an increase in temperature of 3 °C, of which eight transcripts were differentially expressed after acute exposure only. Gene set enrichment analysis revealed that, overall, *T. longicornis* was more impacted by a sustained thermal exposure, rather than an immediate (acute) exposure, with two times as many enriched GO terms in the sustained treatment. We also identified several general stress responses independent of exposure time, such as modified protein synthesis, energy mobilisation, cuticle and chaperone proteins. Finally, we highlighted candidate genes of a possible recovery from heat exposure, identifying similar terms as those enriched in the heat treatments, i.e. related to for example energy metabolism, cuticle genes and extracellular matrix. The data presented in this study provides the first transcriptome available for *T. longicornis* which can be used for future genomic studies.

Keywords: North Sea, BPNS, global warming, global change, climate change, temperature, gene expression, Crustacea, Arthropoda, zooplankton

1. Introduction

Over the past decades, the world’s oceans and seas have been influenced by human induced impacts such as overfishing, invasive species, coastal eutrophication, chemical pollution, oil/gas extraction and global change (Crain et al., 2009; Halpern et al., 2008; Zacharias and
Globally about 41% of our oceans and seas are already impacted to a high or very high degree, including several parts of the North Sea (Halpern et al., 2008). A major aspect of global change is global warming, i.e. the long-term increase in the Earth’s average surface temperature. In the North Sea region, which is characterized by large annual fluctuations in the temperature of the surface water (from ca. 5°C in winter to about 20 °C in summer and an annual mean SST between 10.5 °C and 11.5°C), the average sea surface temperature (SST) of the water has already risen with 1-2 °C over a time-period of twenty-five years (1985 - 2010) (OSPAR, 2010; van der Kooij et al., 2016). In comparison with the warming rate of the western part of the North-east Atlantic (0.4 - 0.8 °C) over the same time-period, the North Sea is warming faster (OSPAR, 2010). Moreover, the global SST is likely to rise further within the range of 1.1–6.4 °C during the 21st century, depending on the selected emission scenario (Aral and Guan, 2016; IPCC, 2007).

Over 99% of all organisms on Earth are ectothermic, making them susceptible to fluctuations of temperature, and hence, to global warming (Atkinson, 1994). Furthermore, an increased temperature also affects the physical and chemical properties of water, e.g. solubility of carbon dioxide (CO₂) and oxygen (O₂) and the salinity, viscosity, pH and the density of the sea water (IPCC, 2014; McNeil and Matear, 2006; Orr et al., 2005; Pierce et al., 2012; Sharqawy et al., 2010). From a food web point of view, primary production by the smallest plankton is expected to increase in the warmer waters, but this often does not translate into secondary production (the zooplankton and smaller fish) which shows a decreased productivity due to the with global change associated ocean acidification (Nagelkerken and Connell, 2015). As a consequence, temperature often has striking impacts on virtually all life processes from a molecular scale to the entire pelagic ecosystem scale and from plankton to higher trophic levels (Beaugrand and Kirby, 2010; Beaugrand et al., 2009).

Temperature is considered the major factor structuring marine ecosystems and shaping zooplankton dynamics in particular (Richardson 2008). Due to their rapid responses to environmental variation, including shifts in temperature (i.e. changes in biomass, community structure and trophic linkages in marine food webs), planktonic organisms are frequently used as bio-indicators of ecosystem changes (Chiba et al., 2018; Ferdous and Muktadir, 2009). Therefore, understanding the impacts of this changing environmental condition, and hence global change, is essential, as alterations in the zooplankton communities can affect entire marine food webs and ecosystems (Richardson 2008). Zooplanktonic abundance in the Belgian Part of the North Sea (BPNS) is dominated year-round by copepods (on average around 66%
and sometimes even up to 84\% (Krause et al. 1995; Van Ginderdeuren, 2014b). Copepods are the most important trophic link between primary producers and higher-level consumers (such as fish and marine mammals) in the pelagic food web (Möllmann et al., 2005; Richardson, 2008; Van Ginderdeuren, 2014a). By representing 35 to 70\% of the total copepod community, the calanoid copepod *Temora longicornis* O.F. Müller (Maxillopoda, Copepoda, Calanoida) is a major component of the copepod assemblage in the coastal waters of the northern hemisphere (Brylinski, 2009; Daan, 1989; Halsband and Hirche, 2001; Van Ginderdeuren, 2014b; Williams et al., 1994). *T. longicornis* has been shown to be able to remove up to 49\% of the daily primary production of phytoplankton, illustrating its ecological importance (Dam and Peterson, 1993). This copepod species is an important primary food source for the pelagic fish stocks of the BPNS, such as the commercially important herring, *Clupea harengus* L. (Casini et al., 2004; Van Ginderdeuren et al., 2014a).

Due to its ecological and economic importance, the impact of temperature (and climate change) on this species has already been investigated numerous times. Several studies suggest that abundance of *T. longicornis* is strongly correlated to climate variability (Deschutter et al., 2017; Dippner et al., 2000; Heyen et al., 1998; Kane and Prezioso, 2008). It was found that *T. longicornis* was not able to withstand temperatures higher than 22.5 °C, explaining its absence in warmer environments and its restriction to the northern hemisphere, with a southern boundary coinciding with the 20 °C isotherm in summer (Halsband-Lenk et al., 2002). Many studies have already reported effects of increased temperature on different life history traits of this species including body size, development and a number of reproductive parameters (Devreker et al., 2005; Dzierzbicka-Glowacka et al., 2011; Halsband-Lenk et al., 2002; Hay et al., 1988; Holste et al., 2009; Klein Breteler and Gonzalez, 1986). Together, the results of all of the studies mentioned above, indicate that *Temora longicornis* can be regarded as a sentinel species for assessing the impact of global change on the marine environment. Despite the abundance, global distribution and importance of copepods, only limited biological information is available at the molecular level for this taxon (e.g. Lee et al., 2015; Lenz et al., 2014; Ning et al., 2013). Nevertheless, molecular studies on the physiological responses to stress are an interesting method to study and predict responses to climate change, since responses of life-history parameters to environmental stress are being coordinated by changes in gene networks and pathways (Bron et al., 2011). For *T. longicornis* in particular, no genomic/transcriptomic studies have been reported so far. To date, there are only 31 nucleotide sequences for *T. longicornis* deposited in GenBank (September, 2018), corresponding to COI and ribosomal...
genes. Global gene expression could identify cellular and molecular processes, which may provide key data on the effect that global warming has and will have on this abundant species at higher levels of organisation at the individual, population and community level. Therefore, in this study, we used RNA sequencing technology to study the transcriptomics responses and mechanisms of the calanoid copepod *T. longicornis* to temperature fluctuations. Our objective is to document gene expression changes in response to different thermal stress regimes in this ecologically relevant marine species. Using realistic temperatures (i.e. 14 °C, the spring and summer mean SST of the BPNS and 17°C, an estimation of the average SST of the next century), we simulated an increase of temperature of 3 °C for different durations to find genetic mechanisms and signatures for both an immediate or short term (acute) and a more sustained (subchronic) exposure. Another aspect of stress responses in invertebrates that has received relatively little attention is stress recovery. Understanding the mechanisms and processes enabling recovery from stress will help to shed light on the occurring processes after stress exposure. We, therefore, investigated the possibility of a short-term recovery of a heat shock, aiming to identify signatures of stress recovery.

2. **Material & Methods**

2.1. **Sampling and animal acclimatization**

Zooplankton were collected with the research vessel (RV) Simon Stevin on September 24th and 25th 2015, from the coastal waters near Ostend (Belgium), North Sea, NO Atlantic Ocean (September 24: Lat: 51° 19’ 9.62” N, Long: 2° 30’ 52” E, September 25: Lat: 51° 28’ 5.55” N, Long: 2° 48’ 27.66” E). Copepods were collected using a WP2 net (200 µm mesh size) and specimens were stored in 30 L vessels until they could be transported to the laboratory. There, live *T. longicornis* individuals were picked out with a Pasteur pipette and transferred into a rigorously aerated 35 L vessel containing filtered natural seawater (FSW) from the sampling location (Lat: 51° 28’ 5.55” N, Long: 2° 48’ 27.66” E, sea surface), at a temperature of 14 °C, a salinity of 34.05 PSU and a conductivity of 4.3 S/m.

During the acclimation period, copepods were fed three times a week *ad libitum* with a 3:1 mixture of live algae (cf. Gonçalves et al., 2014; Klein Breteler,1980), based on cell densities, containing respectively 75% *Rhodomonas salina* (Wislouch) D.R.A.Hill & R.Wetherbee (Cryptophyta) and 25% of the other algae species: *Isochrysis galbana* Parke (Haptophyta), *Oxhyrris marina* Dujardin (Dinoflagellata), *Prorocentrum cordatum* (Ostenfeld) J.D.Dodge (Dinoflagellata) and *Thalassiossira weissflogii* (Grunow) G.Fryxell & Hasle (Bacillariophyta).
After seven days of acclimation, adult (fully developed) *T. longicornis* were selected from the culture for the exposure experiment.

### 2.2. Experimental design

Adult (fully developed) *T. longicornis* individuals were first exposed to a temperature of either 14 °C or 17 °C. A pool of both male and female copepods was used in the experiment in a sex ratio of 7:10 males:females, which was the same sex ratio that we observed in the zooplankton sample. Temperatures were chosen based on 1) the long-term spring/summer mean SST (April to September) of the southern part of the North Sea i.e. 14 °C (Klimaatinfo, 2016) and 2) a temperature three degrees higher, representing an estimation of the average temperature of the next century if global warming is not stopped or slowed down in the next hundred years (IPCC, 2007). Moreover, as mentioned above, the temperature of the surface water of the BPNS can reach to 20°C in summer, indicating the chosen temperatures are already relevant today. Eight replicates were kept for four days at 14 °C and the other eight were acclimatized to 17 °C over the course of 24 h and kept at this temperature for the remaining three days (Figure 1). Subsequently at day four, four randomly selected replicates were moved for 24 hours to 17 °C from the 14 °C group and vice versa. This resulted in four treatment groups (1) a group exposed to 14°C for five days (2) a group exposed to 17°C for five days (below referred to as the ‘sustained exposure’ group) (3) a group exposed to 14°C for four days and then to an additional acute exposure of 17°C (below referred to as acute exposure group) and (4) a group exposed to 17°C for four days and then to an additional recovery exposure of 14°C (below referred to as recovery exposure group). Each treatment thus consisted of four replicates, represented by 1 L Weck jars, containing 50 adult individuals at the beginning of the experiment (Figure 1). Animals were fed on the first and third day of the experiment with 10,000 cells/mL in an 8-to-2 ratio of *R. salina* and *P. cordatum* respectively. At day five, all animals were sampled from each Weck jar for mRNA extraction.
Figure 1: The set-up of the experiment: two incubators (one at 14 °C and one at 17 °C) with each eight Weck jars with 1 L of filtered sea water with an algae concentration of 10 000 algae/mL and 50 adult *T. longicornis* individuals. On day four of the experiment, four random selected jars (marked in grey) were switched randomly from one incubator to the other. Each Weck jar was aerated rigorously by an aeration pump.

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

Total RNA was extracted from adult copepods using the MasterPureTM RNA Purification Kit (Epicentre®, tebu-bio NV), following the manufacturer's protocol. Remaining DNA was removed by adding RNase-Free DNase I (5 µL diluted in 195 µL DNase Buffer) (MasterPureTM RNA Purification Kit). RNA Quality was assessed with the NanoDropTM spectrophotometer (Version 2000c, Thermo ScientificTM, Thermo Fisher Scientific) (quality standards were based on the following absorbance ratios 260/230 and 260/280). The amount of RNA in the replicates was determined by means of a QubitTM fluorometer (Version 2.0, InvitrogenTM, Thermo Fisher Scientific). Exactly 10 µL of three replicates per treatment was provided to NXTGNT lab (Faculty of Pharmaceutical Sciences, UGent) for library preparation and next generation sequencing. An Illumina mRNA sequencing library was made from one µg total RNA of each sample using the TruSeq Stranded mRNA Sample Prep Kit (Illumina, San Diego, USA) with 13 PCR cycles during the enrichment PCR and according to manufacturer's protocol. The size distribution, purity (absence of free adaptors) and quantity of the resulting libraries were measured using a High Sensitivity DNA chip (Agilent Technologies, Santa Clara, CA, US). The libraries were equimolarly pooled and sequenced in three lanes of an Illumina HiSeq 3000 flow cell, generating paired-end 2x150bp reads. After sequencing, the data was demultiplexed using the sample specific nucleotide barcodes.
2.4. **Sequence assembly and gene annotation**

First, the quality of the raw data was assessed with FastQC Version 0.11.5 (Babraham Bioinformatics, 2010). Next, the sequences were subjected to adapter trimming and quality filtering (low-quality bases from ends of the reads) using TrimGalore Version 0.4.1 (Babraham Bioinformatics, 2012) with a threshold of an ASCII+33 Phred quality score and a length threshold of 50 bp. The remaining clean reads were then de novo assembled to form a draft transcriptome of *T. longicornis*, using the Trinity software Version 2.1.1 (Grabherr et. al., 2011) with default parameters and a minimum contig length of 200 bp. Transcriptome assembly accuracy and completeness was analysed using BUSCO (Benchmarking Universal Single-Copy Orthologs; Simão et al., 2015) against a set of 843 metazoan genes and a set of 2,675 arthropod genes to evaluate the quality of the final assembly. BUSCO is a recognized benchmark approach for single copy orthologs providing an assessment of orthologs conserved among species. All transcripts were BLASTX aligned against the NCBI non-redundant (nr) arthropod protein database using BLAST2GO PRO (Conesa et al., 2005) to evaluate sequence similarity to genes in other species at the e-value cutoff of 1E-03. We also blasted (blastn) these transcripts, at the e-value cutoff of 1E-03, against three other crustacean transcriptomes, more specifically those of the calanoid copepods *Eurytemora affinis* (Almada and Tarrant, 2016), *Calanus finmarchicus* (Lenz et al., 2014) and as outgroup, the cladoceran *Daphnia magna* (Orsini et al., 2016). We also blasted the *Eurytemora* and *Calanus* transcripts to the *D. magna* transcriptome for comparison purposes.

Next, the expressed transcripts were translated into potential proteins according to ORF prediction by TransDecoder v5.0.2 (Grabherr et al., 2011). The potential coding transcripts sequences were annotated against the UniProt database (e-value of 1E-03) and InterProScan using Blast2GO PRO Annotation (Conesa et al., 2005; Jones et al., 2014). Translated contigs (peptide sequences) from the resulting functional transcriptome were also annotated with GhostKOALA (Kanehisa et al., 2015) using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. The draft transcriptome has been deposited at DDBJ/EMBL/GenBank/NCBI SRA under the accession GGQN00000000. The version described in this paper is the first version, GGQN01000000.

2.5. **Differential expressed Gene (DEG) analysis & Enrichment analysis of DEGs**
The reference transcriptome index was constructed using Bowtie v2.3.2 (Langmead and Salzberg, 2012) using the constructed draft transcriptome as described in 2.4, and the clean paired-end reads of each sample were aligned to the reference transcriptome using TopHat v2.1.1 (Pertea et al., 2013). BAM files from TopHat and the GTF file from Cuffmerge v2.2.1 (Trapnell et al., 2010, 2013) were used as inputs for HTSeq v0.6.1 (Anders et al. 2015) to calculate gene counts with the “union” model using the Htseq-count function. These count files and the Bioconductor edgeR (v3.12.1) package (Robinson and Smyth 2008; Robinson et al. 2010) were used to normalize the gene counts, to estimate dispersion and to identify the differentially expressed transcripts. Normalization was performed using the “calcNormFactors” function (Robinson and Oshlack, 2010). Only transcripts with an average number counts per million reads (cpm) per sample higher than 1 were considered for DEG analysis. Three DEG analyses were performed: (1) the ‘sustained exposure’ group vs. the control group that stayed at 14 °C for five days, (2) the acute exposure group vs. the control group, and (3) the recovery exposure group vs. the ‘sustained exposure’ group. The differentially expressed transcripts were determined using a likelihood ratio test with the “glmLRT()” function and p-values were adjusted using the Benjamini-Hochberg false discovery rate (FDR) to account for multiple hypothesis testing (Benjamini and Hochberg, 1995). For the identification of significantly differentially expressed genes, only genes with FDR <0.05 were considered. These DEGS were blasted against the NCBI database (BlastX). DEG results were then filtered based on a set of threshold values: log2FoldChange and an a p-value < 0.05. Gene ontology (GO) enrichment analysis on this filtered data set was conducted using the topGO bioconductor package in R (version 2.30.0) (Alexa and Rahnenfuhrer, 2016; Alexa et al., 2016). Enrichment analysis for GO terms against the GO annotation file from the UniProt/InterProScan database was performed exploiting the Fisher’s exact test with the weight algorithm to obtain P values (Fisher, 1922). A P-value of ≤0.01 was considered as a threshold for significant enrichment of the genes. Raw data files and count files were deposited in NCBI GEO under the accession number GSE115472.

3. Results

3.1 Transcriptome Assembly and Annotation

Twelve libraries of T. longicornis were successfully obtained from the Illumina NextSeq® 500 platform, which represented over 60 Gb of data after quality control. Clean sequencing reads
and alignment statistics are shown in Table S1. Transcriptome assembly of the Illumina reads
by Trinity generated 179,569 transcripts with an average length of 540 bp (Figure 2, Table 1).
Half of these transcripts (N50) were at least 694 bp long, while the longest transcript was 20,976
bp long (Table 1). Approximately 78% (660) of the BUSCO complete orthologs from the
metazoan reference group were identified within the transcriptome, with 207 (24%) fragmented
BUSCOs and 453 (54%) complete BUSCOs. Among the transcripts, TransDecoder identified
44,985 putative ORF transcripts that contain candidate coding regions without a stop codon in
the nucleotide sequences, resulting in a translated peptide assembly (Table 2).

Figure 2. Frequency distribution of the contig size

Table 1. Summary statistics for the de novo assembly of the Temora longicornis draft transcriptome. %GC is the
percent of nucleotide bases in sequences that are either G or C. N50 and N90 bp are the mean number of basepairs
in all transcripts that, ordered by length, make up respectively 50% and 90% of the assembly.

<table>
<thead>
<tr>
<th>Assembly statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of trimmed and high quality raw reads</td>
<td>460,460,205</td>
</tr>
<tr>
<td>Total number of assembled contigs</td>
<td>179,569</td>
</tr>
<tr>
<td>Total number of assembled bases (bp)</td>
<td>96,958,655</td>
</tr>
<tr>
<td>Minimum assembled contig length (bp)</td>
<td>201</td>
</tr>
<tr>
<td>Average assembled contig length (bp)</td>
<td>540</td>
</tr>
<tr>
<td>Median assembled contig length (bp)</td>
<td>334</td>
</tr>
<tr>
<td>Maximum assembled contig length (bp)</td>
<td>20,976</td>
</tr>
<tr>
<td>Total GC count (bp)</td>
<td>40,047,936</td>
</tr>
<tr>
<td>--------------------</td>
<td>------------</td>
</tr>
<tr>
<td>GC count for the whole assembly (%)</td>
<td>41.30</td>
</tr>
<tr>
<td>N50 assembled transcripts length (bp)</td>
<td>694</td>
</tr>
<tr>
<td>N90 assembled transcripts length (bp)</td>
<td>253</td>
</tr>
</tbody>
</table>

Table 2. Summary statistics for the annotation of the de novo assembly of the *Temora longicornis* draft transcriptome. Numbers and percentages between brackets are reported.

**Annotation of transcripts encoding proteins**

<table>
<thead>
<tr>
<th>Description</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcripts with BLAST hits against arthropod nr database</td>
<td>60,809</td>
</tr>
<tr>
<td>Transcripts with BLAST hits against <em>Calanus</em> transcriptome</td>
<td>52,004</td>
</tr>
<tr>
<td>Transcripts with BLAST hits against <em>Eurytemora</em> transcriptome</td>
<td>68,385</td>
</tr>
<tr>
<td>Transcripts with BLAST hits against <em>Daphnia</em> transcriptome</td>
<td>24,295</td>
</tr>
<tr>
<td>Putative number ORF contigs after TransDecoder</td>
<td>44,985</td>
</tr>
<tr>
<td>Transcripts with GO terms (UniProt &amp; InterProScan)</td>
<td>20,525 (45.6%)</td>
</tr>
<tr>
<td>Transcripts with KEGG terms</td>
<td>14,973 (33.3%)</td>
</tr>
<tr>
<td>Number of core eukaryotic genes – BUSCO (Metazoa)</td>
<td>660 (78%)</td>
</tr>
<tr>
<td>Complete genes</td>
<td>453 (54%)</td>
</tr>
<tr>
<td>Complete duplicated</td>
<td>82 (9%)</td>
</tr>
<tr>
<td>Fragmented genes</td>
<td>207 (24%)</td>
</tr>
<tr>
<td>Missing genes</td>
<td>183 (22%)</td>
</tr>
<tr>
<td>Number of core eukaryotic genes – BUSCO (Arthropoda)</td>
<td>1676 (63%)</td>
</tr>
<tr>
<td>Complete genes</td>
<td>1209 (45%)</td>
</tr>
<tr>
<td>Complete duplicated</td>
<td>199 (7%)</td>
</tr>
<tr>
<td>Fragmented genes</td>
<td>467 (18%)</td>
</tr>
<tr>
<td>Missing genes</td>
<td>999 (37%)</td>
</tr>
</tbody>
</table>

BLAST searching of the assembly by sequence homology returned 60,809 peptide transcripts (33.9%) with hits from the NCBI non-redundant (nr) arthropod protein database (Table 2; S1). Blasting the transcripts against the *Eurytemora* transcriptome revealed 68,385 different *Temora* transcripts (38% of the *T. longicornis* transcriptome), corresponding with 91,933 blast hits (covering 66.3% of the *Eurytemora* transcriptome). 29% (52,004 transcripts) the *Temora* transcriptome blast back to 75,465 sequences (37%) sequences of the complete *Calanus* transcriptome (Table 2). In comparison, 40% of the *Eurytemora* transcriptome (56,416 unique transcripts) could be blasted successfully against the *Calanus* transcriptome (75,438 blast hits).
in total, also covering 37% of the entire *Calanus* transcriptome). Finally, 13.5% of the *Temora* transcriptome (24,295 sequences) displayed homology (< e-value of 1E-03) when aligned against the *D. magna* transcriptome (Table 2). In comparison, 14.8% (30,519) of the *Calanus* transcripts and 19% (29,129) of the *Eurytemora* transcripts could be aligned with the *D. magna* transcriptome.

Several complementary approaches were used to annotate the translated assembled sequences. In total, 65,756 (2,588 unique) GO terms and 7,875 unique IPS terms were assigned to 20,525 genes (45.6%), being annotated against UniProt (Table 2). The majority of these GO terms (Figure 3) were assigned to molecular functions (MF, 31,140 hits), followed by biological processes (BP, 18,213 hits) and cellular components (CC, 9,339 hits). A total of 33.3% (14,973) of the predicted proteins by TransDecoder were assigned to orthologous groups in the KEGG database using the GhostKoala mapping tool (Table 2, S2, S3). Most of these assigned orthologs were found to be connected to metabolic pathways (S3). Supplementary figure 4 provides a representation of the global functionality of the genes and summarizes the molecular pathways identified from the functional transcriptome. The five most represented categories of pathways were 1) Signal transduction, 2) Metabolic pathways, 3) Immune system, 4) Cell growth and death categories, and, 5) Transport and catabolism. In total, 23,423 transcripts (~52%) were not annotated by any of these tools.
Figure 3. GO term distribution of BLAST hits from the *T. longicornis* transcriptome. Selected level 2 GO categories are shown within the divisions of Biological Process (BP), Molecular Function (MF), and Cellular Component (CC).

### 3.2 Gene expression under temperature

For differential expression analysis, only putative ORF transcripts with a cpm > 1 were included, resulting in 31,180 transcripts for further DEG analysis. We identified 12 differentially expressed transcripts at a FDR of 0.05 in copepods exposed to a constant temperature of 17°C (sustained heat treatment) when compared with the control group, exposed to a constant temperature of 14°C (Suppl. Table 2). The numbers of up-regulated and down-regulated transcripts in the sustained heat treatment were 5 and 7 respectively (Suppl. Table 2). Comparing the control group that was exposed at 14 °C for five days with the acute treatment
group (switched from 14 °C to 17 °C for 24 h), 15 differentially expressed transcripts were detected with 5 transcripts being down-regulated and 10 transcripts up-regulated (Suppl. Table 3). 8 transcripts were found to be significantly up- (5) or down-regulated (3) when comparing the sustained treatment group that stayed at 17 °C for five days versus the recovery group, that switched back from 17 °C to 14 °C on day 4 (Suppl. Table 4). Among all the differentially expressed genes, 20 transcripts were successfully blasted against the NCBI database (BlastX) (Suppl. Table 2, 3, 4). However, the function of most of these transcripts remains elusive (hypothetical proteins). Some of the genes were found to be affected by several treatments (Figure 4, 5): 5 transcripts were differentially expressed in both the acute and the sustained treatment, as depicted in the reaction norm plot (Figure 5, Suppl. Table 2, 3). Finally, one transcript was down-regulated in the sustained treatment, while being up-regulated in the recovery treatment (Suppl. Table 2, 4).

![Gene expression comparisons of the different treatments](image)

Figure 4: Gene expression comparisons of the different treatments: a Venn diagram of the number of DEGs before (a) and after (b) correcting for multiple hypothesis testing. Genes in overlapping sets show the differential expression in two or three comparison pairs.
Figure 5: Thermal reaction norms of the acute heat shock treatment (Switched S17) and the sustained treatment (Continuous C17) comparing the 5 shared DEGs with the control group (Continuous C14). A steeper slope in a reaction norm is a larger log2-fold change.

Next, we conducted GO enrichment analyses across the different exposures to identify biological functions and mechanisms involved in the thermal stress response. Gene ontology analysis revealed 48 enriched GO terms in the sustained treatment (Table S5). Enrichment indicated that DEGs were mostly associated with cellular components and biological processes, with ribosome related terms being the most prominent. In molecular function, most of the annotated unique sequences were assigned to the structural constituent of ribosome and cytoskeleton, mRNA binding and GTP binding. GO term analysis revealed that DE genes in the acute treatment were strongly related to the GO terms “structural constituent of cuticle” (MF), “macromolecular complex” (CC) and “extracellular exosome” (CC) (Table S6). In total, 26 enriched GO terms were found. Finally, gene ontology analysis of the recovery treatment revealed 23 enriched GO terms (Table S7). The most enriched GO terms were related to the extracellular region for the category CC and the exoskeleton (chitin binding and structural constituent of chitin-based cuticle) for MF.

Discussion

4.1 Transcriptome assembly and comparison with other assembled copepod transcriptomes
The present study presents the first draft transcriptome of *T. longicornis*. The number of assembled transcripts by Trinity and the length of the longest transcript are comparable with other de novo copepod transcriptomes reported (Table 3). The smaller average sequence length and N50 value in our study can be partially attributed to a difference in the minimum sequence length parameter among the studies (200 bp in the present study versus 300 bp in some other studies; cf. Table 3). As the N50 metric is a commonly used metric in almost all other transcriptome studies, we chose to report this value for comparison as well. However, a N50 is not as useful in a transcriptome assembly as in a genome assembly (O’Neal and Emrich, 2003). Many metrics such as N50 and overall assembly size were simply repurposed from whole-genome assembly evaluation (O’Neal and Emrich, 2003). The N50 is used to quantify the number of large scaffolds or chromosomes as expected in a genome assembly (O’Neal and Emrich, 2003). A transcriptome however is a collection of genes, which by nature can differ significantly in length and will never form complete scaffolds or chromosomes. Another way to assess the quality of the *Temora longicornis* transcriptome is to perform a reference-based alignment using BUSCO. The BUSCO results were compared against the arthropod transcriptome reference scores provided in the BUSCO supplementary materials and several recently published crustacean transcriptomes (Figure 6) (Northcutt et al., 2016, Seong et al., 2016, Theissinger et al., 2016; Waiho et al., 2017). Our results are comparable with other recent crustacean transcriptomes such as those of *Homarus americanus* and *Cancer borealis*, but are lower compared to other copepod transcriptomes, of which BUSCO results are known (Figure 6). Overall, the annotation statistics suggest that our draft transcriptome is comparable in terms of quality and depth as other copepod transcriptomes.
Figure 6: Horizontal stacked bar plots showing the proportions of gene sets in BUSCO quality categories for several arthropod transcriptomes. Copepod transcriptomes are denoted by a ‘*’ (Lenz et al., 2014; Tarrant et al., 2014; Bailey et al., 2017; Roncalli et al., 2017a; Roncalli et al., 2017b). Both annotations against metazoan (M) and arthropod (A) dataset for the Temora transcriptome have been included. Several other recently published crustacean transcriptomes are also provided for comparison (Northcutt et al., 2016, Seong et al., 2016, Theissinger et al., 2016; Waiho et al., 2017). Quality categories are as follows: 1) Complete BUSCOs: genes that match a single gene in the BUSCO reference group (green); 2) Fragmented BUSCOs: genes only partially recovered with gene length exceeding alignment length cut-off (orange); 3) Missing BUSCOs: non-recovered genes (red).

After a successful de novo assembly of the RNA-seq reads, 21,562 (~48%) of the transcripts present in the T. longicornis draft transcriptome could be annotated with a KEGG and/or a GO term. While this may seem a low percentage, it is comparable to other non-model crustacean
species (incl. other copepod species), with a similar number of annotated genes, ranging from 20 to 55% (Table 3; Bailey et al., 2017; Kim et al., 2016; Ning et al., 2013; Roncalli et al., 2017a). De novo transcriptome studies in poorly annotated taxonomic groups such as Crustacea continue to be challenging because of these difficulties with annotation. This is due to the lack of information available on the genes of interest in closely related organisms. Only limited information exists for this taxon. To date (September 2018), the Ensembl genome database for metazoan species contains mainly insects and nematodes (Kersey et al., 2018). However, data on only three crustacean species is available: for the water fleas, Daphnia pulex and Daphnia magna, and the salmon louse, Lepeophtheirus salmonis, a specialised parasitic copepod species (Kersey et al., 2018). In addition, while these data are available, these genomes are also far from fully annotated. For the Daphnia spp. genome in particular, 35 to 45% of the coding sequences remain unannotated (Colbourne et al., 2011; Orsini et al., 2016). Furthermore, the number of GenBank records shows a threefold under-representation of crustaceans compared to insects. In total, there are 7,751,892 nucleotide sequences linked to the subphylum Crustacea, while 24,891,737 sequences are retrieved for Insecta (insects) to date (September 2018).

Genomic resources available for crustaceans are hence limited, yet they are crucially needed to improve annotation rates (Stillman et al., 2008). Until now annotations are mainly based on insect genes. However, both groups include species with widely different life strategies: insects are almost exclusively terrestrial, at least in the adult stage, while crustaceans are mostly aquatic species (Alfsnes et al., 2017; Schwentner et al., 2017; Stillman et al., 2008; Tamone and Harrison, 2015). This has profound implications for the environmental drivers and life history strategies of the taxa. Comparing genome size between insects and crustaceans, Alfsnes et al. (2017) found that genome size is clearly phylogeny-dependent for insects, reflecting primarily their life history and mode of development. For crustaceans, however, they found a weaker association between genome size and phylogeny, suggesting life cycle strategies and habitat as more important determinants. Moreover, lineage specific genes are often difficult to annotate because their function is specific to both the species itself and the environment it experiences, which might additionally explain the relatively low annotation rate in the mentioned studies (Asselman et al. 2015).

4.2 Shared responses to acute and sustained thermal stress exposure

A total of 27 transcripts, were differentially expressed across heat treatments, demonstrating a pervasive genomically mild response to temperature stress, representing ~0.09% of the genes
considered for DEG analysis. The lack of up-regulation of important stress response genes could indicate that neither one of these treatments were felt as a severe stress for the copepods. While this is indeed a low number, low percentages have also been found in other studies that used non-targeted investigations of the transcriptome in response to thermal stress: in the copepods *Calanus finmarchicus* and *Tigriopus californicus* (~0.73% and ~0.88% respectively; Schoville et al., 2012; Smolina et al., 2015), the coral *Acropora hyacinthus* (~0.40%–0.70%; Barshis et al., 2013), the Antarctic fish *Trematomus bernacchii* (~0.08%–1.00%; Huth & Place, 2016) and the sea grass *Zostera marina* L. (~0.80%; Franssen et al., 2011). The response percentages of the mentioned copepod species (*C. finmarchicus* and *T. californicus*) are a bit higher than those of *T. longicornis*. This is probably due to the fact that the former species were exposed to increased temperatures just outside their natural temperature range (15°C and 35 °C respectively), whereas *Temora* was exposed to 17 °C, a temperature still within its temperature range of 5-20 °C (Halsband-Lenk et al., 2002; Schoville et al., 2012; Smolina et al., 2015). We aimed at identifying possible target genes for acclimation to different thermal regimes (a short and more sustained exposure). General stress responses include DNA damage sensing and repair, transcription regulation, lipid metabolism and energy metabolism and damaged protein degradation (Kültz, 2005). DEG analysis identified several types of genes related to these general stress responses (Table S2, S3), but 17 out of the 27 DEGs (63 %) remain to date unannotated.

We observed seven GO terms, “structural constituent of cuticle”, “lipid metabolic process”, “response to heat”, “axonemal microtubule”, “cytosolic small ribosomal subunit”, “polysomal ribosome” and “extracellular matrix” to be enriched in genes differentially regulated in both the acute and sustained exposure treatments, indicating common transcriptional responses, caused by temperature fluctuations at different time points. Differential regulation of the ribosome and mRNA synthesis suggest an impact of temperature on protein synthesis (Table S5, S6). The differential regulation of protein synthesis in other copepod species, as well as in other marine species such as the Pacific oyster, *Crassostrea gigas*, has already been observed multiple times after exposure to thermal stress (Meistertzheim et. al., 2007; Schoville et al., 2012; Smolina et al., 2015). The “response to heat” term was also found to be enriched in other arthropod studies (e.g. Cabrera-Brandt et al., 2014; Teets et al., 2012) after exposure to pesticides or dehydration respectively, rather than to heat stress.

Enrichment analysis also revealed an overrepresentation of genes involved with the structural constituent of the cuticle in the immediate and sustained heat stress group (Table S5, S6) and
DEG analysis revealed one cuticular DEG in the latter treatment (Table S2). Cuticle proteins are structural proteins that comprise the exoskeleton of arthropods together with chitin (Liu et al. 2010). They have been found DE after thermal exposure in several previous studies as well, e.g. in Drosophila species (Zhao et al., 2015), the nematode C. elegans (Brunquell et al., 2016) and in the copepod Tigriopus californicus (Lima and Willett, 2017; Schoville et al., 2012), although the direction of the expression differs between and within these studies. Comparison with the control group (Table S2), shows a general down-regulation of the cuticle gene in our study. During development, the transcription of cuticle collagens is tightly regulated with molting and growth in arthropods (Johnstone and Barry, 1996). During molting, a new and larger exoskeleton is produced beneath the old cuticula, which is ultimately replaced so that the animal can expand in size (Eichner et al., 2015; Hartnoll, 2001). Yet, molting in the adult phase is suppressed in copepods, a synapomorphy for Copepoda (Khodami et al., 2017). It is known though that external factors such as food and temperature affect molting in some species (Hartnoll, 2001). We do not know the exact function of these cuticular genes/proteins in response to thermal stress, but it is possible that cuticle associated proteins are part of the thermal response in arthropods (Lima and Willett, 2017). Some studies suggest the implication of the cuticle in the regulation of the internal body temperature of arthropods (Amore et al., 2017; Carrascal et al., 2017).

Genes involved in mobilization of energy reserves (e.g. enriched GO term “lipid metabolic process”) also appear to be essential components of the general stress response in T. longicornis. It is known as a species with high metabolic rates but low energy reserves (Helland et al. 2003; Kreibich et. al. 2008). DEG analysis seems to support this, as it revealed a significant up-regulation for transcripts encoding for a subunit of NADH dehydrogenase (also called NADH-ubiquinone oxidoreductase), after exposure to both treatments, i.e. the acute and the more sustained period of thermal stress (Table S2, S3). A significant increase of an enzyme of the respiratory chains possibly leads to an increase in mitochondrial activity to meet an increased energy demand. This hypothesis seems to be supported by the slight up-regulation comparing the immediate and the sustained heat treatment (Figure 5). An up-regulation of this enzyme was also found in the related great spider crab Hyas araneus in response to ocean acidification and warming (Harms et al., 2014). In larvae of the mosquito species, Aedes aegypti, exposed to xenobiotics, NADH dehydrogenase was also found to be up-regulated for example, confirming this DEG is being part of a global stress response (David et al., 2010).
One annotated transcript related to damage control, apolipoprotein D-like isoform, was also DE in both heat treatments (exposure to acute and a more sustained period of thermal stress) (Table S2, S3). Apolipoprotein D (ApoD) and isoforms are ancient members of the lipocalin family with a high degree of sequence conservation from plants to insects to mammals (Ganfornina et al., 2008; Muffat and Walker, 2010). ApoD has emerged as an evolutionarily conserved anti-stress protein, by maintaining low levels of lipid peroxidation, that is induced by both internal and external stressors and inflammation (Ganfornina et al., 2008; Muffat and Walker, 2010). ApoD was found to be up-regulated after induction to several types of stressors in studies on human cells, mammals, fruit flies and plants (Charron et al., 2008; Dassati et al., 2014; Muffat and Walker, 2010). Interestingly, in our study we showed that the gene coding for the ApoD isoform was down-regulated after exposure to thermal stress, especially after exposure to acute stress, with a small increase after exposure to more sustained stress (Table S2, S3). In the study of Xu et al. (2016) two down-regulated apolipoprotein D genes were identified upon cold stress in the white-backed planthopper, Sogatella furcifera.

### 4.3 Signatures of acute thermal stress

“Macromolecular complex”, “extracellular exosome”, “short-chain fatty acid catabolic process” and “glyceraldehyde-3-phosphate dehydrogenase (NAD+) (phosphorylating) activity” are the most significant GO terms enriched in genes differentially regulated by acute heat stress (Table S5, S6). These GO terms were not significantly enriched in genes up-regulated after a more sustained exposure to heat stress, suggesting a temporal specificity of these responses to heat stress. The Go term “chitin binding” though, can be linked to cuticle related GO terms, a response shared between the exposure scenarios (discussed in 4.2). Eight transcripts were DE specifically after exposure to short term heat stress as well, including the down-regulated gene, ganglioside GM2 activator-like. GM2 activators act within lysosomes, where they present GM2 gangliosides to a coenzyme to be broken down (d’Azzo and Sano, 2006). Several crucial cellular responses may be influenced or controlled by gangliosides and ultimately result in either cell growth and division, differentiation, or cell death (d’Azzo and Sano, 2006; Tessitore et al., 2004). Under stress conditions that dramatically increase their intracellular concentration, gangliosides can initiate the induction of an apoptotic program (d’Azzo and Sano, 2006; Tessitore et al., 2004). Thus, a down-regulation of enzymes that break down gangliosides, may possibly lead to an increased abundance of gangliosides, suggesting perceived stress at the cellular level.
Several enriched GO terms (e.g. “short-chain fatty acid catabolic process”, “carbohydrate phosphorylation” and “glyceraldehyde-3-phosphate dehydrogenase”) are associated with energy metabolism. We also found an increase in the expression of the SLC37A2-like gene (also known as glucose-6-phosphate transporter/exchanger, G6PT) (Table S3). Even though there is still little information on its exact protein function, it is described to transport glucose-6-phosphate, a key component of glycolysis, the first fundamental metabolic process of aerobic respiration (Chou and Mansfield, 2014; Olcese and Bartolini, 2013). A possible explanation of the up-regulation of this gene is the additional requirement for energy of the organism as a general stress response. Respiration rates of several marine calanoid copepod species, including T. longicornis and several Centropages species, were indeed found to increase with temperature following an exponential function (Castellani and Altunbas, 2013; Cruz et al., 2013; Gaudy, 1973, 1975). In another study, Vidal (1980) found a linear increase of the metabolic rate with a logarithmic increase in temperature in the marine calanoid copepod Calanus pacificus. These findings, together with our results, indicate indeed an increased energy demand after exposure to increased water temperatures.

4.4 Signatures of sustained thermal stress

Overall, the transcriptome of T. longicornis demonstrated a larger response to a sustained heat exposure, rather than an immediate (acute) exposure, with two times as many enriched GO terms in the sustained treatment. Gene ontology analysis indicated that the differentially expressed genes were enriched with a broad spectrum of biological processes and pathways (48 GO terms in total) among which protein synthesis appears to be the most affected metabolic pathway. After exposing Calanus finmarchicus copepods to realistic, sustained thermal stress (+ 5°C), Smolina et al., (2015) also found the enrichment of the GO terms “structural constituent of the ribosome” and “cytosolic small ribosomal subunit” (cf. our results, Table S5), suggesting an effort to increase translation capacity or to protect the ribosomal function through the replacement/addition of ribosomal proteins (Kültz, 2005). Several GO terms, such as “protein binding”, “mRNA binding”, “nucleosome” and “structural constituent of cytoskeleton”, were found to be significantly enriched in genes differentially regulated after a more sustained exposure only (Table S5). Nucleosome remodelling has already been shown to be an important part of the stress response, e.g. by allowing access to transcription sites of stress responsive genes (de Nadal et al., 2011; Guertin et al., 2010; Jovic et al., 2017). Depletion of nucleosome remodelling leads to a higher thermal sensitivity in C. elegans for example (Satyal et al., 1998). Hence, packaging of DNA into nucleosomes could be an additional protective mechanism
during the stress response (Jovic et al., 2017). In addition to nucleosome remodelling, genes encoding components of the cytoskeleton were also found to be enriched after a sustained exposure to heat stress (Table S5). Cytoskeletal organization is known to go through a profound transformation under thermal stress (Richter et al., 2010). The GO term “structural constituent of cytoskeleton” was also found as enriched in other heat stress exposure studies in for example trypanosomatids and the Hong Kong Oyster, *Crassostrea hongkongensis* (Kelly et al., 2012; Li et al., 2017). Again, several enriched GO terms, related to metabolism and energy (e.g. “ATP synthesis coupled proton transport”, Table S5), suggest that *T. longicornis* individuals experienced a higher metabolic rate when submitted to a sustained heat temperature treatment.

We also identified one GO term (“heat shock protein binding”; interacting selectively and non-covalently with a heat shock protein) associated with heat-shock proteins (Hsps) in the sustained exposure treatment (Table S5). The same GO term was also found to be significantly enriched with a lower pH in *Calanus glacialis* (Bailey et al., 2017). A general cellular reaction to a stressor is the expression of molecular chaperones (De Maio, 1999; Moseley, 1997). As molecular chaperones, heat-shock proteins are involved in the folding and unfolding of other proteins with proteins that are in their non-native conformation (stress denatured) in such a way that they prevent these proteins from interacting inappropriately with each other (De Maio, 1999; Moseley, 1997). Heat shock proteins are typically increasing in abundance when cells are exposed to elevated temperatures or other stressors (De Maio, 1999; Moseley, 1997). Through BLAST searching the nr protein database, a total of 171 Hsps and 31 Hsfs (heat-shock factor) were identified in the functional transcriptome, none of which were significantly differentially expressed (p-value <0.05; Table S5). In addition to their chaperone activity, hsp5 also target damaged proteins to the proteasome to prevent accumulation of dysfunctional proteins and to recycle amino acids (Goldberg, 2003). We detected the enrichment of a GO term related to proteolysis (Table S5) in the DE genes. The lack of an abundant number of differentially expressed chaperone genes after both an acute and a sustained treatment in our study could indicate that (1) only a limited number of proteins is affected by the increase of temperature to 17°C, making an up-regulated expression of Hsps unnecessary, or (2) that up-regulation of the Hsps happened already a few hours (less than 24 hours) after the treatments and was already back to normal regulation when RNA was extracted.

### 4.5 Signatures of recovery after thermal stress
To assess whether *T. longicornis* individuals could recover from exposure to water temperatures of 17°C, we compared gene expression between two groups exposed to 17°C, with one of them having had a day of recovery exposure of 14°C on day 4 (the recovery treatment). We found some enriched GO terms related to energy and protein metabolism (e.g. “alpha, alpha-trehalase activity”, Table S7). Interestingly, the GO terms “chitin binding”, “structural constituent of the chitin-based cuticle”, “structural constituent of the extracellular matrix” and “chaperone-mediated protein folding” were also enriched, confirming their role in the stress response of *T. longicornis*. DE analysis revealed only 8 DEGs. One gene, trypsin-1 like (Table S4), was shown to be down-regulated after recovery. Trypsins, having a proteolytic function, have been established as key enzymes for growth through its role in the protein digestion processes (Rungruangsk Torrissen and Male 2000). Moreover, trypsin is recognized as the most abundant protease present in crustaceans (Jones et al. 1997). Trypsins have already been reported being differentially expressed, in both directions, after exposure to stressors such as cyanobacteria, cadmium and salinity stress in the crustacean *Daphnia magna* (Asselman et al., 2012; Heckman et al., 2008; Poynton et al., 2007). The removal of a stressor might be an explanation for the down-regulation in our experiment. Simultaneously, DEG analysis revealed the up-regulation of a part of the protease inhibitor inter-alpha-trypsin inhibitors (ITI) heavy chain H3-like, suggesting indeed an inhibition of trypsin activity. On the other hand, there is strong evidence that all members of the ITI family play an important role in extracellular matrix stability by covalent linkage to hyaluronan, linking the ITI H3-like gene with the GO term “structural constituent of the extracellular matrix” (Hamm et al., 2008). Recovery from stress is dependent on several factors, such as exposure duration, timing of stress application in terms of development, population densities and other ecological conditions, as has been shown in several studies on *Daphnia* (Barata et al., 2002; Hanazato and Dodson, 1995; Reynaldi and Liess, 2005). While this study presents a first effort to reveal the possibility of stress recovery in *Temora*, more detailed research is necessary to understand the recovery potential of this species from stress.

### 4.6 Linking molecular responses to potential higher-level effects

Our results from the DEG and GO enrichment analyses suggest *Temora longicornis* may display a general cellular stress response when exposed to a water temperature increase of 3°C. More research is needed to confirm the downstream biological and developmental effects of the differential expression of these genes. The effect of temperature-induced gene expression
changes on the fitness of Temora in the BPNS remains thus to be addressed. Our findings suggest a weak response to exposure to a temperature of 17 °C, which might be interpreted as an indication that T. longicornis is naturally tolerant to this temperature, which is consistent with the occurrence of this species above the 20 °C isotherm in summer (Halsband-Lenk et. al., 2002). Moreover, using ad libitum feeding conditions, Holste et al. (2009) measured egg production of T. longicornis in the Baltic Sea at 12 different temperatures between 2.5 and 24°C and observed the highest mean egg production at 16.9°C (12 eggs female⁻¹ day⁻¹), suggesting a tolerance to an increased temperature of the sea water. The findings of Holste et al. (2009) contrast with Halsband-Lenk et. al. (2002), who did not find large differences in reproduction rates over a temperature range of 2-20 °C. Deschutter et al. (2017) applied multimodel inference on generalized additive models to quantify the relative contribution of temperature to the dynamics of calanoid copepod species in the Belgian part of the North Sea. For Temora, they found that temperature is consistently important for the population density of this species: an optimal temperature between 14°C and 16.5°C, together with negative effects below temperatures of ± 11 °C was found for this species. Together, these studies indicate a tolerance to water temperatures of 17°C.

Although a GO enrichment analysis offers initial insights into processes affected by an elevated temperature, a strong bias still exists towards conserved and well-characterized molecular functions and cellular components in model organisms (Harms et al., 2014). Species-specific transcripts are not assessed when they are not annotated and may thus form a bottleneck for the identification of key stressor pathways. This bias particularly applies to non-model organisms, such as Temora, with a lack of GO annotation for about half of the transcripts. Additionally, many genes are grouped into more than one GO term and are thus more difficult to interpret. Considering these limitations, a GO enrichment analysis can only provide a general overview of possibly affected processes, making more detailed analyses desirable. Also, to predict whether T. longicornis will be able to cope with the temperature rise - due to global warming - more elaborate experiments with more extreme temperature stress treatments should be carried out (instead of mild temperature exposure treatments only). In addition, it could also be interesting to assess gender-specific and life-stage specific differences in the molecular stress response of Temora. Boulangé-Lecomte et. al. (2014) found for example evidence for sexual dimorphism in the heat shock protein expression (Grp78 and Hsp90A) in the closely related estuarine copepod Eurytemora affinis, as females showed significantly higher expression for the tested Hsp genes than male individuals, suggesting females might be more stress resistant.
Finally, transcriptomic responses may not necessarily predict changes at the protein level. Limitations of monitoring exclusively gene expression, such as posttranslational modifications and the destabilising mRNA under stress, makes mRNA an inaccurate proxy for the abundance of active products of translation (Androulakis et al., 2007). We acknowledge that mRNA transcript abundance is indeed not equivalent to protein abundance and that ultimately it is the stressed proteome that protects and repairs the cell during stressor exposure. Yet, other studies have shown that there is often a high correlation between gene expression and protein abundance and that gene expression changes in experimental treatments are often relevant predictors of organismal physiology (Feder and Walser, 2005; Maier et al., 2009). Hence, future exposure studies should further try to link gene expression data with phenotypic data.

Conclusions

The first transcriptome analysis on the calanoid copepod Temora longicornis was carried out here. However, the annotation of the gene transcripts proved to be more difficult due to the lack of available information or closely related genomes. The implementation of whole transcriptome technology to study the response to a thermal stressor in T. longicornis has resulted in identifying several general stress responses independent of exposure time, such as modified protein synthesis, energy mobilisation, cuticle and chaperone proteins. Overall, T. longicornis was more impacted by a sustained exposure, rather than an immediate (acute) exposure, to thermal stress with two times as many enriched GO terms in the sustained treatment. Furthermore, we also highlighted several transcripts and enriched GO terms with a temporal specificity, illustrating the importance of exposure time in gene transcription. Finally, we highlighted candidate genes of a possible recovery from heat exposure, identifying similar terms as those enriched in the heat treatments, i.e. related to for example energy metabolism, cuticle genes and extracellular matrix. In conclusion, Temora exhibits a general, mild cellular stress response when exposed to water temperature increase of 3°C. However, more research is needed to confirm the downstream biological, developmental and fitness effects of these differentially expressed transcripts.

Acknowledgements

The authors are grateful to the Flanders Marine Institute (VLIZ) for the assistance, especially André Cattrijse and the crew of the research vessel Simon Stevin. The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC
(Flemish Supercomputer Center), funded by Ghent University, FWO and the Flemish Government – department EWI. The authors thank Josefien Van Landuyt, Pieter Masschelein and Olivier Berteloot for the technical assistance. Funding was received from UGent research fund (BOF15/24J/106). Jana Asselman is an FWO (Fonds voor Wetenschappelijk Onderzoek – Vlaanderen) postdoctoral fellow.

References


Table 3. Comparison of de novo assembled transcriptomes for several copepod species. Summary statistics are reported, unless not provided by the study (‘nr’: not reported). Where possible, summary statistics were calculated (marked with an ‘*’). Respective data sources: 1) Current study; annotation results of the functional draft transcriptome are included in the table; 2) Almada and Tarrant, 2016; 3) Lenz et al., 2014; 4) Tarrant et al., 2014; 5) Bailey et al., 2017; 6) Ning et al., 2013; 7) Roncalli et al., 2017a; 8) Roncalli et al., 2017b; 9) Schoville et al., 2012, San Diego population 10) Kim et al., 2015; 11) Kim et al., 2016; 12) Lee et al., 2015; 13) Gallardo-Escárate et al.. 2014. Percentages are given between brackets.

<table>
<thead>
<tr>
<th></th>
<th>Calanoida</th>
<th>Harpacticoida</th>
<th>Cyclopoida</th>
<th>Siphonostomatoida</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sequencing platform</strong></td>
<td>Illumina HiSeq</td>
<td>Illumina HiSeq</td>
<td>Trinity</td>
<td>MiSeq Illumina</td>
</tr>
<tr>
<td>Raw Reads (#)</td>
<td>460,460,205</td>
<td>216,737,662</td>
<td>93,000,000</td>
<td>154,840,000</td>
</tr>
<tr>
<td><strong>Assembled Transcripts (#)</strong></td>
<td>179,569</td>
<td>138,581</td>
<td>206,041</td>
<td>125,631</td>
</tr>
<tr>
<td><strong>Transcripts with ORF (#)</strong></td>
<td>44,985</td>
<td>82,891</td>
<td>96,090</td>
<td>83,444</td>
</tr>
<tr>
<td><strong>Minimum Length (bp)</strong></td>
<td>201</td>
<td>201</td>
<td>301</td>
<td>300</td>
</tr>
<tr>
<td><strong>Average Length (bp)</strong></td>
<td>540</td>
<td>857</td>
<td>997</td>
<td>819</td>
</tr>
<tr>
<td><strong>Maximum Length (bp)</strong></td>
<td>20,976</td>
<td>23,627</td>
<td>23,068</td>
<td>23,841</td>
</tr>
<tr>
<td><strong>N50</strong></td>
<td>694</td>
<td>2,087</td>
<td>1,418</td>
<td>4,178</td>
</tr>
<tr>
<td><strong>Transcripts with BLAST hits (#)</strong></td>
<td>60,809</td>
<td>nr</td>
<td>81,554</td>
<td>12,474</td>
</tr>
<tr>
<td><strong>Transcripts with GO terms (#)</strong></td>
<td>20,525</td>
<td>nr</td>
<td>78,467 (81%)</td>
<td>23,841</td>
</tr>
<tr>
<td><strong>Transcripts with KEGG terms (#)</strong></td>
<td>14,973</td>
<td>nr</td>
<td>67,673 (70%)</td>
<td>15,314 (19%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Calanoida</th>
<th>Harpacticoida</th>
<th>Cyclopoida</th>
<th>Siphonostomatoida</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Temora longicornis</strong></td>
<td>Illumina HiSeq</td>
<td>Illumina HiSeq</td>
<td>Trinity</td>
<td>MiSeq Illumina</td>
</tr>
<tr>
<td><strong>Eurytemora affinis</strong></td>
<td>Illumina HiSeq</td>
<td>Illumina HiSeq</td>
<td>Trinity</td>
<td>Trinity</td>
</tr>
<tr>
<td><strong>Calanus finmarchicus</strong></td>
<td>Illumina HiSeq</td>
<td>Illumina HiSeq</td>
<td>Trinity</td>
<td>Trinity</td>
</tr>
<tr>
<td><strong>Calanus glacialis</strong></td>
<td>Illumina HiSeq</td>
<td>Illumina HiSeq</td>
<td>Trinity &amp; CAP3</td>
<td>Trinity</td>
</tr>
<tr>
<td><strong>Calanus sinicus</strong></td>
<td>454 GS FLX</td>
<td>Illumina NextSeq</td>
<td>Trinity</td>
<td>Trinity</td>
</tr>
<tr>
<td><strong>Neocalanus flemingeri</strong></td>
<td>Illumina HiSeq</td>
<td>Illumina NextSeq</td>
<td>Trinity</td>
<td>Trinity</td>
</tr>
<tr>
<td><strong>Labidocera madurae</strong></td>
<td>Illumina HiSeq</td>
<td>Illumina NextSeq</td>
<td>Trinity</td>
<td>Trinity</td>
</tr>
<tr>
<td><strong>Tigriopus californicus</strong></td>
<td>Illumina HiSeq</td>
<td>Illumina NextSeq</td>
<td>Trinity</td>
<td>Trinity</td>
</tr>
<tr>
<td><strong>Tigriopus japonicus</strong></td>
<td>Illumina HiSeq</td>
<td>Illumina NextSeq</td>
<td>Trinity</td>
<td>Trinity</td>
</tr>
<tr>
<td><strong>Tigriopus king-sejongensis</strong></td>
<td>nr</td>
<td>nr</td>
<td>Trinity</td>
<td>Trinity</td>
</tr>
<tr>
<td><strong>Paracyclopina nana</strong></td>
<td>245,135</td>
<td>211,002</td>
<td>38,250 (49%)</td>
<td>67,179 (53%)</td>
</tr>
<tr>
<td><strong>Caligus rogercresseyi</strong></td>
<td>245,135</td>
<td>211,002</td>
<td>38,250 (49%)</td>
<td>67,179 (53%)</td>
</tr>
</tbody>
</table>

| **Calanoida**            | 44,985 (26%)               | 82,891 (59%)               | 96,090 (47%)                   | 83,444                          |
| **Calanina finmarchicus**| 124,618 (51%)              | 59,353                     | 108,092 (44%)                  | 83,444                          |
| **Calanus glacialis**    | 59,353                     | 108,092 (44%)              | 72,391 (34%)                   | 83,444                          |
| **Calanus sinicus**      | 124,618 (51%)              | 59,353                     | 108,092 (44%)                  | 83,444                          |
| **Neocalanus flemingeri**| 124,618 (51%)              | 59,353                     | 108,092 (44%)                  | 83,444                          |
| **Labidocera madurae**   | 59,353                     | 108,092 (44%)              | 72,391 (34%)                   | 83,444                          |
| **Tigriopus californicus**| 59,353                    | 108,092 (44%)              | 72,391 (34%)                   | 83,444                          |
| **Tigriopus japonicus**  | 59,353                     | 108,092 (44%)              | 72,391 (34%)                   | 83,444                          |
| **Tigriopus king-sejongensis** | nr                     | nr                        | nr                             | 83,444                          |
| **Paracyclopina nana**   | 245,135                    | 211,002                    | 38,250 (49%)                   | 83,444                          |
| **Caligus rogercresseyi**| 245,135                    | 211,002                    | 38,250 (49%)                   | 83,444                          |

| **Calanoida**            | 201                        | 201                        | 301                            | 300                             |
| **Calanina finmarchicus**| 201                        | 201                        | 301                            | 300                             |
| **Calanus glacialis**    | 301                        | 301                        | 301                            | 300                             |
| **Calanus sinicus**      | 301                        | 301                        | 301                            | 300                             |
| **Neocalanus flemingeri**| 301                        | 301                        | 301                            | 300                             |
| **Labidocera madurae**   | 301                        | 301                        | 301                            | 300                             |
| **Tigriopus californicus**| 301                        | 301                        | 301                            | 300                             |
| **Tigriopus japonicus**  | 301                        | 301                        | 301                            | 300                             |
| **Tigriopus king-sejongensis** | 301                    | 301                        | 301                            | 300                             |
| **Paracyclopina nana**   | 301                        | 301                        | 301                            | 300                             |
| **Caligus rogercresseyi**| 301                        | 301                        | 301                            | 300                             |

| **Calanoida**            | 201                        | 201                        | 301                            | 300                             |
| **Calanina finmarchicus**| 201                        | 201                        | 301                            | 300                             |
| **Calanus glacialis**    | 301                        | 301                        | 301                            | 300                             |
| **Calanus sinicus**      | 301                        | 301                        | 301                            | 300                             |
| **Neocalanus flemingeri**| 301                        | 301                        | 301                            | 300                             |
| **Labidocera madurae**   | 301                        | 301                        | 301                            | 300                             |
| **Tigriopus californicus**| 301                        | 301                        | 301                            | 300                             |
| **Tigriopus japonicus**  | 301                        | 301                        | 301                            | 300                             |
| **Tigriopus king-sejongensis** | 301                    | 301                        | 301                            | 300                             |
| **Paracyclopina nana**   | 301                        | 301                        | 301                            | 300                             |
| **Caligus rogercresseyi**| 301                        | 301                        | 301                            | 300                             |