Molecular physiology of copepods

from biomarker to transcriptomes and back again

Tarrant, Ann M.; Nilsson, Birgitte; Hansen, Benni Winding

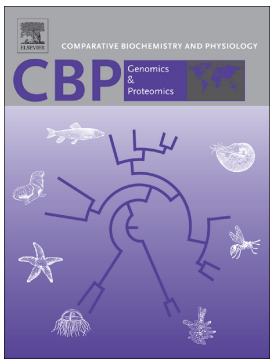
Published in:
Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics

DOI:
10.1016/j.cbd.2019.03.005

Publication date:
2019

Document Version
Peer reviewed version

Citation for published version (APA):
Tarrant, A. M., Nilsson, B., & Hansen, B. W. (2019). Molecular physiology of copepods: from biomarker to transcriptomes and back again. *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics*, *30*(30), 230-247. https://doi.org/10.1016/j.cbd.2019.03.005



Molecular physiology of copepods - from biomarkers to transcriptomes and back again

Ann M. Tarrant, Birgitte Nilsson, Benni Winding Hansen

PII: S1744-117X(18)30161-8
DOI: https://doi.org/10.1016/j.cbd.2019.03.005
Reference: CBD 579

To appear in: *Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics*

Received date: 20 November 2018
Revised date: 14 March 2019
Accepted date: 16 March 2019

Please cite this article as: A.M. Tarrant, B. Nilsson and B.W. Hansen, Molecular physiology of copepods - from biomarkers to transcriptomes and back again, Comparative Biochemistry and Physiology - Part D: Genomics and Proteomics, https://doi.org/10.1016/j.cbd.2019.03.005

Review: Molecular physiology of copepods - from biomarkers to transcriptomes and back again

Ann M. Tarrant[1]*, Birgitte Nilsson[2], Benni Winding Hansen[2]

[1] Biology Department, Woods Hole Oceanographic Institution, 45 Water Street, Woods Hole MA 02544, USA

[2] Department of Science and Environment, Roskilde University. Universitetsvej 1, Roskilde, Denmark

* Corresponding author: atarrant@whoi.edu

Abstract

Planktonic copepods are a diverse and abundant group of small (~mm sized) aquatic animals that play a critical role in linking the base of the food chain with higher trophic levels. These invertebrates are a primary food source for marine fish larvae. Their ubiquitous presence is thus of vital importance for recruitment of fish stocks and also as promising live feed for finfish production in aquaculture. This paper reviews the application of molecular approaches to understanding copepod physiology, particularly in non-parasitic species. The review includes both targeted gene approaches and untargeted transcriptomic approaches, with suggestions for best practices in each case. Issues particularly relevant to studies of copepods include heterogeneity within species, morphologically cryptic species, experimental artifacts associated with sample handling, and limited annotation of copepod genes and transcripts. The emergence of high-throughput sequencing and associated increased availability of genomic and transcriptomic databases has presented a huge opportunity to advance knowledge of copepod physiology. The research community can leverage this opportunity through efforts to maintain or improve data accessibility, database annotation, and documentation of analytical pipelines.

Key words

best practices, biomarker, Copepoda, gene expression, non-model, review, RNA-seq, transcriptome

1. Introduction

Copepods (Subclass Copepoda) are a diverse and ecologically important group of crustaceans that reside freshwater, estuarine and marine environments. They inhabit the water column and the benthos, may be free-living or parasitic, and range from the intertidal to deep ocean basins. They have even been suggested as the most numerous multicellular organisms on earth (Walter and Boxshall, 2019). Several recent molecular studies have indicated that the traditional grouping of Crustacea is paraphyletic, and have suggested a new clade Pancrustacea or Tetraconata that includes hexapods (e.g., insects; Dohle, 2001; Regier et al., 2005). The relationships of copepods to other lineages within the Pancrustacea is still under debate and is considered a matter of active research (Oakley et al., 2012; Regier et al., 2010; Rota-Stabelli et al., 2012). Copepoda is the second largest subclass in the Crustacea, including approximately 12,000 copepod species described to date (Encyclopedia of Life, Accessed 8 June 2018), with the greatest diversity in the marine environment (Boxshall and Defaye, 2008; 11,443 species listed in the World Register of Marine Species, Accessed 8 June 2018).

Within marine and freshwater food webs, small planktonic copepods serve a critical role linking phytoplankton and microzooplankton with larger predators (Sherr and Sherr, 2016; Turner, 2004; Zöllner et al., 2009). Thus, they channel energy from primary producers and the microbial loop up to higher trophic levels. Diel vertical migration of planktonic copepods is a significant conduit for the biological pump, which exports organic carbon below the euphotic zone (reviewed by Steinberg and Landry, 2017). Seasonal dormancy of many species enables efficient grazing of seasonally abundant phytoplankton populations, and within the Calanidae, creates an additional mechanism for export as lipids are respired at depth over a prolonged period

(i.e., the "lipid pump"; Jónasdóttir et al., 2015). In coastal and freshwater ecosystems, many species produce quiescent or diapausing embryos that settle into the sediments, where they remain for months to years until hatching during favorable conditions (Holm et al., 2017). This “egg bank” enables species to adapt to seasonal variability, helps to smooth the effects of variable reproduction across years, and facilitates the coexistence of diverse species and genotypes (Hairston, 1996; Marcus et al., 1994).

In addition to their importance to natural ecosystems, copepods also have significance for aquaculture, due both to their impacts as parasites and to their promising contributions as a source of live feed. Copepods parasitize diverse species in the wild and within aquaculture. Among the parasitic species, caligid copepods (“sea lice”, primarily in the *Lepeophtheirus* and *Caligus* genera) have infested salmonid cultures, where they can both compromise commercial production and spread to local wild populations. Treatments to reduce parasite load can have unintended consequences, such as the development of drug-resistant populations and impacts on non-target species (reviewed by Aaen et al., 2015). On the other hand, cultured free-living copepods can have significant benefits for aquaculture. Relative to more traditional food sources like brine shrimp and rotifers, copepod nauplii can result in increased survival and other quality metrics of fish larvae (Drillet et al., 2011; Nielsen et al., 2017). However, rearing copepods at high densities on a commercial scale requires optimization and is a subject of active research (Nilsson et al., 2017 and references therein; Vu et al., 2017).

Improved understanding of copepod physiology can refine our ability to predict how natural copepod populations will respond to environmental change. Furthermore, increased knowledge of copepod physiology can aid efforts to optimize live-feed production for aquaculture. Aspects of copepod physiology can be monitored using

gross organismal end-points like survival, development, growth, fecundity, respiration and swimming behavior. All of these can provide insight into physiological condition; however, to gain a comprehensive mechanistic understanding of copepod physiology a molecular approach is necessary.

To date, the application of molecular tools to study copepod physiology has been patchy, with different model species (Figure 1) used to study distinct sets of questions and relatively little integration between research communities. Ecologists have focused much attention on the impacts of climate change on species and ecosystems, including the potential for physiological plasticity to mitigate impacts. Elegant studies have demonstrated heritable and plastic components of thermal tolerance in the intertidal copepod *Tigriopus californicus* (Kelly et al., 2017; Lima and Willett, 2017; Pereira et al., 2014, 2017; Schoville et al., 2012; Tangwancharoen et al., 2018). In oceanic environments, investigations of the responses of *Calanus* spp. to the thermal environment have focused on understanding the interactions between physiological tolerances and range shifts in shaping future ecosystems (Ramos et al., 2015; Smolina et al., 2015). Ecotoxicological studies exploring the effects of diverse chemical stressors have most commonly focused on *Tigriopus japonicus* and *Calanus* spp., but a handful of other species have been used (Table 1). Estuarine species, including *Acartia tonsa* and *Eurytemora affinis*, have been studied in many contexts including characterization of responses to salinity changes and handling (Nilsson et al., 2018; Petkeviciute et al., 2015; Rahlff et al., 2017; Xuereb et al., 2012). Many copepods incorporate a dormant stage within their life history, which has consequences for developmental progression and energy utilization. Molecular approaches have been used to study embryonic dormancy in *Acartia tonsa* (Nilsson and Hansen, 2018), juvenile dormancy and lipid utilization in *Calanus finmarchicus* (Tarrant et al., 2008,

2014), and emergence from adult diapause in *Neocalanus flemingeri* (Roncalli et al., 2018b). Finally, molecular approaches have been extensively used to characterize the life history and stress responses of parasitic copepods, (e.g., Núñez-Acuña et al., 2016; Poley et al., 2015). In this case, many of the stress responses studied are intentionally induced with the aim of developing treatments to weaken or impair the propagation of these pests. Overall, most of the species discussed above and throughout this manuscript are calanoids (i.e., members of Order Calanoida), but studies have also been targeted toward a few species of harpacticoids (*Tigriopus* spp. and *Tisbe holothuriae*), siphonostomatiods (especially *Lepeophtheirus salmonis* and *Caligus rogercresseyi*), and cyclopoids (*Apocyclops royi* and *Paracyclopina nana*).

This review seeks to provide a resource that summarizes previous studies of copepod molecular physiology, divided between targeted "candidate gene" approaches and untargeted transcriptomic approaches. To do this, we build upon earlier reviews focused on copepods that articulated the value of emerging genomic resources (Bron et al., 2011; Amato and Carotenuto, 2018) and characterized molecular stress responses (Lauritano et al., 2012). We focus primarily on non-parasitic species that are marine or euryhaline. In synthesizing the results from studies conducted with a wide variety of focal species and for diverse applications, we will both demonstrate how techniques have advanced over time and make suggestions for future study design and data analysis.

2. Common Methodological Considerations

Within this section, we discuss considerations common to both targeted and untargeted gene expression studies. For any gene expression study, it is essential to minimize

artifacts associated with handling of the animals and to maintain high-quality RNA throughout the molecular analysis. Also, in studying the responses of copepods to environmental stressors or other experimental conditions, developmental and sex-specific specificity should be considered. These topics are explored below.

Handling

Studies of copepod physiology have been conducted in a variety of contexts, including direct sampling of wild populations, short-term laboratory manipulations of field-collected animals, and genetically controlled experiments with animals that had been maintained in the laboratory over multiple generations. The potential effects of handling stress have rarely been assessed in copepods. In a limited example, *C. finmarchicus* expression of three small heat shock proteins was shown to increase between the time of collection and 2-3 hours post-collection (Aruda et al., 2011). More broadly, a transcriptomic study of *A. tonsa* demonstrated that intense handling stress created by holding adult copepods outside of water for 10 minutes on Nitex mesh led to substantial changes in gene expression 24 hours later (Nilsson et al., 2018). While the study did not profile handling-induced changes in gene expression under other conditions, elevated mortality was observed in copepods held out of the water for as little as 1 minute. As Nilsson et al. (2018) suggested, the stress imposed by field collections can be high and varies according to factors such as tow speed, mesh size, the temperature during retrieval and processing, and any additional manipulations associated with isolation and preservation. This study highlights the need to minimize stress effects during sampling and experimental manipulation. This can be done through the use of gentle towing methods, maintaining constant temperature and salinity, and minimizing the total time from collection until preservation. To the extent possible, the

effectiveness of any laboratory acclimatization periods should be experimentally validated, and controls for handling should be included.

Selection of Developmental Stage

Previous studies of copepod physiology have varied in the developmental stage(s) that were tested, along with the experimental duration, sampling times, season, and other environmental factors. This diversity in study designs is naturally driven by the diverse objectives of the individual studies, but such differences also make it difficult to compare across species or studies. Among these many factors, developmental stage merits additional discussion in the context of sensitivity to environmental stressors. Within many groups of marine invertebrates, early life stages exhibit increased sensitivity to abiotic stressors, including hypercapnia, extreme temperatures (reviewed by Kurihara, 2008). Among crustaceans, studies in decapods have identified early developmental stages that are particularly vulnerable to thermal stress (e.g., Schiffer et al., 2014; Storch et al., 2011).

In copepods, relatively little research has been devoted to comparing the dynamics of gene expression across developmental stages, but among the available studies, there has been no consistent pattern in stage sensitivity to diverse stressors. For example, Nilsson et al. (2018) found that *A. tonsa* adults were much more sensitive to handling stress than nauplii, while nauplii were more sensitive to salinity stress than adults. Tangwancharoen and Burton (2014) showed that *T. californicus* adults were more sensitive to thermal stress than nauplii and copepodites. Jager et al. (2016) found that within *C. finmarchicus,* adult males were the most sensitive to exposure to fresh and weathered oil, followed by late copepodites. They reviewed numerous studies showing the differential sensitivity of copepod developmental stages to environmental toxicants, including several showing the increased sensitivity of nauplii (e.g., Lotufo and Fleeger,

1997; Saiz et al., 2009). The authors pointed toward the need for additional empirical observations in multiple species and also suggested that the molt from last naupliar stage to the first copepodite stage might be particularly energetically demanding and sensitive to external stressors. The physiological basis for stage-specific sensitivity of copepods to environmental stressors is not generally known, but transcriptional profiling can provide some insight. Using RNA-seq, Roncalli et al. (2017b) observed increased sensitivity of *C. finmarchicus* nauplii to saxitoxin, and also noted that, unlike adults, nauplii did not upregulate digestive enzymes in response to saxitoxin exposure. They hypothesized that upregulation of digestive enzymes by adults reduces assimilation of the toxin and provides increased tolerance.

For lipid-storing copepods, such as *Calanus* spp., accumulation of lipophilic compounds in the oil sac and potential mobilization of these contaminants into adult tissues and offspring may be a significant route of exposure. While studies in this area are just beginning, Hansen et al. (2016) showed that polycyclic aromatic hydrocarbons (PAHs) could be transferred from oil-exposed mothers to offspring and that maternal exposure resulted in mild but measurable effects on naupliar hatching and gene expression. Toxværd et al. (2018) found that exposure of females to pyrene during overwintering lead to a reduced rebuilding of lipid reserves, as well as decreased survival and egg production.

Sex is an additional consideration in studies of adult copepods. Female copepods usually have longer lifespans and higher stress resistances than males (e.g., Foley et al., 2019; Parrish and Wilson, 1978). This difference in resistance, and other physiological differences associated with reproduction may result in different transcriptional responses to stressors. When studying adult copepods, the selection of sex or the choice to include a mixture of sexes, should be carefully considered in light of the scientific

question to be answered. More broadly, understanding of developmental changes in environmental sensitivity is improving, but considerable work is needed to create an integrative view of molecular physiology during copepod development.

Sample preparation

Any measurement of gene expression requires that high RNA quality is maintained throughout sampling, storage, extraction, and subsequent analysis. Copepods have been successfully stored in liquid nitrogen for at least 10 years with high RNA yields, and no evidence on degradation (Hassett et al., 2010); however, preservation in liquid nitrogen is not always tractable. Obtaining liquid nitrogen in remote field locations can be difficult, liquid nitrogen levels must be maintained, and liquid nitrogen storage can pose problems during shipment. Alternative suitable storage methods include guanidine thiocyanate/phenol-based reagents and RNAlater. Zhang et al. (2013) reported that samples preserved in guanidine thiocyanate/phenol-based reagents (e.g. TRI Reagent, TRIzol), could be stored without degradation at 4°C for up to two weeks, or at -80°C for two years. They also reported satisfactory extraction of RNA following storage in RNAlater but noted that copepods stored in RNAlater sometimes become transparent and easy to lose during the extraction protocols. Asai et al. (2015) reported higher RNA yield and improved quality following storage in RNAlater compared with storage in TRIzol, but they did not specify the length of storage prior to extraction. In samples stored in RNAlater, Nilsson et al. (2018) obtained high-quality RNA following 1 week of storage at -20°C, but a noted decreased quality after transport on dry ice and a total of 3 months of storage at -20°C.

With copepod samples, several extraction protocols have been used to obtain total RNA that is of suitable quality for downstream measurements of gene expression (reviewed by Asai et al., 2015; Zhang et al., 2013). While DNAse is frequently used during the

extraction protocol to remove residual genomic DNA before transcript quantification, this step can lead to RNA degradation, and effects on sample quality must be monitored (Nilsson, 2018; Zhang et al., 2013).

Assessment of RNA quality is a critical best practice for both targeted and untargeted gene expression studies; however, RNA quality metrics are not always reported. Spectrophotometry (e.g., NanoDrop™ by ThermoScientific) and fluorometry (e.g., Qubit™ by Invitrogen) enable assessment of RNA yield, and spectrophotometric absorbance ratios indicate sample purity. RNA integrity can be assessed through visualization on denaturing agarose gels, or automated electrophoresis (via Bioanalyzer™ or TapeStation™, both produced by Agilent Technologies). Automated electrophoresis has been used to assess copepod RNA quality for many years (e.g., Voznesensky et al. 2004), and the approach is now common, particularly in association with RNA-seq studies. In automated electrophoresis of total RNA, the most commonly used single metric of quality in the RNA Integrity Number (RIN), which is derived from an algorithm that compares the relative proportions of 28S and 18S rRNA; values range from 1 to 10, with lower values indicating degradation. In copepods, like many other arthropods, the 28S band is fragile and can break ("the hidden break") during sample preparation (Asai et al., 2015; McCarthy et al., 2015). Thus, many copepod studies disregard the RIN metrics and rely on a subjective visual analysis of an electropherogram trace, including the presence of a strong discrete 18S band, the absence of larger bands indicating contamination by genomic DNA, and limited smearing within the smaller size ranges (e.g., Almada and Tarrant 2016, Zhou et al. 2018). In many cases 28S breakage occurs during heat denaturing of samples immediately before analysis, so integrity can sometimes be preserved by omitting this step (e.g, Figure 2 within Asai et al. 2015).

3. Biomarkers

Broadly, biomarkers are detectable molecular, biochemical, and tissue-level changes that indicate physiological effects (Smit et al., 2009). Compared to organismal metrics, biomarkers can provide increased sensitivity to detect changes and specific insights into their likely causes (reviewed by Hook et al., 2014). Quantitative real-time RT-PCR (qPCR) has been the primary approach to measure the expression of individual biomarker genes. Biomarker expression is typically normalized to the expression of one or more reference genes that exhibit stable expression. Below we first discuss criteria for selection of reference genes and methods of normalization. We then review the desirable characteristics of biomarkers and their historical application to studies of copepod physiology. Finally, we point toward additional considerations for future studies.

Reference genes

Relative changes in mRNA levels can be estimated through a variety of methods, including the comparative threshold cycle ($2^{-\Delta\Delta Ct}$) (Livak and Schmittgen, 2001), Pffaffl (Pfaffl, 2001) and LinRegPCR (Ruijter et al., 2009) methods. With each of these methods, users typically account for systematic variation (e.g., differences in starting material, RNA quality, and PCR efficiencies) by normalizing expression against one or more reference genes (Chervoneva et al., 2010; Livak and Schmittgen, 2001).

Commonly used reference genes are often carry-overs from older studies that used semi-quantitative methods, e.g., Northern blots, RNase protection assays, and conventional reverse-transcription PCR assays (Huggett et al., 2005). Suitable reference genes should exhibit stable expression across experimental conditions or

groups to be compared, such as various developmental stages and tissue types. Unfortunately, the stability of reference genes is often insufficiently assessed and the requirement for stability is frequently violated (Dheda et al., 2005; Huggett et al., 2005; Kozera and Rapacz, 2013; Pfaffl et al., 2004; Svingen et al., 2015). Commonly used reference genes for copepods (Table 2) might exhibit stable expression within a set of experimental conditions, but will not necessarily be stable across a different set of conditions or in other species. Thus, it is important to carefully select and validate the reference genes to ensure optimal normalization.

Some recent studies with copepods have used only a single, or few, reference genes for normalization (e.g., Nilsson et al., 2017; Petkeviciute et al., 2015; Rahlff et al., 2017). The limited selection of references genes has historically been due to difficulties in generating suitable primers from species with no available genomic or transcriptomic resources. With the continued improvement of sequencing technologies and lower costs, copepod sequence resources are increasing, which makes it easier to identify and generate new primers for reference genes (e.g., Nilsson and Hansen, 2018). It is strongly recommended to normalize gene expression of target genes against the geometric mean of multiple reference genes (Vandesompele et al., 2002).

Several algorithms are available for selection of the most suitable reference genes, including geNorm (Perkins et al., 2012; Vandesompele et al., 2002), BestKeeper (Pfaffl et al., 2004) and NormFinder (Andersen et al., 2004). Of these, geNorm calculates a gene stability value (M), which is defined as the average pairwise variation of gene expression. The procedure is iterative, where the least-desirable reference gene is discarded, with subsequent recalculation of the M-values. The ranking of M-values is carried out in a step-wise manner starting with the two genes having the lowest pairwise variation. M-values lower than 1.5 are recommended for selecting stable reference

genes (Vandesompele et al., 2002). Furthermore, geNorm is able to estimate how many reference genes should be used for normalization in a given study (Perkins et al., 2012). BestKeeper assumes that reference genes have similar expression patterns; hence suitable reference genes should have highly-correlated expression patterns. From the geometric mean of Ct values and their standard deviation (SD), a "BestKeeper index" is estimated. Genes that are stably expressed have an SD below 1. The genes are compared pairwise, and those with the lowest SD values and highest coefficients of correlation (r) are assumed to exhibit the most stable expression among the candidate genes (Pfaffl et al., 2004). NormFinder uses a statistical linear mixed-effect model to estimate intra- and inter-group variation of gene expression and combines the two into a stability value. The genes with the lowest stability value are assumed to be the most stable across the experimental conditions (Andersen et al., 2004).

Across copepod species and conditions, *elongation factor 1α* (*EFA*, Table 1) and *Histone H3* (*HIST*, Table 1), have been validated as some of the most stable reference genes (Christie et al., 2016; Hansen et al., 2008, 2010; Jeong et al., 2015; Lee et al., 2017). Another commonly used reference gene, that often is validated as stable, is the *18S ribosomal RNA* (*18S*) (e.g., Jeong et al., 2015). However, the expression of *18S* has been shown to be very high expression compared with other candidate reference genes and biomarker genes in studies with multiple copepod species and tested conditions (e.g., Lauritano et al., 2015; Nilsson and Hansen, 2018). This suggests that *18S* is not generally suitable as a reference gene for copepods. Where possible, reference genes should be selected from distinct functional groups to avoid co-regulation (Riemer et al., 2012).

Biomarker selection and application in copepods

Desirable characteristics in a biomarker include sensitivity, a large signal-to-noise ratio, consistency in responses, and known specificity for environmental stressor or other drivers of response. Numerous studies of copepods have reported the expression of small numbers of target genes, which were selected as putative biomarkers of processes of interest, including detoxification, antioxidant activity, apoptosis, and protein refolding (Table 1). Many of the individual genes that have been used as biomarkers belong to larger families (e.g., heat shock protein, cytochrome 450 oxidases). Within large gene families, gene function typically diverges and diversifies, with individual genes developing distinct expression patterns (developmental, tissue-specific and/or subcellular) and functionality (e.g., substrate specificity). These features contribute to the dynamic range of expression for each gene and the environmental conditions that affect that expression.

Initially, due to a lack of genomic resources, copepod genes needed to be individually cloned and sequenced, using degenerate primers based on known sequences in other animals. The genes selected for these studies were necessarily evolutionarily conserved and typically were widely used as biomarkers of similar processes in other animals. For example, heat shock proteins (HSPs) are a deeply conserved superfamily of molecular chaperones that enable proper three-dimensional folding of nascent proteins, help to repair or recycle damaged proteins, contribute to subcellular localization and prevent aggregation (reviewed by Kregel, 2002; Lanneau et al., 2010). While these proteins play essential roles in cellular maintenance, HSPs are also frequently up-regulated in response to diverse cellular stressors. As a biomarker, the best-studied form is the highly inducible cytosolic HSP70. In copepods, induced expression of HSP70 has been reported in response to elevated temperature, crowding, handling, embryonic transition between subitaneous and quiescence states, abnormal salinity, and various chemical

contaminants (e.g., Aruda et al., 2011; Nilsson et al., 2014; Petkeviciute et al., 2015; Rahlff et al., 2017; Rhee et al., 2009, VanderLugt, 2009). In addition to HSP70, several other HSP molecules display changes expression in response to temperature (Seo et al., 2006c, but see also Rhee et al. 2009), handling (Aruda et al., 2011), as well as exposure to endocrine disruptors (Seo et al., 2006b) or toxic diatoms (Lauritano et al., 2011b).

Another broad class of biomarkers is related to antioxidant activity. Reactive oxygen and reactive nitrogen compounds are produced through normal cellular metabolism, through exposure to and metabolism of environmental contaminants, and as a result of exposure to ultraviolet radiation or other physical stressors. Animals have developed several classes of antioxidant enzymes to neutralize these reactive compounds, and accordingly, studies in copepods have measured expression of antioxidant enzymes, including superoxide dismutases (Jiang et al., 2013; Kim et al., 2011), catalases (Hansen et al., 2008; Lauritano et al., 2011b, 2016), glutathione peroxidases (Zhuang et al., 2017) and peroxiredoxins (Zhuang et al., 2017). While exposure to cellular oxidants is broadly expected to lead to induction of antioxidant defenses, the observed patterns are complicated and dependent upon the duration, concentration, and type of stressor, as well as the specific genes, measured.

Genes are often selected as biomarkers based on their specific mode of action to indicate the disruption of a process process or exposure to a specific stressor. For example, xenobiotic metabolizing enzymes often indicate exposure to chemical stressors. Among these, cytochrome P450 oxidases and glutathione S-transferases have been most widely studied in copepods (see Table 2). Vitellogenins are precursors to major egg yolk proteins and have been proposed as markers of the reproductive condition in copepods. Studies in *L. salmonis* have demonstrated that vitellogenins are produced in subcuticular tissues of adult females, secreted into the hemolymph, and

deposited in the maturing oocytes (Dalvin et al., 2011). Vitellogenins have been identified in other copepod species, with measurable expression in late copepodid stages that greatly increases in adult females (e.g., Hwang et al., 2009). Two studies have demonstrated induction of vitellogenin expression in response to metal exposure, but the mechanism of disruption and links to reproductive endpoints are still unclear (Hwang et al., 2010b; Lee et al., 2008a).

Challenges and opportunities

To date, there have been limited instances where biomarkers developed for copepods have been adopted for studies by distinct research groups or across species. As with many physiological studies, experimental differences in factors such as handling protocols, the nutritional status and developmental stage of the animals, and the duration and intensity of any experimental exposure complicate direct comparisons across studies. Two additional consideration merit additional discussion: homologous relationships of biomarkers and genetic complexity of study species.

Homologous relationships

Because full copepod transcriptomic databases have only recently become available, earlier biomarker studies frequently required cloning and sequencing of individual genes. In the case of multi-gene families, this could lead to an analysis of paralogous genes that might not be directly comparable to one another. For example, in studies of *Calanus finmarchicus*, Voznesensky et al. (2004) reported the induction of HSP70 following thermal stress, but Hansen et al. (2008) found no effect of naphthalene exposure on HSP70 expression. While it is enitirely plausible that the two different stressors would induce distinct physiological responses, it would not be evident to a casual reader that the two studies measured different HSP70 family members (Aruda

et al., 2011). This issue of homology becomes even more complex in cross-species comparisons, in which the roles of various gene family members may have diverged.

High-throughput sequencing and the increased availability of copepod transcriptomes and genomes have also provided an opportunity to place biomarkers within a gene family, and more broadly to study gene diversification and loss. As an example, Porter et al. (2017) characterized the evolutionary relationships among phototransduction genes in 10 copepod species from diverse lineages and identified four primary groups of copepod opsins, two of which were broadly distributed, and two of which were restricted to a subset of species.

Genetic complexity

It is becoming increasingly apparent that morphological identification may be insufficient for many copepod species. This is particularly problematic for studies of natural populations, where cryptic species may co-occur. For example, it has recently been demonstrated that morphological characters do not reliably discriminate *C. finmarchicus* and *C. glacialis* and that these species widely co-occur, particularly within fjord environments (Choquet et al., 2017, 2018). Smolina et al. (2015) incorporated this consideration into their methodology, using genetic techniques to verify species identity before pooling RNA from multiple individuals and measuring gene expression.

Heritable physiological variability has been described in both *Acartia tonsa* and *A. hudsonica* (Avery, 2005; Cournoyer, 2013), along with substantial genetic diversity within each group. It has been suggested that the major genetic lineages represent cryptic species, which share broadly overlapping ranges (Chen and Hare, 2011; Milligan et al., 2011). No studies have yet compared these physiological

characteristics of these *Acartia* lineages in a controlled genetic context. In contrast, studies conducted in *Eurytemora affinis* and *Tigriopus californicus*, have frequently incorporated genetic variation into their experimental design, comparing expression in genetically distinct populations or genetically-controlled lineages to study gene by environment interactions, adaptation to novel environments, and predicted responses to climate change (e.g., Kelly et al. 2017; Lee et al. 2011; Pereira et al. 2014, 2017).

Thus, moving forward with natural populations, it will be necessary in many cases to conduct molecular species identifications alongside any other biomarker analyses. This has consequences for experimental design. For example, addressing this concern requires nucleic acid extraction and analysis before pooling any material from individual animals.

4. Transcriptomes

The development of high-throughput sequencing and associated bioinformatic pipelines has revolutionized our understanding of copepod physiology. As copepod physiologists have adopted these new methods, best practices for experimental design, analysis, documentation and data availability have also developed. This section reviews transcriptional profiling approaches that have been applied to copepods, including technical recommendations and suggestions for best practices.

Early untargeted approaches to expression profiling

An advantage of whole-transcriptome profiling is that it enables candidate biomarkers to be identified based on observed expression patterns, without requiring *a priori* selection of candidate genes. Before the development of high-throughput sequencing, gene sequences were typically determined by targeted amplification and cloning using

degenerate primers, or by mining libraries of expressed sequence tags (ESTs) produced through Sanger sequencing. EST libraries also facilitated *de novo* identification of candidate biomarkers, both within suppressive subtractive hybridization (SSH) experiments and by enabling probe design for microarray analysis (e.g., Lenz et al. 2012). Because ESTs are derived from longer Sanger sequences, they have sometimes provided a useful check on the accuracy of transcripts predicted from the assembly of shorter reads produced by high-throughput sequencing (e.g., Christie et al., 2013). Overall, while these older methods are becoming less common, microarrays, in particular, are still useful for some applications and continue to be used.

With SSH, complementary DNA fragments from two different libraries (e.g., treatment and control) are hybridized, and fragments overrepresented within one of the libraries are amplified, cloned and sequenced. This method has been used to identify biomarkers associated with copepod energetics and diapause (*Calanus finmarchicus*, Tarrant et al., 2008), as well as with exposure to nickel (*Pseudodiaptomus annandalei*, Jiang et al., 2013), diethanolamine (*C. finmarchicus*, Hansen et al., 2010), toxic diets (*C. helgolandicus*, Carotenuto et al., 2014), an organophosphate (*Lepeophtheirus salmonis*, Walsh et al., 2007) and multiple stressors (*C. finmarchicus*, Hansen et al., 2007). This approach has produced useful biomarkers: for example, genes identified during an SSH screen of active lipid-storing copepods were subsequently shown to parallel changes in oil sac volume during juvenile copepod development (Tarrant et al., 2014). Limitations of the method include a high rate of false positives and a bias toward genes with high overall expression. For example, Jiang et al. (2013) tested 8 randomly-selected genes from an SSH screen and was able to verify consistent expression patterns for only 5 of them using quantitative PCR (qPCR). In bi-directional screens comparing active and dormant copepods, Tarrant et al. (2008) found that ribosomal genes represented 18%

of the clone libraries, and the highly expressed myosin transcripts represented 32% of the annotated mRNA sequences.

For microarray analysis of gene expression, DNA probes are arrayed onto glass slides and hybridized with fluorescently-labeled cDNA. The probes may be prepared from cDNA libraries or synthetic oligonucleotides, and the hybridization may be with one sample or two distinctly labeled samples. Microarrays have been extensively used to characterize responses of the intertidal copepod *Tigriopus japonicus* to environmental stressors including copper (Ki et al., 2009), manganese (Kim et al., 2013a), ultraviolet radiation (Rhee et al., 2012), and β-naphthoflavone (Rhee et al., 2015). Lee et al. (2011) used a custom cDNA microarray to study changes in Na+K+ ATPase expression associated with invasion of freshwater habitats by *Eurytemora affinis*. A targeted "physiological" microarray was used to measure changes in *Calanus finmarchicus* gene expression associated with food availability, lipid storage, development and vertical migration (Lenz et al., 2012; Unal et al., 2013). A distinct broad-scale array was used to evaluate the effects of components of petroleum extraction on *C. finmarchicus* (Jensen et al., 2016). For the parasitic *Lepeophtheirus salmonis,* microarrays have been used to characterize adult female maturation (Eichner et al., 2008) and abiotic stress responses of free-living larvae (Sutherland et al., 2012). For some species, multiple arrays have been designed, generally increasing in complexity over time. Successive *T. japonicus* arrays produced by the same research group increased from ~6000 35-mer probes (Ki et al., 2009) to ~55,000 60-mer probes (Rhee et al., 2012). For *C. finmarchicus* and *L. salmonis*, earlier arrays were built using cDNA amplified from normalized clone libraries, which were enriched for rare genes (Eichner et al., 2008; Lenz et al., 2012). Later arrays constructed for both of these species used larger numbers of oligonucleotide probes (Jensen et al., 2016; Sutherland et al., 2012).

While microarray is a powerful technique, it requires a substantial initial investment in array design and synthesis. In addition, sample preparation, hybridization, scanning and downstream data processing all require careful optimization and quality control. The popularization of microarray approaches eventually led to the development of analytical pipelines, standards for data reporting, and best practices for data analysis (e.g., Brazma et al., 2001; Knapen et al., 2009; Shi et al., 2008). This increased attention toward data management provided some of the initial frameworks for data analysis and storage associated with high-throughput sequencing approaches. Within NCBI, the Gene Expression Omnibus (GEO) is a public repository for curated gene expression datasets produced using either microarray or sequencing-based platforms (Barrett et al., 2013).

De novo Transcriptome Assembly

To identify current and recent practices in transcriptome assembly, we searched the NCBI transcriptome shotgun assembly (TSA) in August 2018 for available copepod transcriptomes. From each of these, we compiled details of the source biological material, methods of sequencing and assembly, and statistics regarding the number of contigs, and BUSCO score (Table 3, Figure 2). We also included the *Apocyclops royi* transcriptome, which had not yet been publicly released. In total, we identified 19 transcriptome assemblies, corresponding to 13 non-parasitic marine species. These efforts were not uniformly distributed across taxonomic groups; for example, five of the transcriptomes corresponded to two species of *Calanus*, and four of the transcriptomes corresponded to three species of *Tigriopus.* In addition to the TSA's, we identified 6 whole-genome shotgun assemblies (WGS) (*Eurytemora affinis*: GCA_000591075.2, GCA_000591075.2; *Calanus finmarchicus*: GCA_002740975.1,

GCA_002740985.1; *Oithona nana*: GCA_900157175.1; *Acartia tonsa*: GCA_900241095.1) for 4 copepod species.

Sequencing technologies included Illumina-based methods (HiSeq, MiSeq, and NextSeq), 454 pyrosequencing (e.g., 454 GS FLX), and Ion-Torrent. Assembly methods included Trinity, CLC, Mira, Newbler, and CAP3. Of the most recently published assemblies (9 transcriptomes published or posted from 2016-2018), the majority used Illumina-based methods (HiSeq and NextSeq). All of these used paired-end reads. In most of these, (7 of 9) the read length was 150 bp, and in the other cases, the read length was 100 or 125 bp. Trinity was used for all of these assemblies, and one of the studies included both Trinity and CAP3 assemblies.

In the 18 published transcriptomes, the number of contigs ranged dramatically from 28,954 to 554,991. Some of this range represents variation in the extent of coverage and completeness of the transcriptome assembly. One approach to assess the completeness of an assembled transcriptome is the comparison with a curated set of well-conserved single copy eukaryotic genes. To apply this method, we used BUSCO ver. 2 (Benchmarking Universal Single-Copy Orthologs; Simão et al., 2015) to assess the completeness of the same 18 copepod transcriptomes shown in Table 3. Of these, three (*C. finmarchicus,* GBXU*; C. glacialis* GBXT and HACJ) had a large proportion of missing or fragmented genes (79-87% combined). These three assemblies were produced from relatively shallow sequencing (0.7-5 Mb total) and using either Ion-Torrent or 454 GS FLX methods. While these smaller assemblies miss a large number of genes, they can still be useful for targeted applications, such as profiling expression of highly expressed genes and identification of biomarker sequences for qPCR studies (e.g., Smolina et al., 2015). In contrast to the smaller transcriptome assemblies, nine transcriptomes had less than two percent missing BUSCO genes, but also had 20-54%

of the genes represented by more than one transcript. While some of these may represent true lineage-specific duplications, the BUSCO gene set was curated to include genes represented by single orthologs in a vast majority of the diverse animal taxa studied. Thus, a large proportion of the apparent duplications are thought to represent incorrectly assembled haplotypes (Simão et al., 2015). Of the two *C. finmarchicus* transcriptomes with nearly complete BUSCO sets (GAXK and GBFB), the proportion of duplicated genes was lower in the assembly by Lenz et al. (2015; GAXK, 26% duplicated BUSCO genes), which included 206,012 contigs, compared with the assembly by Tarrant et al. (2014; GBFB, 36% duplicated BUSCO genes), which contained 241,140 contigs. Thus, the BUSCO analysis suggests that many of the additional contigs in the second transcriptome may represent duplicates. A likely explanation for this difference is that the smaller transcriptome was filtered post-assembly to retain only the longest contig associated with each clustered component ("comp") produced by the Trinity assembler. While this filtering approach undoubtedly removes many duplicates, it can also result in the removal of some distinct genes (e.g., Lenz et al., 2014). Finally, it should also be noted that the BUSCO score does not directly indicate the completeness of the transcriptome; transcriptome assemblies with high BUSCO scores may still be missing large proportions of rare or conditionally expressed transcripts.

Transcriptome-wide Differential Gene Expression via RNA-seq

To evaluate recent methods for analyzing differential gene expression in copepods, we identified 18 RNA-seq-based studies published between 2012 and 2019 (Table 4; new studies continue to emerge, and this list is not comprehensive). The studies were conducted in a total of 8 species, and they investigated responses to a range of environmental or experimental conditions, including salinity shock, handling stress,

pH, and temperature, as well as experimental exposure to cultured bacteria, toxic dinoflagellates, or organic contaminants. In addition, some studies addressed variation among geographically isolated populations or across developmental stages. All of these studies utilized pools of copepods to construct each library, with the specific number ranging from 3 to 500. Most of the studies (13 of 18 in Table 4) included biological replication, typically with 3 or 4 replicates per treatment. Replication occurred at different levels. For example, Kelly et al. (2017) produced only one library per genetic line and treatment, but the experimental conditions were fully replicated across three genetic lines (i.e., providing full biological replication). Bailey et al. (2017) included 5-6 replicates per treatment, including three experimental replicates (separate aquaria), each with 1-2 libraries. Five studies had only one biological replicate per treatment; however, these often used the differential expression analysis in a more exploratory manner. For example, Smolina et al. (2015) conducted a small-scale transcriptomic study of responses in *Calanus* spp. to different temperature conditions and used the results to develop hypotheses regarding thermal sensitivity and to select genes for more detailed expression profiling by qPCR.

The studies in Table 4 used a variety of methods for read-alignment, calculation of counts per read and identification of differentially expressed sequences; however, not all studies indicated the method used for each step. For alignment, methods included bowtie, bowtie2, bwa, Rsubread, and the proprietary CLC genomics workbench. Two studies used the Kallisto alignment-free method (described below). From the mapping results, matrices of counts per transcript can be generated using RSEM, Rsubread, and other custom scripts; in these analyses, transcripts with very low levels of expression are frequently removed prior to differential expression analysis. Differentially expressed genes are typically identified using R-based packages including

DESeq/DESeq2, edgeR, limma, and Sleuth (Anders and Huber, 2010; Love et al., 2014; Pimentel et al., 2016; Ritchie et al., 2015; Robinson et al., 2010).

Overall, most of these pipelines have included *de novo* transcriptome assembly followed by read mapping, abundance estimation, and differential expression analysis. These methods are time-consuming, require high computational capacity, and can be limited by the quality of the reference genome or transcriptome (Bray et al., 2016; Pimentel et al., 2016). "Alignment-free" methods (e.g., Sailfish, Kallisto, Salmon) have recently been developed, in which the reference transcriptomes are shredded into kmers (Conesa et al., 2016). The kmers from experimental reads can then be matched to the kmers from the transcriptome, resulting in fast and accurate estimations of abundance. Kallisto (e.g., used by Nilsson et al., 2018), constructs a *de Bruijn* graph from transcriptome kmers (abbreviated as t-DBG), and then a pseudo-alignment is constructed in which the kmers from experimental reads are evaluated for compatibility with the t-DBG (Bray et al., 2016). By skipping kmers for which compatibility does not change with the t-DBG, the process is accelerated. Because Kallisto only accounts for exact k-mer matches, most sequencing errors are discarded (Bray et al., 2016). Expression of genes or transcripts is determined by quantifying the k-mers associated with each component of the indexed reference transcriptome. The R package Sleuth is designed for processing the output from Kallisto and analyzing differential expression at the transcript or gene level (Pimentel et al., 2016). Analysis methods continue to develop, and future studies will most likely include a balance between investigators using established pipelines and those choosing to incorporate new methods.

Best Practices and Challenges

Sequencing method and experimental design

Most considerations of sequencing depth and replication are not unique to copepods. Specific concerns are mostly related to working with heterogeneous samples and lacking a well-annotated reference genome for scaffolding. Illumina-based sequencing is the most widely used technique for *de novo* transcriptome assembly, and paired-end, stranded reads of at least 100 bp are recommended (Haas et al., 2013). Studies in diverse animals have shown that *de novo* assemblies derived from 20-40 M reads can typically recover most transcripts (Francis et al., 2013; MacManes, 2016). In evaluating a *Calanus finmarchicus* transcriptome assembly, Lenz et al. (2014) progressively subsampled their data set and found steep increases in the number of assembled contigs obtained when increasing from 6 M to 50 M reads (100 bp, paired-end). They report that a good quality assembly can be constructed from as few as 50 M reads, but suggest based on rarefaction analysis that rare transcripts may be missing even with up to 400 M reads.

In comparing transcript expression among groups of samples, choices must be made to optimize the statistical power of the study within the constraints of funding and sample availability (reviewed by Todd et al., 2016). Several studies have demonstrated that increased replication is more important than sequencing depth in maximizing statistical power. Depending on the type of sample, once 10-20 M reads are obtained, it is generally much more beneficial to increase the number of replicates rather than sequencing depth (Ching et al., 2014; Liu et al., 2013). Another possible element of the experimental design is the use of paired samples or blocking factors. For example, copepods from replicate cultures (or collection sites) may each be split into treatment and control groups that are treated as pairs. In this case, such a design would account for variability among cultures (or collection sites), can increase the signal-to-noise ratio, and thereby can increase statistical power in gene expression studies (Ching et

al., 2014). Paired or blocked study designs should be considered in future studies with copepods.

Assessment of Transcriptome Quality

The best approach for transcriptomic or targeted gene expression studies depends on the available resources as well as the goals of the study. While a variety of methods can be successfully used, most recent high-throughput sequencing studies have used paired-end Illumina-based sequencing coupled with *de novo* assembly using the Trinity software suite. Until recently, few tools were available to assess and compare the quality of various transcriptome assemblies, with many investigators citing the number of contigs along with some transcript size metrics. The most common size metric has been the N50, which was developed for genome assemblies where very long contigs (full chromosomes) are desirable, and sequence representation should be uniform. More recently, the ExN50, which is weighted toward the most abundant transcripts, has been proposed as a more relevant metric for transcriptome assemblies (https://github.com/trinityrnaseq/trinityrnaseq/wiki, Accessed 8 March 2019). Overall, such general assembly statistics provide a useful metric of comparison, but they do not necessarily provide direct insight into the completeness or accuracy of the assembly. BUSCO analysis, as we have used here, provides one means to assess transcriptome completeness and duplication. Other methods include DETONATE (Li et al., 2014) and TransRate (Smith-Unna et al., 2016). The selection of specific tools will depend on factors such as the sequencing platform and the availability of a sequenced reference genome (e.g., Moreton et al., 2016). As sequencing technologies and assembly algorithms continue to develop and improve, these tools can and should be used to inform choices about sequencing platforms, assembly parameters, and post-assembly filtering.

Variability

For eventual comparison across samples, it has been recommended to pool all sequence reads from all samples prior to transcriptome assembly (Haas et al., 2013). However, sequence polymorphism increases the complexity of the *de Bruijn* graph and can negatively affect the assembly (Iqbal et al., 2012; Studholme, 2010). To account for this, MacManes (2016) recommended assembling sequences derived from a single individual. In cases where distinct sets of transcripts may be present in different groups of animals (e.g., developmentally restricted or sex-specific transcripts), sequences should be assembled from one individual per group. For larger-bodied copepods and later developmental stages, it is possible to obtain sufficient RNA for library construction from individual animals, and this relatively new recommendation has been adopted into some of the most recent copepod transcriptome assemblies (e.g., Nilsson et al., 2018; Roncalli et al., 2018a).

In studies that encompass divergent populations, investigators have sometimes utilized the alternative approach of independently assembling population-specific transcriptomes. In several studies comparing isolated and divergent populations of the intertidal copepod *Tigriopus californicus*, investigators have utilized custom analysis pipelines to integrate the transcriptomic databases and enable analysis of orthologous transcript sets (Barreto et al., 2014; DeBiasse et al., 2018; Kelly et al., 2017; Lima and Willett, 2017).

Genomic resources

As with other non-model organisms, a challenge with molecular physiology studies of copepods is the limited availability of genomic resources, including genome assemblies, annotation, and integration of data types. Aspects of this situation are

rapidly improving, particularly with recent successes in genome assembly. At the time of writing, sequenced genomes are available for at least 7 non-parasitic copepod species: *Acartia tonsa, Apocyclops royi, Eurytemora affinis, Oithona nana, Tigriopus californicus, T. japonicus* and *T. kinsejongensis* (Barreto et al., 2018; Eyun et al., 2017; Jørgensen et al., in press; Kang et al., 2017; Lee et al., 2010; Madoui et al., 2017). Among these, the *T. californicus* genome assembly is of particularly high quality, with >94% complete predicted transcripts when compared to the BUSCO arthropod gene set, and >99% of assembled sequence contained within 12 chromosomal scaffolds. The relatively compact genome size of *T. californicus*, thought to be around 200 Mb, undoubtedly facilitated its assembly (Barreto et al. 2018).

In other cases, assembly of copepod genomes has proven more challenging, for reasons including low GC content, small organism size, and sometimes large genomes with high proportions of repetitive DNA (Bron et al., 2011; Jørgensen et al., in press). For example, the complete *A. tonsa* genome is estimated to be nearly 2.5 Gb, with only ~0.5 Gb assembled and non-repetitive sequence (Jørgensen et al., in press). Concerning genome size, another potentially complicating factor is chromatin diminution, the process by which selected heterochromatin regions are eliminated from somatic cells during early embryogenesis. Chromatin diminution has been identified in 23 species of freshwater cyclopoid copepods (reviewed by Grishanin, 2014). The process creates large differences in genome size between the somatic and germ lineages and differences in somatic genome structure among related copepod species (reviewed by Bron et al., 2011). Other considerations affecting genome assembly have included variation associated with pooling heterogeneous individuals and contamination with prey or epibiont sequences. A variety of approaches are being used to address these challenges,

including low-input sequencing, rearing of inbred stocks, careful selection of input material, and improved bioinformatic pipelines for filtering of foreign sequences.

The utility of these newly available genomes and transcriptomes will increase as efforts continue to improve their annotation. In January 2016, a symposium "Tapping the Power of Crustacean Transcriptomes to Address Grand Challenges in Comparative Biology" was convened as part of the annual meeting of the Society of Integrative and Comparative Biology. A key recommendation emerging from an associated workshop was to improve integration of genomic and transcriptomic assemblies to facilitate visualization of gene-specific expression patterns, cross-species comparisons, and identification of novel genes (Mykles et al., 2016). Toward this end, if appropriately leveraged, RNA-seq studies can help to improve gene predictions, reveal alternative splicing and allelic variants, and provide functional insights. Annotation of genomic and transcriptomic databases would benefit from integration with ongoing efforts toward annotating other crustacean databases (e.g., RNA-seq studies in decapods, reviewed by Nguyen et al., 2018) and incorporating emerging results from functional studies conducted in crustaceans (e.g., knockout and knockdown approaches).

Data Availability and Analytical Reproducibility

As the application of high-throughput sequencing has matured, expectations for data availability and analytical reproducibility have increased. These expectations are not unique to studies of copepods (Conesa et al., 2016; Das et al., 2016), but studies published by the community (e.g., those within Tables 3 and 4) have shown increasing documentation of workflows and improved availability of data. Minimally, publications must provide access to raw sequence data and reference databases. Each step of the analysis pipeline must be clearly described, including names and versions of software programs, along with details of options specified within the analysis.

Expectations for sharing of custom code have varied. In some cases, the code is provided only for complex analyses, but increasingly investigators have documented complete analytical pipelines and referenced version-controlled scripts and workflows in repositories such as GitHub.

5. Back Again

Now that both targeted and untargeted approaches at expression profiling have become widely accessible, it is possible to reflect on some lessons learned and directions for future research. In the following sections, we first consider how high-throughput sequencing has greatly accelerated our ability to identify candidate biomarkers that can be used in targeted physiological studies. We then provide some "food for thought" as to how additional physiological insight can be gained from targeted functional studies, characterization of taxonomically restricted genes, and integration of databases.

Building a Better Biomarker

The increasing availability of annotated transcriptomes and transcriptome-wide expression data has greatly informed the selection of biomarkers. For example, Tarrant et al. (2014) compared transcriptomic patterns between two times within a copepodite developmental stage. From this dataset, they identified variable genes related to development and molting and then conducted detailed profiling of a small number of genes during progression through the stage. Similarly, Roncalli et al. (2016) used both transcriptome-wide expression profiling and targeted qPCR to characterize the effects of consumption of toxic algae on the expression of glutathione-S-transferase (GST) enzymes. From the 41 predicted GSTs in *C. finmarchicus*, three were profiled over time in two independent experiments; expression of one of these three genes was induced by

exposure to toxic algae. RNA-seq analysis of a subset of time points corroborated the qPCR analysis and revealed two other GSTs that were induced by exposure to toxic algae. These additional genes could serve as useful biomarkers in future experiments.

From Sequences to Functional Characterization

Beyond selecting biomarkers based on their expression pattern and the known functions of homologous genes in other organisms, new physiological understanding will come from the functional characterization of genes in copepods. Among the possible knockdown/knockout approaches, RNAi has been successfully used for targeted knockdown of copepod transcripts. To date, this approach has primarily been applied to parasitic species (e.g., Eichner et al., 2015; Trösße et al., 2014), but RNAi was also used to confirm the role of HSPb1 in conferring thermal tolerance to the free-living *T. californicus* (Barreto et al. 2014). Direct functional studies provide a much deeper understanding of the roles played by biomarker genes, and have the potential to transform our current understanding of copepod physiology.

Where next?

As curated sequence databases (e.g., Swiss-Prot/Uniprot) have grown, the ability to annotate copepod transcriptomes has improved. As a rough example (significance thresholds and other methodological details varied among studies), early efforts to annotate *Calanus* transcriptomes reported 33-40% of transcripts with positive BLAST hits and only 11-28% of these homologous sequences associated with GO terms (Lenz et al., 2014; Tarrant et al., 2014). More recently, Roncalli et al. (2018a) annotated 57% of coding transcripts against Swiss-Prot, and over 90% of the annotated transcripts were associated with GO terms. Still, the GO annotations are based primarily on knowledge of gene function in model organisms, and many genes remain unannotated. This limited

annotation of copepod genes and transcripts has represented a challenge in harnessing the full power of transcriptomic approaches. Homology-based annotation is by definition biased toward evolutionarily conserved genes, yet within eukaryotes, taxonomically restricted genes (TRGs, also called lineage-specific genes) provide an important source of developmental, physiological and regulatory diversity (Lespinet et al., 2002). TRGs can evolve in association with lineage-specific traits, such as honeybee sociality (Johnson and Tsutsui, 2011) or coral calcification (Moya et al., 2012). In addition, TRGs are strongly associated with adaptation to changing environments (Schlötterer, 2015). For example, TRGs are overrepresented in responses by *C. elegans* to extreme environments (Zhou et al., 2015) and *Daphnia magna* to a suite of environmental perturbations (Orsini et al., 2018). Weighted gene co-expression network analysis (WGCNA) methods can provide powerful insight into regulatory patterns of TRGs by identifying clusters or modules of genes with similar expression patterns. The annotation of conserved genes within individual modules can provide some clues as to potential functions for unannotated genes and can facilitate the selection of biomarkers. For example, investigators can use WGCNA results to identify sets of candidate genes that belong to distinct modules and reflect different physiological processes, or they can select multiple genes within a module to enable more detailed analysis of the association. WGCNA-based methods are increasingly being applied to non-model organisms (e.g., Fuess et al., 2018; Johnson et al., 2018), and will likely become prevalent within future studies in copepods.

Mining underutilized data from older studies, including data generated with older technologies (e.g. microarrays), with machine learning approaches (reviewed by Golestan Hashemi et al., 2018) may enable researchers to further leverage existing data to unlock new insights into copepod physiology. In many cases, data was generated for

answering questions within a targeted study and used in a relatively narrow context. Re-analyzing publicly available data in new ways with new questions has the potential to entrain a broader community of scientists and could lead to novel discoveries without the need for additional sequencing costs and labor. A major challenge in this approach is the harmonization of heterogeneous data from multiple platforms into a unified computational framework to extract the signals. When appropriately harmonized, combining multiple datasets can aid in the construction of expression atlases and identification of regulatory relationships. Expression atlases are frequently built to synthesize spatiotemporal expression patterns within an organism or tissue (e.g., Papatheodorou et al., 2017; Zhang et al., 2014). To date, little is known regarding spatial gene expression patterns within copepods, but a great deal of data is being amassed regarding the environmental, developmental, and experimental conditions associated with expression. Using methods such as WGCNA, described above, genes with similar expression profiles within expression atlases can be associated with a shared function and provisionally annotated (Carnielli et al., 2015; Oliver, 2000). With increasing availability of ‘omics data (e.g., transcriptomic, proteomic and metabolomic) for copepods, another possibility will be to concatenate the information across distinct types of ‘omics analysis (i.e., multi-omics) to gain more information about biological processes, to identify regulatory networks, and to search for robust biomarkers across datasets (Bersanelli et al., 2016).

In conclusion, the tools and databases available for expression profiling studies in copepods have radically advanced over the past ten years. These technological advances are being matched with increased sophistication in analytical approaches as well as improved practices for experimental design, documentation, and data

accessibility. Leveraging these rich datasets will lead to a greatly improved understanding of copepod physiology and copepod responses to environmental change.

Acknowledgements This work was supported by the National Science Foundation (Award Number OPP-1746087) to A.M.T., and the Villum Foundation (Project AMPHICOP no. 8960) to B.W.H.

Figure Legends

Figure 1. Examples of diverse marine copepods utilized in physiological studies. (A) The calanoid *Acartia tonsa* male, (B) The calanoid *Calanus glacialis* C5 copepodite with prominent oil sac, (C) The cyclopoid *Apocyclops royi* egg-bearing female (D) The harpacticoid *Tigriopus japonicus* egg-bearing female. Of these, *Tigriopus spp.* (particularly *T. californicus*, not shown) have been extensively developed as a model for studies of molecular evolution and plasticity. The others represent a growing diversity of species for which molecular physiology studies are being driven by their ecological importance. Photos courtesy of Dr. Minh Thi Thui Vu (A), A.M.T. (B), Dr. Hans van Someren Gréve (C), and Professor Hans Uwe Dahms (D).

Figure 2. BUSCO analysis of the following copepod transcriptomes (with NCBI accession numbers, alphabetized by scientific name as in Table 3): **GFWY**: *Acartia tonsa* (GFWY00000000.1, 27-sep-2017); **HAGX**: *Acartia tonsa* (HAGX00000000.1, 29-sep-2017); **GAXK**: *Calanus finmarchicus* (GAXK00000000.1, 14-may-2018); **GBFB**: *Calanus finmarchicus* (GBFB00000000.1, 30-jan-2015); **GBXU**: *Calanus finmarchicus* (GBXU00000000.1, 13-jan-2015); **GBXT**: *Calanus glacialis* (GBXT00000000.1, 13-jan-2015); **HACJ**: *Calanus glacialis* (HACJ00000000.1, 29-sep-2017); **GBGO**: *Eurytemora affinis* (GBGO00000000.1, 07-jul-2015); **GEAN**: *Eurytemora affinis* (GEAN00000000.1, 16-nov-2016); **GFWO**: *Labidocera madurae* (GFWO00000000.1, 14-may-2018); **GFUD**: *Neocalanus flemingeri* (GFUD00000000.1, 14-may-2018) ; **GCJT**: *Paracyclopina nana* (GCJT00000000.1, 20-jul-2015); **GFCI**: *Pleuromamma xiphias* (GFCI00000000.1, 18-dec-2017); **GBSZ**: *Tigriopus californicus* (GBSZ00000000.1, 02-feb-2015); **GBTC**: *Tigriopus californicus* (GBTC00000000.1, 02-feb-2015); **GCHA**: *Tigriopus japonicus* (GCHA00000000.1, 20-jul-2015) ; **GDFW**: *Tigriopus kingsejongensis* (GDFW00000000.1, 18-apr-2016); **HAHV**: *Tisbe holothuriae* (HAHV00000000.1, 23-jan-2018).

References Cited

Aaen, S.M., Helgesen, K.O., Bakke, M.J., Kaur, K., Horsberg, T.E., 2015. Drug resistance in sea lice: a threat to salmonid aquaculture. Trends Parasitol. 31, 72-81.

Aguilera, V.M., Vargas, C.A., Lardies, M.A., Poupin, M., 2016. Adaptive variability to low-pH river discharges in *Acartia tonsa* and stress responses to high pCO_2 conditions. Mar. Ecol. 37, 215-226.

Almada, A.A., Tarrant, A.M., 2016. *Vibrio* colonists elicit targeted transcriptional responses from copepod hosts. FEMS Microbiol. Ecol. 92, fiw072.

Amato, A., and Carotenuto, Y., 2018. Planktonic calanoids embark into the 'omics' era. In: Trends in Copepod Studies - Distribution, Biology and Ecology, edited by M. Uttieri. New York: Nova Science Publishers, Inc. pp 287-314. ISBN: 978-1 53612-593-1.

Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. Genome Biol. 11, R106.

Andersen, C.L., Jensen, J.L., Ørntoft, T.F., 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. 64, 5245-5250.

Aruda, A.M., Baumgartner, M.F., Reitzel, A.M., Tarrant, A.M., 2011. Heat shock protein expression during stress and diapause in the marine copepod *Calanus finmarchicus*. J. Insect Physiol. 57, 665-675.

Asai, S., Ianora, A., Lauritano, C., Lindeque, P.K., Carotenuto, Y., 2015. High-quality RNA extraction from copepods for Next Generation Sequencing: a comparative study. Mar. Genom. 24, 115-118.

Avery, D.E., 2005. Induction of embryonic dormancy in the calanoid copepod *Acartia hudsonica*: heritability and phenotypic plasticity in two geographically separated populations. J. Exp. Mar. Biol. Ecol. 314, 215-225.

Bailey, A., De Wit, P., Thor, P., Browman, H.I., Bjelland, R., Shema, S., Fields, D.M., Runge, J.A., Thompson, C., Hop, H., 2017. Regulation of gene expression is associated with tolerance of the Arctic copepod *Calanus glacialis* to CO_2-acidified sea water. Ecol. Evol. 7, 7145-7160.

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A., 2012. NCBI GEO: archive for functional genomics data sets—update. Nucl. Acids Research. 41, D991-5.

Barreto, F.S., Pereira, R.J., Burton, R.S., 2014. Hybrid dysfunction and physiological compensation in gene expression. Mol. Biol. Evol. 32, 613-622.

Barreto, F.S., Schoville, S.D., Burton, R.S., 2015. Reverse genetics in the tide pool: knock-down of target gene expression via RNA interference in the copepod *Tigriopus californicus*. Mol. Ecol. Res. 15, 868-879.

Barreto, F.S., Watson, E.T., Lima, T.G., Willett, C.S., Edmands, S., Li, W., Burton, R.S., 2018. Genomic signatures of mitonuclear coevolution across populations of *Tigriopus californicus*. Nature Ecol. Evol. *2*, 1250-1257.

Bersanelli, M., Mosca, E., Remondini, D., Giampieri, E., Sala, C., Castellani, G., Milanesi, L., 2016. Methods for the integration of multi-omics data: mathematical aspects. BMC Bioinformatics 17, S15.

Borchel, A., Komisarczuk, A.Z., Rebl, A., Goldammer, T., Nilsen, F., 2018. Systematic identification and characterization of stress-inducible heat shock proteins (HSPs) in the salmon louse (*Lepeophtheirus salmonis*). Cell Stress Chaperones 23, 127-139.

Boxshall, G.A., Defaye, D., 2008. Global diversity of copepods (Crustacea: Copepoda) in freshwater. Hydrobiologia 595, 195-207.

Bray, N.L., Pimentel, H., Melsted, P., Pachter, L., 2016. Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34, 525.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., 2001. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. Nat. Genet. 29, 365.

Bron, J.E., Frisch, D., Goetze, E., Johnson, S.C., Lee, C.E., Wyngaard, G.A., 2011. Observing copepods through a genomic lens. Front. Zool. 8, 22.

Carnielli, C.M., Winck, F.V., Leme, A.F.P., 2015. Functional annotation and biological interpretation of proteomics data. Biochimica et Biophysica Acta -Proteins Proteomics 1854, 46-54.

Carotenuto, Y., Dattolo, E., Lauritano, C., Pisano, F., Sanges, R., Miralto, A., Procaccini, G., Ianora, A., 2014. Insights into the transcriptome of the marine copepod *Calanus helgolandicus* feeding on the oxylipin-producing diatom *Skeletonema marinoi*. Harmful Algae 31, 153-162.

Chan, C., Pankey, S., Kelly, M., 2014. Identifying the gene (s) that allow *Tigriopus californicus* to survive under thermal stress. Can. Young Sci. J. 2014, 18-23.

Chen, G., Hare, M.P., 2011. Cryptic diversity and comparative phylogeography of the estuarine copepod *Acartia tonsa* on the US Atlantic coast. Mol. Ecol. 20, 2425-2441.

Chervoneva, I., Li, Y., Schulz, S., Croker, S., Wilson, C., Waldman, S.A., Hyslop, T., 2010. Selection of optimal reference genes for normalization in quantitative RT-PCR. BMC Bioinformatics 11, 253.

Ching, T., Huang, S., Garmire, L.X., 2014. Power analysis and sample size estimation for RNA-Seq differential expression. RNA 20, 1-13.

Choquet, M., Hatlebakk, M., Dhanasiri, A.K., Kosobokova, K., Smolina, I., Søreide, J.E., Svensen, C., Melle, W., Kwaśniewski, S., Eiane, K., 2017. Genetics redraws pelagic biogeography of *Calanus*. Biol. Lett. 13, 20170588.

Choquet, M., Kosobokova, K., Kwaśniewski, S., Hatlebakk, M., Dhanasiri, A.K., Melle, W., Daase, M., Svensen, C., Søreide, J.E., Hoarau, G., 2018. Can morphology reliably distinguish between the copepods *Calanus finmarchicus* and *C. glacialis*, or is DNA the only way? Limnol. Oceanogr. Methods 16, 237-252.

Christie, A.E., Roncalli, V., Lenz, P.H., 2016. Diversity of insulin-like peptide signaling system proteins in *Calanus finmarchicus* (Crustacea; Copepoda)–Possible contributors to seasonal pre-adult diapause. Gen. Comp. Endocrinol. 236, 157-173.

Christie AE, Roncalli V, Wu LS, Ganote CL, Doak T, Lenz PH., 2013. Peptidergic signaling in *Calanus finmarchicus* (Crustacea, Copepoda): in silico identification of putative peptide hormones and their receptors using a de novo assembled transcriptome. Gen. Comp. Endocrinol. 187,117-35.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M.W., Gaffney, D.J., Elo, L.L., Zhang, X., 2016. A survey of best practices for RNA-seq data analysis. Genome Biol. 17, 13.

Cournoyer, B.L., 2013. Can *Acartia* spp. adapt to global warming? Heritable within-population genetic variation in life history traits. M.S. Thesis, University of Connecticut, 48 pp.

Dalvin, S., Frost, P., Loeffen, P., Skern-Mauritzen, R., Baban, J., Rønnestad, I., Nilsen, F., 2011. Characterisation of two vitellogenins in the salmon louse *Lepeophtheirus salmonis*: molecular, functional and evolutional analysis. Dis. Aquat. Org. 94, 211-224.

Das, S., Shyamal, S., Durica, D.S., 2016. Analysis of annotation and differential expression methods used in RNA-Seq studies in crustacean systems. Integr. Comp. Biol. 56, 1067-1079.

De Wit, P., Dupont, S., Thor, P., 2016. Selection on oxidative phosphorylation and ribosomal structure as a multigenerational response to ocean acidification in the common copepod *Pseudocalanus acuspes*. Evol. App. 9, 1112-1123.

DeBiasse, M.B., Kawji, Y., Kelly, M.W., 2018. Phenotypic and transcriptomic responses to salinity stress across genetically and geographically divergent *Tigriopus californicus* populations. Mol. Ecol. 27, 1621-1632.

Dheda, K., Huggett, J., Chang, J., Kim, L., Bustin, S., Johnson, M., Rook, G., Zumla, A., 2005. The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. Anal. Biochem. 344, 141-143.

Dohle, W., 2001. Are the insects terrestrial crustaceans? A discussion of some new facts and arguments and the proposal of the proper name 'Tetraconata'for the monophyletic unit Crustacea+ Hexapoda. J Ann. Soc. Entomol. Fr. 37, 85-103.

Drillet, G., Frouël, S., Sichlau, M.H., Jepsen, P.M., Højgaard, J.K., Joarder, A.K., Hansen, B.W., 2011. Status and recommendations on marine copepod cultivation for use as live feed. Aquaculture 315, 155-166.

Eichner C, Dalvin S, Skern-Mauritzen R, Malde K, Kongshaug H, Nilsen F., 2015. Characterization of a novel RXR receptor in the salmon louse (*Lepeophtheirus salmonis,* Copepoda) regulating growth and female reproduction. BMC Genomics. 16, 81.

Eichner, C., Frost, P., Dysvik, B., Jonassen, I., Kristiansen, B., Nilsen, F., 2008. Salmon louse (*Lepeophtheirus salmonis*) transcriptomes during post molting maturation and egg production, revealed using EST-sequencing and microarray analysis. BMC Genomics 9, 126.

Eyun, S.I., Soh, H.Y., Posavi, M., Munro, J.B., Hughes, D.S., Murali, S.C., Qu, J., Dugan, S., Lee, S.L., Chao, H., Dinh, H., 2017. Evolutionary history of chemosensory-related gene families across the Arthropoda. Mol. Biol. Evol. 34, 1838-1862.

Foley, H. B., Sun, P. Y., Ramirez, R., So, B. K., Venkataraman, Y. R., Nixon, E. N., Davies, K.J.A., Edmands, S., 2019. Sex-specific stress tolerance, proteolysis, and lifespan in the invertebrate *Tigriopus californicus*. Exp. Gerontol. 119, 146-156.

Francis, W.R., Christianson, L.M., Kiko, R., Powers, M.L., Shaner, N.C., Haddock, S.H., 2013. A comparison across non-model animals suggests an optimal sequencing depth for *de novo* transcriptome assembly. BMC Genomics 14, 1.

Fuess, L.E., Mann, W.T., Jinks, L.R., Brinkhuis, V., Mydlarz, L.D., 2018. Transcriptional analyses provide new insight into the late-stage immune response of a diseased Caribbean coral. R. Soc. Open Sci. 5, 172062.

Golestan Hashemi, F.S., Razi Ismail, M., Rafii Yusop, M., Golestan Hashemi, M.S., Nadimi Shahraki, M.H., Rastegari, H., Miah, G., Aslani, F., 2018. Intelligent mining of large-scale bio-data: Bioinformatics applications. Biotechnol. Biotechnol. Equip. 32, 10-29.

Grishanin A., 2014. Chromatin diminution in Copepoda (Crustacea): pattern, biological role and evolutionary aspects. Comp. Cytogenetics 8:1.

Haas, B.J., Papanicoulaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8, 1494-1512.

Hairston, N.G.J., 1996. Zooplankton egg banks as biotic reservoirs in changing environments. Limnol. Oceanogr. 41, 1087-1092.

Han, J., Won, E.-J., Kim, H.-S., Nelson, D.R., Lee, S.-J., Park, H.G., Lee, J.-S., 2015a. Identification of the full 46 cytochrome P450 (CYP) complement and modulation of CYP expression in response to water-accommodated fractions of crude oil in the cyclopoid copepod *Paracyclopina nana*. Environ. Sci. Technol. 49, 6982-6992.

Han, J., Won, E.-J., Lee, M.-C., Seo, J.S., Lee, S.-J., Lee, J.-S., 2015b. Developmental retardation, reduced fecundity, and modulated expression of the defensome in the intertidal copepod *Tigriopus japonicus* exposed to BDE-47 and PFOS. Aquat. Toxicol. 165, 136-143.

Hansen, B., Altin, D., Vang, S.-H., Nordtug, T., Olsen, A., 2008. Effects of naphthalene on gene transcription in *Calanus finmarchicus* (Crustacea: Copepoda). Aquat. Toxicol. 86, 157-165.

Hansen, B.H., Altin, D., Booth, A., Vang, S.-H., Frenzel, M., Sørheim, K.R., Brakstad, O.G., Størseth, T.R., 2010. Molecular effects of diethanolamine exposure on *Calanus finmarchicus* (Crustacea: Copepoda). Aquat. Toxicol. 99, 212-222.

Hansen, B.H., Altin, D., Nordtug, T., Olsen, A.J., 2007. Suppression subtractive hybridization library prepared from the copepod *Calanus finmarchicus* exposed to a sublethal mixture of environmental stressors. Comp. Biochem. Physiol. D: Genom. Proteom. 2, 250-256.

Hansen, B.H., Altin, D., Rørvik, S.F., Øverjordet, I.B., Olsen, A.J., Nordtug, T., 2011. Comparative study on acute effects of water accommodated fractions of an artificially weathered crude oil on *Calanus finmarchicus* and *Calanus glacialis* (Crustacea: Copepoda). Sci. Total Environ. 409, 704-709.

Hansen, B.H., Jager, T., Altin, D., Øverjordet, I.B., Olsen, A.J., Salaberria, I., Nordtug, T., 2016. Acute toxicity of dispersed crude oil on the cold-water copepod *Calanus finmarchicus*: elusive implications of lipid content. J. Toxicol. Environ. Health Part A 79, 549-557.

Hansen, B.H., Nordtug, T., Altin, D., Booth, A., Hessen, K.M., Olsen, A.J., 2009. Gene expression of GST and CYP330A1 in lipid-rich and lipid-poor female *Calanus finmarchicus* (Copepoda: Crustacea) exposed to dispersed oil. J. Toxicol. Environ. Health Part A 72, 131-139.

Hassett, R. P., P. H. Lenz, and D. W. Towle., 2010. Gene expression and biochemical studies of the marine copepod *Calanus finmarchicus*. MDI Biol. Lab. Bull. 49: 115–117.

Holm, M.W., Kiørboe, T., Brun, P., Licandro, P., Almeda, R., Hansen, B.W., 2017. Resting eggs in free living marine and estuarine copepods. J. Plank. Res. 40, 2-15.

Hook, S.E., Gallagher, E.P., Batley, G.E., 2014. The role of biomarkers in the assessment of aquatic ecosystem health. Integr. Environ. Assess. Manage. 10, 327-341.

Huggett, J., Dheda, K., Bustin, S., Zumla, A., 2005. Real-time RT-PCR normalisation; strategies and considerations. Genes and Immunity 6, 279.

Hwang, D.-S., Han, J., Won, E.-J., Kim, D.-H., Jeong, C.-B., Hwang, U.-K., Zhou, B., Choe, J., Lee, J.-S., 2016. BDE-47 causes developmental retardation with down-regulated expression profiles of ecdysteroid signaling pathway-involved nuclear receptor (NR) genes in the copepod *Tigriopus japonicus*. Aquat. Toxicol. 177, 285-294.

Hwang, D.-S., Lee, J.-S., Rhee, J.-S., Han, J., Lee, Y.-M., Kim, I.-C., Park, G.S., Lee, J., Lee, J.-S., 2010a. Modulation of p53 gene expression in the intertidal copepod *Tigriopus japonicus* exposed to alkylphenols. Mar. Environ. Res. 69, S77-S80.

Hwang, D.-S., Lee, K.-W., Han, J., Park, H.G., Lee, J., Lee, Y.-M., Lee, J.-S., 2010b. Molecular characterization and expression of vitellogenin (Vg) genes from the cyclopoid copepod, *Paracyclopina nana* exposed to heavy metals. Comp. Biochem. Physiol. C: Pharmacol. Toxicol. Endocrinol.
151, 360-368.

Hwang DS, Lee KW, Lee JS. 2009. Cloning and expression of vitellogenin 2 gene from the intertidal copepod *Tigriopus japonicus*. Annals New York Acad. Sci. 1163, 417-20.

Jager, T., Altin, D., Miljeteig, C., Hansen, B.H., 2016. Stage-dependent and sex-dependent sensitivity to water-soluble fractions of fresh and weathered oil in the marine copepod *Calanus finmarchicus*. Environ. Toxicol. Chem. 35, 728-735.

Jensen, L.K., Halvorsen, E., Song, Y., Hallanger, I.G., Hansen, E.L., Brooks, S.J., Hansen, B.H., Tollefsen, K.E., 2016. Individual and molecular level effects of produced water contaminants on nauplii and adult females of *Calanus finmarchicus*. J. Toxicol. Environ. Health Part A 79, 585-601.

Jeong, C.-B., Kang, H.-M., Seo, J.S., Park, H.G., Rhee, J.-S., Lee, J.-S., 2016. Identification and molecular characterization of nitric oxide synthase (NOS) gene in the intertidal copepod *Tigriopus japonicus*. Gene 577, 47-54.

Jeong, C.-B., Kim, B.-M., Kim, R.-K., Park, H.G., Lee, S.-J., Shin, K.-H., Leung, K.M.Y., Rhee, J.-S., Lee, J.-S., 2014. Functional characterization of P-glycoprotein in the intertidal copepod *Tigriopus japonicus* and its potential role in remediating metal pollution. Aquat. Toxicol. 156, 135-147.

Jeong, C.-B., Lee, M.C., Lee, K.-W., Seo, J.S., Park, H.G., Rhee, J.-S., Lee, J.-S., 2015. Identification and molecular characterization of dorsal and dorsal-like genes in the cyclopoid copepod *Paracyclopina nana*. Mar. Genom. 24, 319-327.

Jiang, J.-L., Wang, G.-Z., Mao, M.-G., Wang, K.-J., Li, S.-J., Zeng, C.-S., 2013. Differential gene expression profile of the calanoid copepod, *Pseudodiaptomus annandalei*, in response to nickel exposure. Comp. Biochem. Physiol. C: Toxicol. Physiol. 157, 203-211.

Johnson, B.R., Tsutsui, N.D., 2011. Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. BMC Genomics 12, 164.

Johnson, K.M., Wong, J.M., Hoshijima, U., Sugano, C.S., Hofmann, G.E., 2018. Seasonal transcriptomes of the Antarctic pteropod, *Limacina helicina antarctica*. Mar. Environ. Res.

Jónasdóttir, S.H., Vissir, A.W., Richardson, K., Heath, M.R., 2015. Seasonal copepod lipid pump promotes carbon sequestration in the deep North Atlantic. Proc Nat Acad Sci 112, 12122-12126.

Jørgensen, T.S., Nielsen, B.L.H., Petersen, B., Browne, P.D., Hansen, B.W., Hansen, L.H., In press. The whole genome sequence and mRNA transcriptome of the tropical cyclopoid copepod *Apocyclops royi*. G3, manuscript in press.

Jørgensen, T.S., Petersen, B., Hansen, L.H., Hansen, B.W., unpublished data. The genome and transcriptome of the cosmopolitan calanoid copepod *Acartia tonsa* Dana expand the understanding of copepod genome size evolution. Manuscript in revision, submitted to Genome Biol. Evol.

Kang, S., Ahn, D.H., Lee, J.H., Lee, S.G., Shin, S.C., Lee, J., Min, G.S., Lee, H., Kim, H.W., Kim, S., Park, H., 2017. The genome of the Antarctic-endemic copepod, *Tigriopus kingsejongensis*. Gigascience, 6, giw010.

Kelly, M.W., Pankey, M.S., DeBiasse, M.B., Plachetzki, D.C., 2017. Adaptation to heat stress reduces phenotypic and transcriptional plasticity in a marine copepod. Funct. Ecol. 31, 398-406.

Ki, J.-S., Raisuddin, S., Lee, K.-W., Hwang, D.-S., Han, J., Rhee, J.-S., Kim, I.-C., Park, H.G., Ryu, J.-C., Lee, J.-S., 2009. Gene expression profiling of copper-induced responses in the intertidal copepod *Tigriopus japonicus* using a 6K oligochip microarray. Aquat. Toxicol. 93, 177-187.

Kim, B.-M., Choi, B.-S., Lee, K.-W., Ki, J.-S., Kim, I.-C., Choi, I.-Y., Rhee, J.-S., Lee, J.-S., 2013a. Expression profile analysis of antioxidative stress and developmental pathway genes in the manganese-exposed intertidal copepod *Tigriopus japonicus* with 6K oligochip. Chemosphere 92, 1214-1223.

Kim, B.-M., Jeong, C.-B., Han, J., Kim, I.-C., Rhee, J.-S., Lee, J.-S., 2013b. Role of crustacean hyperglycemic hormone (CHH) in the environmental stressor-exposed intertidal copepod *Tigriopus japonicus*. Comp. Biochem. Physiol. C: Toxicol. Physiol. 158, 131-141.

Kim, B.-M., Jeong, C.-B., Lee, M.C., Rhee, J.-S., Lee, J.-S., 2015a. Identification of the retinoblastoma (Rb) gene and expression in response to environmental stressors in the intertidal copepod *Tigriopus japonicus*. Mar. Genom. 24, 387-396.

Kim, B.-M., Rhee, J.-S., Jeong, C.-B., Seo, J.S., Park, G.S., Lee, Y.-M., Lee, J.-S., 2014. Heavy metals induce oxidative stress and trigger oxidative stress-mediated heat shock protein (hsp) modulation in the intertidal copepod *Tigriopus japonicus*. Comp. Biochem. Physiol. C: Toxicol. Physiol. 166, 65-74.

Kim, B.-M., Rhee, J.-S., Park, G.S., Lee, J., Lee, Y.-M., Lee, J.-S.J.C., 2011. Cu/Zn- and Mn-superoxide dismutase (SOD) from the copepod *Tigriopus japonicus:* molecular cloning and expression in response to environmental pollutants. Chemosphere 84, 1467-1475.

Kim, H.-S., Lee, B.-Y., Han, J., Lee, Y.H., Min, G.-S., Kim, S., Lee, J.-S., 2016. De novo assembly and annotation of the Antarctic copepod (*Tigriopus kingsejongensis*) transcriptome. Mar. Genom. 28, 37-39.

Kim, H.-S., Lee, B.-Y., Won, E.-J., Han, J., Hwang, D.-S., Park, H.G., Lee, J.-S., 2015b. Identification of xenobiotic biodegradation and metabolism-related genes in the copepod *Tigriopus japonicus* whole transcriptome analysis. Mar. Genom. 24, 207-208.

Knapen, D., Vergauwen, L., Laukens, K., Blust, R., 2009. Best practices for hybridization design in two-colour microarray analysis. Trends Biotechnol. 27, 406-414.

Kozera, B., Rapacz, M., 2013. Reference genes in real-time PCR. J. Appl. Genet. 54, 391-406.

Kregel, K.C., 2002. Invited review: heat shock proteins: modifying factors in physiological stress responses and acquired thermotolerance. J. Appl. Physiol. 92, 2177-2186.

Kurihara, H., 2008. Effects of CO_2-driven ocean acidification on the early developmental stages of invertebrates. Mar. Ecol. Prog. Ser. 373, 275-284.

Lanneau, D., Wettstein, G., Bonniaud, P., Garrido, C., 2010. Heat shock proteins: cell protection through protein triage. Sci. World J. 10, 1543-1552.

Lauritano, C., Borra, M., Carotenuto, Y., Biffali, E., Miralto, A., Procaccini, G., Ianora, A., 2011a. First molecular evidence of diatom effects in the copepod *Calanus helgolandicus*. J. Exp. Mar. Biol. Ecol. 404, 79-86.

Lauritano, C., Borra, M., Carotenuto, Y., Biffali, E., Miralto, A., Procaccini, G., Ianora, A., 2011b. Molecular evidence of the toxic effects of diatom diets on gene expression patterns in copepods. PLoS ONE 6, e26850.

Lauritano, C., Carotenuto, Y., Procaccini, G., Turner, J.T., Ianora, A., 2013. Changes in expression of stress genes in copepods feeding upon a non-brevetoxin-producing strain of the dinoflagellate *Karenia brevis*. Harmful Algae 28, 23-30.

Lauritano, C., Carotenuto, Y., Vitiello, V., Buttino, I., Romano, G., Hwang, J.-S., Ianora, A., 2015. Effects of the oxylipin-producing diatom *Skeletonema marinoi* on gene expression levels of the calanoid copepod *Calanus sinicus*. Mar. Genom. 24, 89-94.

Lauritano, C., Procaccini, G., Ianora, A., 2012. Gene expression patterns and stress response in marine copepods. Mar. Environ. Res. 76, 22-31.

Lauritano, C., Romano, G., Roncalli, V., Amoresano, A., Fontanarosa, C., Bastianini, M., Braga, F., Carotenuto, Y., Ianora, A., 2016. New oxylipins produced at the end of a diatom bloom and their effects on copepod reproductive success and gene expression levels. Harmful Algae 55, 221-229.

Lee, B.-Y., Kim, H.-S., Choi, B.-S., Hwang, D.-S., Choi, A.Y., Han, J., Won, E.-J., Choi, I.-Y., Lee, S.-H., Om, A.-S., 2015. RNA-seq based whole transcriptome analysis of the cyclopoid copepod *Paracyclopina nana* focusing on xenobiotics metabolism. Comp. Biochem. Physiol. D: Genom. Proteom. 15, 12-19.

Lee, C.E., Kiergaard, M., Gelembiuk, G.W., Eads, B.D., Posavi, M., 2011. Pumping ions: rapid parallel evolution of ionic regulation following habitat invasions. Evolution 65, 2229-2244.

Lee, K.-W., Hwang, D.-S., Rhee, J.-S., Ki, J.-S., Park, H.G., Ryu, J.-C., Raisuddin, S., Lee, J.-S., 2008a. Molecular cloning, phylogenetic analysis and developmental expression of a vitellogenin (Vg) gene from the intertidal copepod *Tigriopus japonicus*. Comp. Biochem. Physiol. B: Biochem. Mol. Biol. 150, 395-402.

Lee, K.-W., Raisuddin, S., Rhee, J.-S., Hwang, D.-S., Yu, I.T., Lee, Y.-M., Park, H.G., Lee, J.-S., 2008b. Expression of glutathione S-transferase (GST) genes in the marine copepod *Tigriopus japonicus* exposed to trace metals. Aquat. Toxicol. 89, 158-166.

Lee, K.W., Rhee, J.S., Raisuddin, S., Park, H.G., Lee, J.S., 2008c. A corticotropin-releasing hormone binding protein (CRH-BP) gene from the intertidal copepod, *Tigriopus japonicus*. Gen. Comp. Endocrinol. 158, 54-60.

Lee, K.-W., Rhee, J.-S., Han, J., Park, H.G., Lee, J.-S., 2012. Effect of culture density and antioxidants on naupliar production and gene expression of the cyclopoid copepod, *Paracyclopina nana*. Comp. Biochem. Physiol. A: Mol. Integr. Physiol. 161, 145-152.

Lee, J.S., Rhee, J.S., Kim, R.O., Hwang, D.S., Han, J., Choi, B.S., Park, G.S., Kim, I.C., Park, H.G., Lee, Y.M., 2010. The copepod *Tigriopus japonicus* genomic DNA information (574 Mb) and molecular anatomy. Mar. Environ. Res. 69, S21-S23.

Lee, M.-C., Puthumana, J., Lee, S.-H., Kang, H.-M., Park, J.C., Jeong, C.-B., Han, J., Hwang, D.-S., Seo, J.S., Park, H.G., 2016. BDE-47 induces oxidative stress, activates MAPK signaling pathway, and elevates de novo lipogenesis in the copepod *Paracyclopina nana*. Aquat. Toxicol. 181, 104-112.

Lee, Y.-M., Lee, K.-W., Park, H., Park, H.G., Raisuddin, S., Ahn, I.-Y., Lee, J.-S., 2007. Sequence, biochemical characteristics and expression of a novel Sigma-class of glutathione S-transferase from the intertidal copepod, *Tigriopus japonicus* with a possible role in antioxidant defense. Chemosphere 69, 893-902.

Lee, Y.-M., Park, T.-J., Jung, S.-O., Seo, J.S., Park, H.G., Hagiwara, A., Yoon, Y.-D., Lee, J.-S., 2006. Cloning and characterization of glutathione S-transferase gene in the intertidal copepod *Tigriopus japonicus* and its expression after exposure to endocrine-disrupting chemicals. Mar. Environ. Res. 62, S219-S223.

Lee, Y.H., Kang, H.-M., Kim, D.-H., Wang, M., Jeong, C.-B., Lee, J.-S., 2017. Adverse effects of methylmercury (MeHg) on life parameters, antioxidant systems, and MAPK signaling pathways in the copepod *Tigriopus japonicus*. Aquat. Toxicol. 184, 133-141.

Legrand, E., Forget-Leray, J., Duflot, A., Olivier, S., Thomé, J.-P., Danger, J.-M., Boulangé-Lecomte, C., 2016. Transcriptome analysis of the copepod *Eurytemora affinis* upon exposure to endocrine disruptor pesticides: Focus on reproduction and development. Aquat. Toxicol. 176, 64-75.

Lenz, P.H., Roncalli, V., Hassert, R.P., Wu, L.S., Cieslak, M.C., Hartline, D.K., Christie, A.E., 2014. *De novo* assembly of a transcriptome for *Calanus finmarchicus* (Crustacea, Copepoda) - The dominant zooplanker of the North Atlantic Ocean. PLoS ONE 9, e88589.

Lenz, P.H., Unal, E., Hassett, R.P., Smith, C.M., Bucklin, A., Christie, A.E., Towle, D.W., 2012. Functional genomics resources for the North Atlantic copepod, *Calanus finmarchicus:* EST database and physiological microarray. Comp. Biochem. Physiol. D: Genom. Proteom. 7, 110-123.

Lespinet, O., Wolf, Y.I., Koonin, E.V., Aravind, L., 2002. The role of lineage-specific gene family expansion in the evolution of eukaryotes. Genome Res. 12, 1048-1059.

Li, B., Fillmore, N., Bai, Y., Collins, M., Thomson, J.A., Stewart, R., Dewey, C.N., 2014. Evaluation of de novo transcriptome assemblies from RNA-Seq data. Genome Biol. 15, 553.

Lima, T.G., Willett, C.S., 2017. Locally adapted populations of a copepod can evolve different gene expression patterns under the same environmental pressures. Ecol. Evol. 7, 4312-4325.

Liu, Y., Zhou, J., White, K.P., 2013. RNA-seq differential expression studies: more sequence or more replication? Bioinformatics 30, 301-304.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-DDC(T)) Method. Methods 25, 402-408.

Lotufo, G.R., Fleeger, J.W., 1997. Effects of sediment-associated phenanthrene on survival, development and reproduction of two species of meiobenthic copepods. Mar. Ecol. Prog. Ser. 151, 91-102.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.

Maas, A.E., Blanco-Bercial, L., Lo, A., Tarrant, A.M., Timmins-Schiffman, E., 2018. Variations in copepod proteome and respiration rate in association with diel vertical migration and circadian cycle. Biol. Bull. 235, 30-42.

MacManes, M.D., 2016. Establishing evidenced-based best practice for the de novo assembly and evaluation of transcriptomes from non-model organisms. bioRxiv, 035642.

Madoui, M.A., Poulain, J., Sugier, K., Wessner, M., Noel, B., Berline, L., Labadie, K., Cornils, A., Blanco- Bercial, L., Stemmann, L. and Jamet, J.L., 2017. New insights into global biogeography, population structure and natural selection from the genome of the epipelagic copepod *Oithona*. Mol. Ecol. 26(17), 4467-4482.

Marcus, N.H., Lutz, R., Burnett, W., Cable, P., 1994. Age, viability, and vertical distribution of zooplankton resting eggs from an anoxic basin: Evidence of an egg bank. Limnol. Oceanogr. 39, 154-158.

Mauchline, J., 1998. The Biology of Calanoid Copepods. Advances in Marine Biology, Volume 33, Academic Press, London UK ISBN 30-12-105545-105540, 105710 pp.

McCarthy, S.D., Dugon, M.M., Power, A.M., 2015. 'Degraded' RNA profiles in Arthropoda and beyond. PeerJ 3, e1436.

Milligan, P.J., Stahl, E.A., Schizas, N.V., Turner, J.T., 2011. Phylogeography of the copepod *Acartia hudsonica* in estuaries of the northeastern United States. Hydrobiologia 666, 155-165.

Moreton, J., Izquierdo, A., Emes, R.D., 2016. Assembly, assessment, and availability of de novo generated eukaryotic transcriptomes. Front. Genet. 6, 361.

Moya, A., Huisman, L., Ball, E., Hayward, D., Grasso, L., Chua, C., Woo, H., Gattuso, J.P., Foret, S., Miller, D.J., 2012. Whole transcriptome analysis of the coral *Acropora millepora* reveals complex responses to CO2- driven acidification during the initiation of calcification. Mol. Ecol. 21, 2440-2454.

Mykles, D.L., Burnett, K.G., Durica, D.S., Joyce, B.L., McCarthy, F.M., Schmidt, C.J., Stillman, J.H., 2016. Resources and recommendations for using transcriptomics to address grand challenges in comparative biology. Integr. Comp. Biol. 56, 1183-1191.

Nguyen, T.V., Jung, H., Rotllant, G., Hurwood, D., Mather, P., Ventura, T., 2018. Guidelines for RNA-seq projects: applications and opportunities in non-model decapod crustacean species. Hydrobiologia, 1-23.

Nielsen, R., Nielsen, M., Abate, T.G., Hansen, B.W., Jepsen, P.M., Nielsen, S.L., Støttrup, J.G., Buchmann, K., 2017. The importance of live- feed traps–farming marine fish species. Aquaculture Research 48, 2623-2641.

Nilsson, B., 2018. Molecular Stress Physiology in the Calanoid Copepod *Acartia tonsa.* PhD Dissertation, Roskilde University, 148 pp.

Nilsson, B., Hansen, B.W., 2018. Timing of embryonic quiescence determines viability of embryos from the calanoid copepod, *Acartia tonsa* (Dana). PLoS ONE 13, e0193727.

Nilsson, B., Jakobsen, H.H., Stief, P., Drillet, G., Hansen, B.W., 2017. Copepod swimming behavior, respiration, and expression of stress-related genes in response to high stocking densities. Aquacult. Rep. 6, 35-42.

Nilsson, B., Jepsen, P.M., Bucklin, A., Hansen, B.W., 2018. Environmental stress responses and experimental handling artifacts of a model organism, the copepod *Acartia tonsa* (Dana). Front. Mar. Sci. 5, 156.

Nilsson, B., Jepsen, P.M., Rewitz, K., Hansen, B.W., 2014. Expression of hsp70 and ferritin in embryos of the copepod *Acartia tonsa* (Dana) during transition between subitaneous and quiescent state. J. Plank. Res. 36, 513-522.

Núñez-Acuña, G., Vera-Bizama, F., Boltaña, S., Hawes, C., Marambio, J.P., Wadsworth, S., Gallardo-Escárate, C., 2016. In-feed additives modulate ionotropic receptor genes from the sea louse *Caligus rogercresseyi*: a comparative analysis in two host salmonid species. Aquaculture 451, 99-105.

Oakley, T.H., Wolfe, J.M., Lindgren, A.R., Zaharoff, A.K.J.M.b., 2012. Phylotranscriptomics to bring the understudied into the fold: monophyletic Ostracoda, fossil placement, and pancrustacean phylogeny. Mol. Biol. Evol. 30, 215-233.

Oliver, S., 2000. Proteomics: guilt-by-association goes global. Nature 403, 601.

Orsini, L., Brown, J.B., Shams Solari, O., Li, D., He, S., Podicheti, R., Stoiber, M.H., Spanier, K.I., Gilbert, D., Jansen, M., 2018. Early transcriptional response pathways in *Daphnia magna* are coordinated in networks of crustacean-specific genes. Mol. Ecol. 27, 886-897.

Papatheodorou, I., Fonseca, N.A., Keays, M., Tang, Y.A., Barrera, E., Bazant, W., Burke, M., Füllgrabe, A., Fuentes, A.M.-P., George, N., 2017. Expression Atlas: gene and protein expression across multiple studies and organisms. Nucleic Acids Res. 46, D246-D251.

Park, J.C., Han, J., Lee, M.-C., Seo, J.S., Lee, J.-S., 2017. Effects of triclosan (TCS) on fecundity, the antioxidant system, and oxidative stress-mediated gene expression in the copepod *Tigriopus japonicus*. Aquat. Toxicol. 189, 16-24.

Parrish, K.K., Wilson, D.F., 1978. Fecundity studies on *Acartia tonsa* (Copepoda: Calanoida) in standardized culture. Mar. Biol. 46, 65–81.

Pereira, R.J., Barreto, F.S., Burton, R.S., 2014. Ecological novelty by hybridization: experimental evidence for increased thermal tolerance by transgressive segregation in *Tigriopus californicus*. Evolution 68, 204-215.

Pereira, R.J., Sasaki, M.C., Burton, R.S., 2017. Adaptation to a latitudinal thermal gradient within a widespread copepod species: the contributions of genetic divergence and phenotypic plasticity. Proc. Royal Soc. London B: Biol. Sci. 284, 20170236.

Perkins, J.R., Dawes, J.M., McMahon, S.B., Bennett, D.L., Orengo, C., Kohl, M., 2012. ReadqPCR and NormqPCR: R packages for the reading, quality checking and normalisation of RT-qPCR quantification cycle (Cq) data. BMC Genomics 13, 296.

Petkeviciute, E., Kania, P.W., Skovgaard, A., 2015. Genetic responses of the marine copepod *Acartia tonsa* (Dana) to heat shock and epibiont infestation. Aquacult. Rep. 2, 10-16.

Pfaffl, M., 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29, e45.

Pfaffl, M.W., Tichopad, A., Prgomet, C., Neuvians, T.P., 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper–Excel-based tool using pair-wise correlations. Biotechnol. Lett. 26, 509-515.

Pimentel, H., Sturmfels, P., Bray, N., Melsted, P., Pachter, L., 2016. The Lair: a resource for exploratory analysis of published RNA-Seq data. BMC Bioinformatics 17, 490.

Poley, J.D., Igboeli, O.O., Fast, M.D., 2015. Towards a consensus: Multiple experiments provide evidence for constitutive expression differences among sexes and populations of sea lice (*Lepeophtheirus salmonis*) related to emamectin benzoate resistance. Aquaculture 448, 445-450.

Puthumana, J., Lee, M.-C., Han, J., Kim, H.-S., Hwang, D.-S., Lee, J.-S., 2017. Ecdysone receptor (EcR) and ultraspiracle (USP) genes from the cyclopoid copepod *Paracyclopina nana*: Identification and expression in response to water accommodated fractions (WAFs). Comp. Biochem. Physiol. C: Toxicol. Physiol. 192, 7-15.

Rahlff, J., Peters, J., Moyano, M., Pless, O., Claussen, C., Peck, M.A., 2017. Short-term molecular and physiological responses to heat stress in neritic copepods *Acartia tonsa* and *Eurytemora affinis*. Comp. Biochem. Physiol. A: Mol. Integr. Physiol. 203, 348-358.

Ramos, A.A., Weydmann, A., Cox, C.J., Canario, A.V.M., Serrao, E.A., Pearson, G.A., 2015. A transcriptome resource for the copepod *Calanus glacialis* across a range of culture temperatures. Mar. Genomics 23, 27-29.

Regier, J.C., Shultz, J.W., Kambic, R.E., 2005. Pancrustacean phylogeny: hexapods are terrestrial crustaceans and maxillopods are not monophyletic. Proc. Royal Soc. London B: Biol. Sci. 272, 395-401.

Regier, J.C., Shultz, J.W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J.W., Cunningham, C.W., 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. Nature 463, 1079.

Rhee, J.-S., Kim, B.-M., Choi, B.-S., Lee, J.-S., 2012. Expression pattern analysis of DNA repair-related and DNA damage response genes revealed by 55 K oligomicroarray upon UV-B irradiation in the intertidal copepod, *Tigriopus japonicus*. Comp. Biochem. Physiol. C: Toxicol. Pharmacol. 155, 359-368.

Rhee, J.-S., Raisuddin, S., Lee, K.-W., Seo, J.S., Ki, J.-S., Kim, I.-C., Park, H.G., Lee, J.-S., 2009. Heat shock protein (Hsp) gene responses of the intertidal copepod *Tigriopus japonicus* to environmental toxicants. Comp. Biochem. Physiol. C: Toxicol. Pharmacol. 149, 104-112.

Rhee, J.-S., Yu, I.T., Kim, B.-M., Jeong, C.-B., Lee, K.-W., Kim, M.-J., Lee, S.-J., Park, G.S., Lee, J.-S., 2013. Copper induces apoptotic cell death through reactive oxygen species-triggered oxidative stress in the intertidal copepod *Tigriopus japonicus*. Aquat. Toxicol. 132, 182-189.

Rhee, J.S., Lee, Y.M., Kim, B.M., Leung, K.M.Y., Kim, I.C., Yim, J.H., Lee, J.S., 2015. β- Naphthoflavone induces oxidative stress in the intertidal copepod, *Tigriopus japonicus*. Environ. Toxicol. 30, 332-342.

Riemer, A.B., Keskin, D.B., Reinherz, E.L., 2012. Identification and validation of reference genes for expression studies in human keratinocyte cell lines treated with and without interferon- γ–a method for qRT- PCR reference gene determination. Exp. Dermatology 21, 625-629.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47-e47.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139-140.

Roncalli, V., Christie, A.E., Sommer, S.A., Cieslak, M.C., Hartline, D.K., Lenz, P.H., 2017a. A deep transcriptomic resource for the copepod crustacean *Labidocera madurae*: A potential indicator species for assessing near shore ecosystem health. PloS ONE 12, e0186794.

Roncalli, V., Cieslak, M.C., Lenz, P.H., 2016a. Transcriptomic responses of the calanoid copepod *Calanus finmarchicus* to the saxitoxin producing dinoflagellate *Alexandrium fundyense*. Sci. Rep. 6, 25708.

Roncalli, V., Cieslak, M.C., Sommer, S.A., Hopcroft, R.R., Lenz, P.H., 2018a. *De novo* transcriptome assembly of the calanoid copepod *Neocalanus flemingeri*: A new resource for emergence from diapause. Mar. Genom. 37, 114-119.

Roncalli, V., Jungbluth, M.J., Lenz, P.H., 2016b. Glutathione S-Transferase regulation in *Calanus finmarchicus* feeding on the toxic dinoflagellate *Alexandrium fundyense*. PLoS ONE 11, e0159563.

Roncalli, V., Lenz, P.H., Cieslak, M.C., Hartline, D.K., 2017b. Complementary mechanisms for neurotoxin resistance in a copepod. Sci. Rep. 7, 14201.

Roncalli, V., Sommer, S.A., Cieslak, M.C., Clarke, C., Hopcroft, R.R., Lenz, P.H., 2018b. Physiological characterization of the emergence from diapause: A transcriptomics approach. Sci. Rep. 8, 12577.

Rota-Stabelli, O., Lartillot, N., Philippe, H., Pisani, D., 2012. Serine codon-usage bias in deep phylogenomics: pancrustacean relationships as a case study. Syst. Biol. 62, 121-133.

Ruijter, J., Ramakers, C., Hoogaars, W., Karlen, Y., Bakker, O., Van den Hoff, M., Moorman, A., 2009. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res. 37, e45-e45.

Saiz, E., Movilla, J., Yebra, L., Barata, C., Calbet, A., 2009. Lethal and sublethal effects of naphthalene and 1, 2-dimethylnaphthalene on naupliar and adult stages of the marine cyclopoid copepod *Oithona davisae*. Environ. Pollut. 157, 1219-1226.

Schiffer, M., Harms, L., Lucassen, M., Mark, F.C., Pörtner, H.-O., Storch, D., 2014. Temperature tolerance of different larval stages of the spider crab *Hyas araneus* exposed to elevated seawater pCO_2. Front. Zool. 11, 87.

Schlötterer, C., 2015. Genes from scratch–the evolutionary fate of *de novo* genes. Trends Genet. 31, 215-219.

Schoville, S.D., Barreto, F.S., Moy, G.W., Wolff, A., Burton, R.S., 2012. Investigating the molecular basis of local adaptation to thermal stress: population differences in gene expression across the transcriptome of the copepod *Tigriopus californicus*. BMC Evol. Biol. 12, 170.

Seo, J., Park, T.-J., Lee, Y.-M., Park, H., Yoon, Y.-D., Lee, J.-S., 2006a. Small heat shock protein 20 gene (Hsp20) of the intertidal copepod *Tigriopus japonicus* as a possible biomarker for exposure to endocrine disruptors. Bull. Environ. Contam. Toxicol. 76.

Seo, J.S., Lee, K.-W., Rhee, J.-S., Hwang, D.-S., Lee, Y.-M., Park, H.G., Ahn, I.-Y., Lee, J.-S., 2006b. Environmental stressors (salinity, heavy metals, H2O2) modulate expression of glutathione reductase (GR) gene from the intertidal copepod *Tigriopus japonicus*. Aquat. Toxicol. 80, 281-289.

Seo, J.S., Lee, Y.-M., Park, H.G., Lee, J.-S., 2006c. The intertidal copepod *Tigriopus japonicus* small heat shock protein 20 gene (Hsp20) enhances thermotolerance of transformed *Escherichia coli*. Biochem. Biophys. Res. Comm. 340, 901-908.

Sherr, E.B., Sherr, B.F., 2016. Phagotrophic protists: central roles in microbial food webs, Aquatic Microbial Ecology and Biogeochemistry: A Dual Perspective. Springer, 3-12.

Shi, L., Perkins, R.G., Fang, H., Tong, W., 2008. Reproducible and reliable microarray results through quality control: good laboratory proficiency and appropriate data analysis practices are essential. Curr. Opin. Biotechnol. 19, 10-18.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210-3212.

Smit, M.G., Bechmann, R.K., Hendriks, A.J., Skadsheim, A., Larsen, B.K., Baussant, T., Bamber, S., Sanni, S., 2009. Relating biomarkers to whole-organism effects using species sensitivity distributions: A pilot study for marine species exposed to oil. Environ. Toxicol. Chem. 28, 1104-1109.

Smith-Unna, R., Boursnell, C., Patro, R., Hibberd, J.M., Kelly, S., 2016. TransRate: reference-free quality assessment of *de novo* transcriptome assemblies. Genome Res. 26, 1134-1144.

Smolina, I., Kollias, S., Moller, E.F., Lindeque, P., Arvind, Y.M., Fernandes, J.M.O., Hoarau, G., 2015. Contrasting transcriptome response to thermal stress in two key zooplankton species *Calanus finmarchicus* and *C. glacialis*. Mar. Ecol. Prog. Ser. 534, 79-93.

Steinberg, D.K., Landry, M.R., 2017. Zooplankton and the ocean carbon cycle. Ann. Rev. Mar. Sci. 9, 413-444.

Storch, D., Fernández, M., Navarrete, S.A., Pörtner, H.-O., 2011. Thermal tolerance of larval stages of the Chilean kelp crab *Taliepus dentatus*. Mar. Ecol. Prog. Ser. 429, 157-167.

Semmouria, I., Asselmana J., Nieuwerburgh, F., Deforce, D., Janssen C.R., De Schamphelaere, K.A.C., 2019. The transcriptome of the marine calanoid copepod *Temora longicornis* under heat stress and recovery. Mar. Environ. Res. 143, 10-23.

Sutherland, B.J., Jantzen, S.G., Yasuike, M., Sanderson, D.S., Koop, B.F., Jones, S.R., 2012. Transcriptomics of coping strategies in free-swimming *Lepeophtheirus salmonis* (Copepoda) larvae responding to abiotic stress. Mol. Ecol. 21, 6000-6014.

Svingen, T., Letting, H., Hadrup, N., Hass, U., Vinggaard, A.M., 2015. Selection of reference genes for quantitative RT-PCR (RT-qPCR) analysis of rat tissues under physiological and toxicological conditions. PeerJ 3, e855.

Tangwancharoen, S., Burton, R.S., 2014. Early life stages are not always the most sensitive: heat stress responses in the copepod *Tigriopus californicus*. Mar. Ecol. Prog. Ser. 517, 75-83.

Tangwancharoen, S., Moy, G.W., Burton, R.S., 2018. Multiple modes of adaptation: regulatory and structural evolution in a small heat shock protein gene. Mol. Biol. Evol. 35, 2110-2119.

Tarrant, A., Baumgartner, M., Verslycke, T., Johnson, C., 2008. Differential gene expression in diapausing and active *Calanus finmarchicus* (Copepoda). Mar. Ecol. Prog. Ser. 355, 193-207.

Tarrant, A.M., Baumgartner, M.F., Hansen, B.H., Altin, D., Nordtug, T., Olsen, A.J., 2014. Transcriptional profiling of reproductive development, lipid storage and molting throughout the last juvenile stage of the marine copepod *Calanus finmarchicus*. Front. Zool. 11, 91.

Todd, E.V., Black, M.A., Gemmell, N.J., 2016. The power and promise of RNA-seq in ecology and evolution. Mol. Ecol. 25, 1224-1241.

Toxværd, K., Van Dinh, K., Henriksen, O., Hjorth, M., Nielsen, T., 2018. Impact of pyrene exposure during overwintering of the Arctic copepod *Calanus glacialis*. Environ. Sci. Technol.

Tribble, N.D., Burka, J.F., Kibenge, F.S.B., 2007. Evidence for changes in the transcription levels of two putative P-glycoprotein genes in sea lice (*Lepeophtheirus salmonis*) in response to emamectin benzoate exposure. Mol. Biochem. Parasitol. 153.

Trösße C, Nilsen F, Dalvin S., 2014. RNA interference mediated knockdown of the KDEL receptor and COPB2 inhibits digestion and reproduction in the parasitic copepod *Lepeophtheirus salmonis*. Comp. Biochem. Physiol. B: Biochem. Mol. Biol. 170:1-9.

Turner, J.T., 2004. The importance of small planktonic copepods and their roles in pelagic marine food webs. Zool. Stud. 43, 255-266.

Unal, E., Bucklin, A., Lenz, P.H., Towle, D.W., 2013. Gene expression of the marine copepod *Calanus finmarchicus:* Responses to small-scale environmental variation in the Gulf of Maine (NW Atlantic Ocean). J. Exp. Mar. Biol. Ecol. 446, 76-85.

VanderLugt, K. , Cooney, M. J., Lechner, A. and Lenz, P. H., 2009, Cultivation of the paracalanid copepod, *Bestiolina similis* (Calanoida: Crustacea). J. World Aquacult. Soc., 40: 616-628.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3, RESEARCH0034.

Voznesensky, M., Lenz, P.H., Spanings-Pierrot, C., Towle, D.W., 2004. Genomic approaches to detecting thermal stress in *Calanus finmarchicus* (Copepoda: Calanoida). J. Exp. Mar. Biol. Ecol. 311, 37-46.

Vu, M.T., Hansen, B.W., Kiørboe, T., 2017. The constraints of high density production of the calanoid copepod *Acartia tonsa* Dana. J. Plank. Res. 39, 1028-1039.

Walsh, T.K., Lyndon, A.R., Jamieson, D.J., 2007. Identification of cDNAs induced by the organophosphate trichlorphon in the parasitic copepod *Lepeophtheirus salmonis* (Copepoda; Caligidae). Pestic. Biochem. Physiol. 88, 26-30.

Walter, T.C., Boxshall, G., 2019. World of Copepods database. Accessed at http://www.marinespecies.org/copepoda on 2019-01-31.

Weaver, R.J., Hill, G.E., Kuan, P.-L., Tseng, Y.-C., 2016. Copper exposure reduces production of red carotenoids in a marine copepod. Ecol. Indic. 70, 393-400.

Willett, C.S., Burton, R.S., 2002., Proline biosynthesis genes and their regulation under salinity stress in the euryhaline copepod *Tigriopus californicus*. Comp. Biochem.Physiol.B: Biochem. Mol. Biol. 132,739-750.

Willett, C.S., Burton, R.S., 2003. Characterization of the glutamate dehydrogenase gene and its regulation in a euryhaline copepod. Comp. Biochem. Physiol. B: Biochem. Mol. Biol. 135, 639-646.

Xuereb, B., Forget-Leray, J., Souissi, S., Glippa, O., Devreker, D., Lesueur, T., Marie, S., Danger, J.-M., Boulangé-Lecomte, C., 2012. Molecular characterization and mRNA expression of grp78 and hsp90A in the estuarine copepod *Eurytemora affinis*. Cell Stress Chaperones 17, 457-472.

Yi, A.X., Han, J., Lee, J.-S., Leung, K.M., 2014. Ecotoxicity of triphenyltin on the marine copepod *Tigriopus japonicus* at various biological organisations: from molecular to population-level effects. Ecotoxicology 23, 1314-1325.

Zhang, H., Finiguerra, M., Dam, H.G., Huang, Y., Xu, D., Liu, G., Lin, S., 2013. An improved method for achieving high-quality RNA for copepod transcriptomic studies. J. Exp. Mar. Biol. Ecol. 446, 57-66.

Zhang, R., Lahens, N.F., Ballance, H.I., Hughes, M.E., Hogenesch, J.B., 2014. A circadian gene expression atlas in mammals: implications for biology and medicine. Proc. Nat. Acad. Sci. 111, 16219-16224.

Zhou, C., Carotenuto, Y., Vitiello, V., Wu, C., Zhang, J. Buttino, I., 2018. De novo transcriptome assembly and differential gene expression analysis of the calanoid copepod *Acartia tonsa* exposed to nickel nanoparticles. Chemosphere 209, 163-172.

Zhou, K., Huang, B., Zou, M., Lu, D., He, S., Wang, G., 2015. Genome-wide identification of lineage-specific genes within *Caenorhabditis elegans*. Genomics 106, 242-248.

Zhou, K., Wang, M., Sun, S., 2016. Effects of elevated temperature and food supply on the termination of over-summering and subsequent development of the calanoid copepod *Calanus sinicus*: morphology, physiology and gene expression. PLoS ONE 11, e0161838.

Zhuang, Y., Yang, F., Xu, D., Chen, H., Zhang, H., Liu, G., 2017. Spliced leader-based analyses reveal the effects of polycyclic aromatic hydrocarbons on gene expression in the copepod Pseudodiaptomus poplesia. Aquat. Toxicol. 183, 114-126.

Zöllner, E., Hoppe, H.-G., Sommer, U., Jürgens, K., 2009. Effect of zooplankton-mediated trophic cascades on marine microbial food web components (bacteria, nanoflagellates, ciliates). Limnol. Oceanogr. 54, 262-275.

Table 1. Commonly used reference genes in copepods.

Gene	Species	Reference	Validation
ß-actin (ACT)	*Acartia tonsa*	(Nilsson et al., 2018)	Yes
		(Aguilera et al., 2016)	No
		(Nilsson et al., 2014, 2017; Petkeviciute et al., 2015; Rahlff et al., 2017)	No
	Calanus finmarchicus	(Hansen et al., 2008; Roncalli et al., 2016b)	Yes
	Calanus helgolandicus	(Lauritano et al., 2011a, 2013, 2016)	Yes
	Calanus sinicus	(Lauritano et al., 2015)	Yes
	Eurytemora affinis	(Rahlff et al., 2017)	No
	Paracyclopina nana	(Hwang et al., 2010b)	No
		(Jeong et al., 2015)	Yes
	Pseudodiaptomus annandalei	(Jiang et al., 2013)	No
	Pseudodiaptomus poplesia	(Zhuang et al., 2017)	Yes
	Tigriopus californicus	(Chan et al., 2014)	No
	Tigriopus japonicus	(Lee et al., 2007, 2008b; Rhee et al., 2009; Seo et al., 2006b)	No
		(Jeong et al., 2014; Lee et al., 2017)	Yes
Elongation factor 1α (EFA)	*Acartia tonsa*	(Nilsson et al., 2014, 2017; Petkeviciute et al., 2015)	No
	Calanus finmarchicus	(Hansen et al., 2008, 2010; Roncalli et al., 2016b)	Yes
	Calanus helgolandicus	(Lauritano et al., 2011a, 2013, 2016)	Yes
	Calanus sinicus	(Zhou et al., 2016)	No
		(Lauritano et al., 2015)	Yes
	Lepeophtheirus salmonis	(Tribble et al., 2007)	No
		(Borchel et al., 2018; Park et al., 2017)	Yes
	Paracyclopina nana	(Jeong et al., 2015)	Yes
	Pseudodiaptomus poplesia	(Zhuang et al., 2017)	Yes
	Tigriopus japonicus	(Park et al., 2017)	No
		(Jeong et al., 2015; Lee et al., 2017)	Yes
18S rRNA (18S)	*Calanus helgolandicus*	(Lauritano et al., 2011a., 2013, 2016)	Yes
		(Lee et al., 2017)	Yes
	Calanus sinicus	(Lauritano et al., 2015; Lee et al., 2017)	Yes
	Paracyclopina nana	(Han et al., 2015a; Lauritano et al., 2015; Lee et al., 2012, 2016; Puthumana et al., 2017)	No
		(Jeong et al., 2015; Lauritano et al., 2011a, 2013, 2016)	Yes
	Tigriopus japonicus	(Han et al., 2015a, 2015a; Hwang et al., 2016; Jeong et al., 2014, 2015, 2016; Kim et al., 2011, 2013a, 2013b, 2014, 2015a; Lee et al., 2012, 2016; Puthumana et al., 2017; Rhee et al., 2013; Yi et al., 2014)	No
Ribosomal protein S16 (S16)	*Calanus finmarchicus*	(Aruda et al., 2011; Roncalli et al., 2016b)	No
		(Roncalli et al., 2016b)	Yes
	Calanus sinicus	(Zhou et al., 2016)	No
Ribosomal protein S20 (S20)	*Calanus helgolandicus*	(Lauritano et al., 2011a., 2013, 2016)	Yes
	Calanus sinicus	(Lauritano et al., 2015)	Yes
	Pseudodiaptomus poplesia	(Zhuang et al., 2017)	Yes
Ribosomal protein S7 (S7)	*Calanus helgolandicus*	(Lauritano et al., 2011a., 2013, 2015, 2016)	Yes
	Calanus sinicus	(Lauritano et al., 2015)	Yes
Glyceraldehyde -3-phosphate dehydrogenase (GAPDH)	*Calanus helgolandicus*	(Lauritano et al., 2011a, 2013, 2016)	Yes
	Calanus sinicus	(Zhou et al., 2016)	No
		(Jeong et al., 2015; Lauritano et al., 2015)	Yes
	Paracyclopina nana	(Jeong et al., 2015)	Yes
	Pseudodiaptomus annandalei	(Jiang et al., 2013)	No
	Tigriopus californicus	(Barreto et al., 2015)	No
	Tigriopus japonicus	(Lee et al., 2006, 2017; Seo et al., 2006a, 2006c)	No
Ubiquitin (UBI)	*Calanus helgolandicus*	(Lauritano et al., 2011a., 2013)	Yes

	Calanus sinicus	(Lauritano et al., 2015)	Yes
Histone H3 (HIST)	*Acartia tonsa*	(Nilsson and Hansen, 2018)	Yes
	Calanus helgolandicus	(Lauritano et al., 2011a, 2013, 2015, 2016)	Yes
	Calanus sinicus	(Lauritano et al., 2015)	Yes
ATP synthase (ATPS)	*Acartia tonsa*	(Nilsson and Hansen, 2018)	Yes
	Calanus helgolandicus	(Lauritano et al., 2011a, 2013, 2015)	Yes
	Calanus sinicus	(Lauritano et al., 2015)	Yes

Table 2. Commonly-used biomarkers for transcriptional analysis with real-time quantitative PCR in copepods. Copepod species abbreviations: *Acartia tonsa (A. tonsa); Calanus finmarchicus (C. finmarchicus); Calanus glacialis (C. glacialis); Calanus helgolandicus (C. helgolandicus); Eurytemora affinis (E. affinis)*; *Paracyclopina nana (P. nana); Pseudodiaptomus annandalei (P. annandalei); Pseudodiaptomus poplesia (P. poplesia); Tigriopus californicus (T. californicus); Tigriopus japonicus (T. japonicus);*

Biomarker	Stressors	Effect	Species	References
Aldehyde dehydrogenases (ALDHs)	Diatom toxins	Isoform dependent	*C. helgolandicus*	(Lauritano et al., 2011b., 2016)
Catalases (CATs)	Naphthalene	No significant change	*C. finmarchicus*	(Hansen et al., 2008)
	Toxic diatoms	No significant change	*C. helgolandicus*	(Lauritano et al., 2011b)
		Elevated expression	*C. helgolandicus*	(Lauritano et al., 2016)
Cell cycle and apoptosis regulatory 1 protein	Toxic diatoms	No significant change	*C. helgolandicus*	(Lauritano et al., 2011b)
Cellular apoptosis susceptibility protein	Toxic diatoms	Decreased expression	*C. helgolandicus*	(Lauritano et al., 2016)
Corticotropin Releasing Hormone Binding Protein	Salinity	Elevated expression	*T. japonicus*	(Lee et al., 2008c)
COI	Copper	Elevated expression	*T. japonicus*	(Weaver et al., 2016)
Cytochromes P450 (CYPs)	Diatom toxins	Isoform dependent	*C. helgolandicus*	(Lauritano et al., 2011b, 2016)
	Naphthalene	Concentration dependent down regulation	*C. finmarchicus*	(Hansen et al., 2008)
	Diethanolamine	Decreased expression	*C. finmarchicus*	(Hansen et al., 2010)
	Water-soluble fractions of crude oil and oil droplets	Stressor dependent	*C. finmarchicus*	(Hansen et al., 2009)
	Polycyclic aromatic hydrocarbons, water accommodated fractions of crude oil	Elevated expression	*P. nana*	(Han et al., 2015a)
Delta-1 pyrroline-5-carboxylase reductase	Salinity	No significant change	*T. californicus*	(Willett and Burton, 2002)
Delta-pyrroline-5-carboxylate synthase	Salinity	No significant change	*T. californicus*	(Willett and Burton, 2002)
DnaJ homolog	Increasing temperature	No significant change	*C. finmarchicus*	(Smolina et al., 2015)
Ferritin	Crowding	No significance	*A. tonsa*	(Nilsson et al., 2017)
	Quiescence	Time dependent peaks	*A. tonsa*	(Nilsson et al., 2014)
	Epibiont infestation	Elevated expression	*A. tonsa*	(Petkeviciute et al., 2015)

	CO_2 pressure	Acclimatization dependent (elevated expression for coastal copepods compared to estuarine copepods)	*A. tonsa*	(Aguilera et al., 2016)
	Nickel	Elevated expression	*P. annandalei*	(Jiang et al., 2013)
Glucose-Regulated Protein, 78kDa	Temperature; salinity shock	Elevated expression	*E. affinis*	(Xuereb et al., 2012)
Glutamate Dehydrogenase	Increasing temperature	Elevated expression, time dependent	*C. finmarchicus*	(Smolina et al., 2015)
	Salinity	No significant change	*T. californicus*	(Willett and Burton, 2003)
Glutathione Peroxidases (GPxs)	Pyrene; naphthalene	No significant change	*P. poplesia*	(Zhuang et al., 2017)
Glutathione Reductase (GR)	Salinity	Elevated for high salinities, decreased for low	*T. japonicus*	(Seo et al., 2006b)
	Heavy metals (Cu, Mn)	Elevated expression	*T. japonicus*	(Weaver et al., 2016)
	Hydrogen peroxide	Concentration and time dependent	*T. japonicus*	(Seo et al., 2006b)
Glutathione S-transferases (GSTs)	Endocrine disrupting chemicals	Concentration and stressor dependent	*T. japonicus*	(Lee et al., 2006)
	Toxic diatoms and dinoflagellates	Time and/or isoform dependent	*C. helgolandicus*, *C. finmarchicus*	(Lauritano et al., 2011b., 2016; Roncalli et al., 2016b)
	Heavy metals (Cu, Mn, Ag, As, Cd)	Time, metal-type and isoform dependent	*T. japonicus*	(Lee et al., 2007, 2008b)
	Naphthalene	Elevated expression	*C. finmarchicus*	(Hansen et al., 2008)
		Elevated expression (peak)	*P. poplesia*	(Zhuang et al., 2017)
	Hydrogen peroxide	Time dependent	*T. japonicus*	(Lee et al., 2007)
	Pyrene	Time and isoform dependent	*P. poplesia*	(Zhuang et al., 2017)
	Water-soluble fractions and water accommodated fractions of crude oil, oil droplets	Concentration - and time dependent	*C. finmarchicus*, *C. glacialis*	(Hansen et al., 2009, 2011)
	Diethanolamine	Concentration and time dependent	*C. finmarchicus*	(Hansen et al., 2010)
	Triphenyltin chloride	Decreased expression	*T. japonicus*	(Yi et al., 2014)
Glutathione Synthase	Toxic diatoms	No significant change	*C. helgolandicus*	(Lauritano et al., 2011b)
	Diethanolamine	Concentration dependent	*C. finmarchicus*	(Hansen et al., 2010)
Heat-shock protein 10, 20, 21, 22, 40, 60, 94 or 105 kDa	Temperature	No significant change	*T. japonicus*	(Rhee et al., 2009)
		Elevated expression for high and, decreased expression for low temperatures	*T. japonicus*	(Seo et al., 2006c)
	Salinity	No significant change	*T. japonicus*	(Seo et al., 2006b)
	Endocrine disrupter chemicals	Stressor dependent	*T. japonicus*	(Seo et al., 2006b)
	Handling	Elevated expression	*C. finmarchicus*	(Aruda et al., 2011)
	Diapause	Elevated expression	*C. finmarchicus*	(Aruda et al., 2011)
	Toxic diatoms	Decreased expression	*C. helgolandicus*	(Lauritano et al., 2011b)
Heat-shock protein 70kDa (HSP70)	Heat-shock / increasing temperature	Elevated expression (acclimatization result in lower expression)	*A. tonsa*	(Petkeviciute et al., 2015; Rahlff et al., 2017)
			E. affinis	(Rahlff et al., 2017)
			T. japonicus	(Rhee et al., 2009)
			T. californicus	(Chan et al., 2014)
			C. finmarchicus	(Voznesensky et al., 2004)
		No significance	*C. finmarchicus*	(Smolina et al., 2015)

	Handling	Elevated expression	*C. finmarchicus* *E. affinis*	(Aruda et al., 2011; Rahlff et al., 2017) (Rahlff et al., 2017)
	Salinity	Concentration dependent, but in general elevated expression outside acclimatization range	*A. tonsa*	(Petkeviciute et al., 2015)
	Crowding	No significance	*A. tonsa*	(Nilsson et al., 2017)
	Quiescence	Time dependent	*A. tonsa*	(Nilsson et al., 2014)
	CO_2 pressure	Acclimatization dependent (elevated expression for coastal copepods compared to estuarine copepods)	*A. tonsa*	(Aguilera et al., 2016)
	Shallow active vs. deep diapausing copepods	Elevated expression for active copepods in shallow waters	*C. finmarchicus*	(Aruda et al., 2011)
	Naphthalene	No significant change	*C. finmarchicus*	(Hansen et al., 2008)
	Toxic diatoms	Elevated expression	*C. helgolandicus*	(Lauritano et al., 2016)
		No significant change	*C. helgolandicus*	(Lauritano et al., 2011b)
	Heavy metals (Cu, Ag, Zn)	Elevated expression	*T. japonicus*	(Rhee et al., 2009)
	Endocrine disrupting chemicals	Concentration and /or stressor dependent	*T. japonicus*	(Rhee et al., 2009; Yi et al., 2014)
Heat-shock protein 90kDa (HSP90)	Heat shock / increasing temperature	No significant change	*T. japonicus*	(Rhee et al., 2009)
		Elevated expression	*E. affinis*	(Xuereb et al., 2012)
		Elevated expression	*A. tonsa*	(Petkeviciute et al., 2015)
	Salinity shock	Concentration dependent, elevated expression outside acclimatization range	*A. tonsa*	(Petkeviciute et al., 2015)
			E. affinis	(Xuereb et al., 2012)
	Naphthalene	No significant change	*C. finmarchicus*	(Hansen et al., 2008)
	Crowding	No significant change	*A. tonsa*	(Nilsson et al., 2017)
Inhibitor of apoptosis protein	Toxic diatoms	Decreased expression	*C. helgolandicus*	(Lauritano et al., 2011b)
Methylmalonate-semialdehyde dehydrogenase	Pyrene	Elevated expression	*P. poplesia*	(Zhuang et al., 2017)
Myohemerythrin-1	Nickel	Elevated expression	*P. annandalei*	(Jiang et al., 2013)
Nucleosome Assembly Protein 1	Increasing temperature	Elevated expression, time dependent	*C. finmarchicus*	(Smolina et al., 2015)
p53 tumor suppressor protein	Endocrine disrupting chemicals	Elevated expression	*T. japonicus*	(Hwang et al., 2010a)
Peroxiredoxin-6	Naphthalene	Elevated expression	*P. poplesia*	(Zhuang et al., 2017)
Ras-related C3 botulinum toxin substrate 1	Naphthalene	Elevated expression	*P. poplesia*	(Zhuang et al., 2017)
Retinoid X receptor	Triphenyltin chloride	Decreased expression	*T. japonicus*	(Yi et al., 2014)

Ribosomal protein L13	Nickel	Elevated expression	*P. annandalei*	(Jiang et al., 2013)
Ribosomal Protein S11	Increasing temperature	No significant change	*C. finmarchicus*	(Smolina et al., 2015)
Separase	Nickel	Elevated expression	*P. annandalei*	(Jiang et al., 2013)
Superoxide dismutases (SODs)	Heavy metals (Cu, Zn, Ag)	Elevated expression at high concentrations	*T. japonicus*	(Kim et al., 2011)
	Endocrine disrupting chemicals	Concentration and stressor dependent	*T. japonicus*	(Kim et al., 2011)
	Naphthalene	No significant change	*C. finmarchicus*	(Hansen et al., 2008)
	Toxic diatoms	No significant change	*C. helgolandicus*	(Lauritano et al., 2011b)
		Decreased expression	*C. helgolandicus*	(Lauritano et al., 2016)
	Nickel	Elevated expression (dose response)	*P. annandalei*	(Jiang et al., 2013)
Toll-like receptor	Increasing temperature	Elevated expression	*T. californicus*	(Chan et al., 2014)
Tubulins	Toxic diatoms	Decreased expression	*C. helgolandicus*	(Lauritano et al., 2011b)
	Nickel	Elevated expression	*T. japonicus*	(Jiang et al., 2013)
Ubiquitin	Naphthalene	No significant change	*C. finmarchicus*	(Hansen et al., 2008)
	Diethanolamine	Concentration dependent	*C. finmarchicus*	(Hansen et al., 2010)
Vitellogenin	Heavy metals (Cd, Ag, As, Cu)	Metal dependent	*T. japonicus*	(Lee et al., 2011)
		Elevated expression	*P. nana*	(Hwang et al., 2010b)

Table 3. Copepod transcriptome assemblies. Type is the type of Illumina sequencing in terms of read length (bp) and if it is paired-end (PE), or single-end (SE) sequencing. Reads used in assembly are reported in millions; the number of pairs is reported when paired reads were generated. Software gives the used assembly software ("CLC" indicates CLC Genomics Workbench, Trinity versions indicated if reported). Contigs is the number of resulting contigs within the assembly. BUSCO analyses were performed in August 2018 using publicly accessible NCBI Transcriptome Shotgun Assemblies (TSA); the rounded percent of complete (C, includes both single and duplicated) and complete single-copy (S) transcripts is shown.

Species	Sample	Platform	Type	Reads Used	Software	Contigs	BUSCO %	NCBI accession	Resources
Acartia tonsa	1 adult female	Illumina NextSeq	150 PE	~350 M	Trinity 2.3.2	60,662	99% C 45% S	**GFWY**00000000	(Nilsson et al., 2018)
Acartia tonsa	Multiple eggs, nauplii, copepodites, and adults	Illumina Next/Mi-Seq	150 PE	111 M	Trinity 2.5.1	119,439	88% C 45% S	**HAGX**00000000	(T.S. Jørgensen et al., unpublished data)
Apocyclops royi	Multiple eggs, nauplii, copepodites, and adults	Illumina NextSeq	150 SE	203 M	Trinity v. 2.5.1	75,477	na	**GHAJ**00000000	(Jørgensen et al., in press),
Calanus finmarchicus	Multiple eggs, nauplii, copepodites, and adults	Illumina HiSeq	100 PE	28 M per stage 640 M total	Trinity r2012-03-17-IU ZIH TUNED	206,012	99% C 72% S	**GAXK**00000000	(Lenz et al., 2014)
Calanus finmarchicus	3 CV stage copepodites	Illumina HiSeq	100 PE	93 M	Trinity r2012-06-08	241,140	97% C 61% S	**GBFB**00000000	(Tarrant et al., 2014)
Calanus finmarchicus	5 individuals exposed to short – and long-term thermal stress (3 temperatures)	Ion-Torrent	N/A	5 M	Trinity r2013-08-14	28,954	13% C 10% S	**GBXU**00000000	(Smolina et al., 2015)
Calanus glacialis	5 individuals short – and long-term stress exposed to 3 temperatures.	Ion-Torrent	N/A	3.5 M	Trinity r2013-08-14	36,880	21% C 14% S	**GBXT**00000000	(Smolina et al., 2015)
Calanus glacialis	10 CV stage copepodites	454 GS FLX	N/A	720 K	Mira v 3.0	54,344	17% C 12% S	**HACJ**00000000	(Ramos et al., 2015)
Eurytemora affinis	Females exposed to two strains of Vibrio and females not exposed	Illumina HiSeq	100 PE	300 M	Trinity r2013-08-14	138,088	100% C 65% S	**GBGO**00000000	(Almada and Tarrant, 2016)
Eurytemora affinis	Pooled adult males and females	Illumina HiSeq	100 PE	Not reported	Trinity r2013-11-10	107,445	71% C 29% S	**GEAN**00000000	Unpublished, Munro, J.B., Posavi, M., Brady, A., Orvis, J., Nadendla, S., Abolude, K., Kumari, P., Shetty, A. Lee, C.E. and Silva, J.C.
Labidocera madurae	5-6 adult females 15-26 CIII-CV stage copepodites	Illumina NextSeq	150 PE	530 M	Trinity v. 2.0.6	211,002	99% C 59% S	**GFWO**00000000	(Roncalli et al., 2017a)
Neocalanus flemingeri	1 adult female	Illumina NextSeq	150 PE	150 M	Trinity v. 2.0.6 & CAP3	140,841	98% C 62% S	**GFUD**00000000	(Roncalli et al., 2018a)

Paracyclopina nana	Unknown - adults	Illumina HiSeq	100 PE	200 M	Trinity	60,687	85% C 64% S	**GCJT**00000000	(Lee et al., 2015)
Pleuromamma xiphias	Pooled adult males and females	Illumina HiSeq	125 PE	267 M	Trinity v. 2.1.1	554,991	85% C 76% S	**GFCI**00000000	(Maas et al., 2018)
Temora longicornis	Pooled adult males and females (7:10 males:females ratio)	Illumina HiSeq	150 PE	460 M	Trinity v. 2.1.1	179,569	80% C 67% S	**GGQN**01000000	(Semmouria et al., 2019)
Tigriopus californicus	4-500 mixed developmental stages. San Diego population	Illumina HiSeq	100 PE	128 M	CLC v. 5.1	36,620	63% C 61% S	**GBSZ**00000000	(Barreto et al., 2014)
Tigriopus californicus	4-500 mixed developmental stages. Santa Cruz population	Illumina HiSeq	100 PE	49 M	CLC v. 5.1	43,077	62% C 61% S	**GBTC**00000000	(Barreto et al., 2014)
Tigriopus japonicus		Illumina HiSeq	100 PE	108 M	Trinity	54,758	99% C 55% S	**GCHA**00000000	(Kim et al., 2015b)
Tigriopus kingsejongensis	200 adults	Illumina HiSeq	150 PE	140 M	Trinity v. 2.0.6	38,250	98% C 77% S	**GDFW**00000000	(Kim et al., 2016)
Tisbe holothuriae	Multiple eggs, nauplii, copepodites, and adults	Illumina NextSeq	150 PE	162 M	Trinity v. 2.2.0	218,711	46% C 28% S	**HAHV**00000000	Roskilde University, 2017

Table 4. Overview of sequencing depth and number of replicates for differential gene expression assessment in copepods by RNA sequencing. M: million reads. # rep: replicates used per treatment. Type: type of used animals, field-caught or culture-reared. Seq. depth: sequencing depth.

Species	Treatments	# Individuals	Seq. depth	Method	# rep.	Type	Reference
Acartia tonsa	Control, salinity shock, handling stress	10 adults	~25 M	Kallisto/Sleuth	3	Culture	(Nilsson et al., 2018)
Calanus finmarchicus	Control and feeding on the neurotoxic *Alexandrium fundyense*	~74-80 pooled nauplii	~20 M	Bowtie, Custom script, edgeR	3	Field	(Roncalli et al. 2017b)
Calanus finmarchicus	Control, low-dose – and high-dose of dinoflagellate, two times	10 adult females	~26 M	Bowtie, edgeR	3	Field	(Roncalli et al., 2016a)
Calanus finmarchicus	Short – and long-term stress exposed to 3 temperatures	5 individuals	~0.8M	Rsubread, featureCounts, DESeq2	1	Field	(Smolina et al., 2015)
Calanus finmarchicus	CIV to CV molt (collected on day 3 and 10)	3 CV stage copepodites	~12M	Bowtie, RSEM, edgeR	4	Field/Culture	(Tarrant et al., 2014)
Calanus glacialis	Four pH treatments	10 nauplii	~28 M	BWA, Custom script (De Wit et al. 2012), DESeq2	5-6	Field	(Bailey et al., 2017)
Calanus glacialis	Control and 5 temperature treatments	10 CV stage copepodites	~60–180K	IDEG6	1	Field	(Ramos et al., 2015)
Calanus glacialis	Short – and long-term stress exposed to 3 temperatures	5 individuals	~0.6M	Rsubread, featureCounts, DESeq2	1	Field	(Smolina et al., 2015)
Eurytemora affinis	Control, exposure to acetone, pyriproxyfen and chlordecone. 3 experimental replicates pooled for sequencing	400-500 females or males	~11- 50 M	Bowtie2, RSEM, EBseq	1	Field	(Legrand et al., 2016)
Eurytemora affinis	Control, exposure to *Vibrio sp.* (F10) and *V. ordali*	20 females	~25 M	Bowtie, RSEM, edgeR	4	Culture	(Almada and Tarrant, 2016)
Neocalanus flemingeri	Diapause, emergence phase from diapause (10 weeks)	6 females (diapause), 4 females (emergence phase)	~10-22 M	Kallisto/ edgeR	3 per week	Field	(Roncalli et al., 2018b)
Pseudocalanus acuspes	7 pCO_2 exposure conditions	28-74 adults	~12-25M	Custom ANCOVA analysis	2	Field	(De Wit et al., 2016)
Temora longicornis	4 temperature treatments	50 adults (7:10 males:females ratio)	Unknown	Bowtie, TopHat, Cuffmerge, HTseq, edgeR	3	Field	(Semmouria et al., 2019)
Tigriopus californicus	2 populations, 2 temperatures	>300 copepods, mixed stages	2.5-7 M	CLC genomics workbench, Z-test	1	Field	(Schoville et al., 2012)
Tigriopus californicus	2 distinct populations and 2 interpopulation crosses	400-500 pooled	~3-40 M	CLC genomics workbench, edgeR	2-3	Field	(Barreto et al., 2014)
Tigriopus californicus	2 populations, 3 salinities	50 pooled	~24 M	RSEM, DESeq2	3	Cultured 2 generations	(DeBiasse et al., 2018)

Tigriopus californicus	Control (parent populations) and F4 crossed population exposed control and high-temperature selection conditions.	30 adults	Unknown	RSEM, limma	3 genetic lines	Field/Culture	(Kelly et al., 2017)
Tigriopus californicus	4 sites, 2 thermal regimes, 2 times	100 pooled	~7.6-29M	CLC genomics workbench, edgeR	2	Field	(Lima and Willett, 2017)

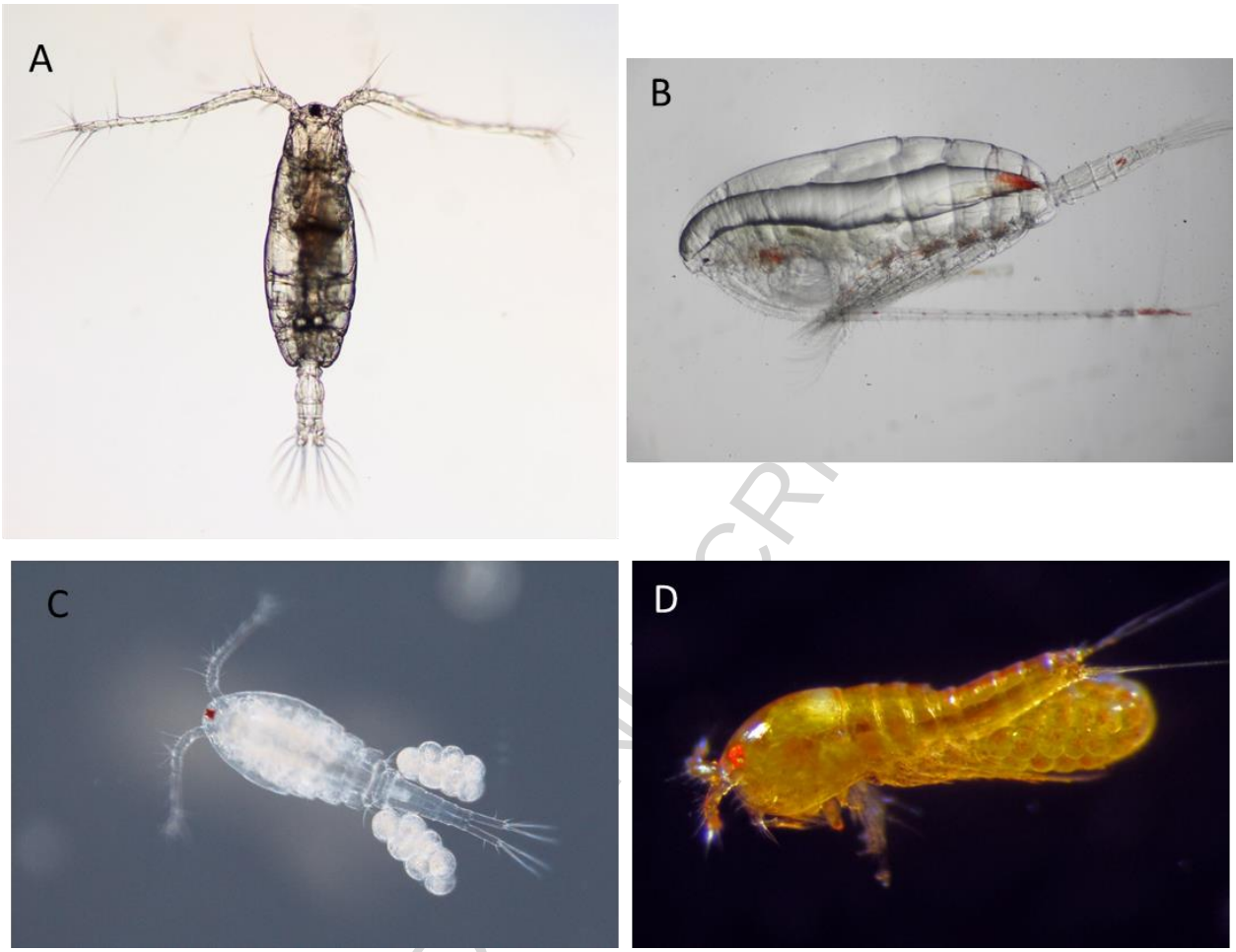


Figure 1. Examples of diverse marine copepods utilized in physiological studies. (A) The calanoid *Acartia tonsa* male, (B) The calanoid *Calanus glacialis* C5 copepodite with prominent oil sac, (C) The cyclopoid *Apocyclops royi* egg-bearing female (D) The harpacticoid *Tigriopus japonicus* egg-bearing female. Of these, *Tigriopus spp.* (particularly *T. californicus*, not shown) have been extensively developed as a model for studies of molecular evolution and plasticity. The others represent a growing diversity of species for which molecular physiology studies are being driven by their ecological importance. Photos courtesy of Dr. Minh Thi Thui Vu (A), A.M.T. (B), Dr. Hans van Someren Gréve (C), and Professor Hans Uwe Dahms (D).

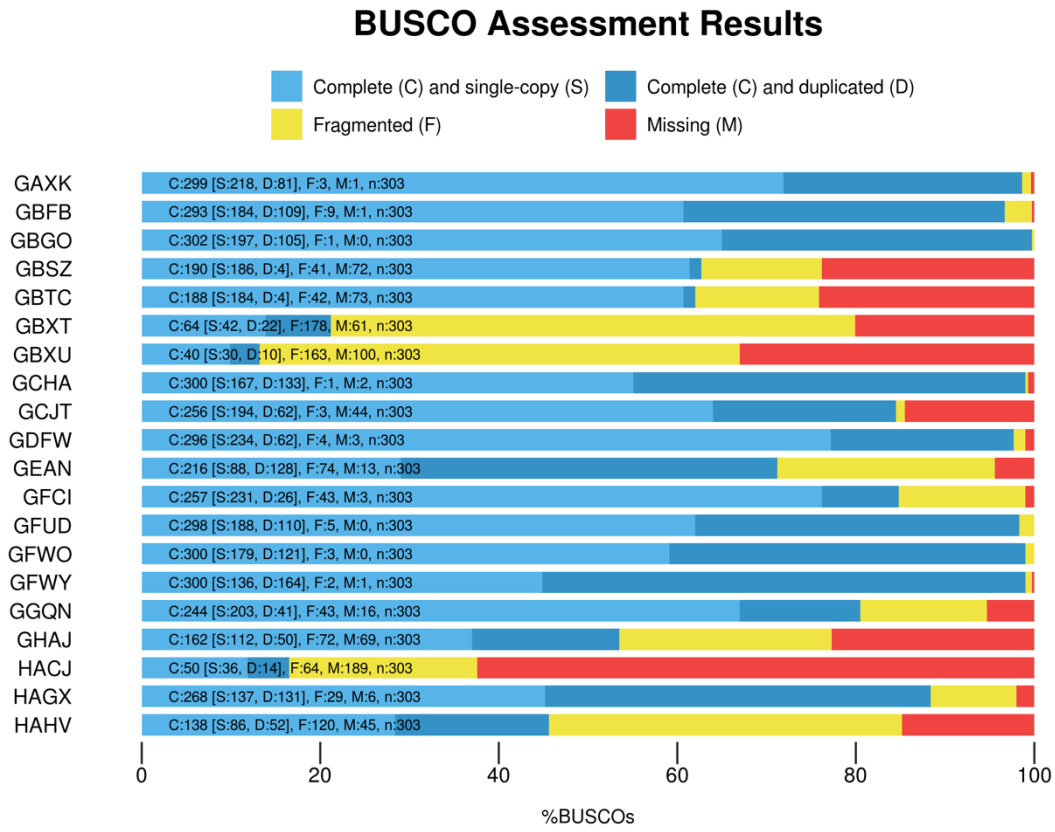


Figure 2. BUSCO analysis of the following copepod transcriptomes (with NCBI accession numbers, alphabetized by scientific name as in Table 3): **GFWY**: *Acartia tonsa* (GFWY00000000.1, 27-sep-2017); **HAGX**: *Acartia tonsa* (HAGX00000000.1, 29-sep-2017); **GAXK**: *Calanus finmarchicus* (GAXK00000000.1, 14-may-2018); **GBFB**: *Calanus finmarchicus* (GBFB00000000.1, 30-jan-2015); **GBXU**: *Calanus finmarchicus* (GBXU00000000.1, 13-jan-2015); **GBXT**: *Calanus glacialis* (GBXT00000000.1, 13-jan-2015); **HACJ**: *Calanus glacialis* (HACJ00000000.1, 29-sep-2017); **GBGO**: *Eurytemora affinis* (GBGO00000000.1, 07-jul-2015); **GEAN**: *Eurytemora affinis* (GEAN00000000.1, 16-nov-2016); **GFWO**: *Labidocera madurae* (GFWO00000000.1, 14-may-2018); **GFUD**: *Neocalanus flemingeri* (GFUD00000000.1, 14-may-2018) ; **GCJT**: *Paracyclopina nana* (GCJT00000000.1, 20-jul-2015); **GFCI**: *Pleuromamma xiphias* (GFCI00000000.1, 18-dec-2017); **GGQN**: *Temora longicornis* (GGQN00000000, 12-jun-2018); **GBSZ**: *Tigriopus californicus* (GBSZ00000000.1, 02-feb-2015); **GBTC**: *Tigriopus californicus* (GBTC00000000.1, 02-feb-2015); **GCHA**: *Tigriopus japonicus* (GCHA00000000.1, 20-jul-2015) ; **GDFW**: *Tigriopus kingsejongensis* (GDFW00000000.1, 18-apr-2016); **HAHV**: *Tisbe holothuriae* (HAHV00000000.1, 23-jan-2018).

Graphical abstract